Université Lumière Lyon 2
Ecole Doctorale ECLIPS
Faculté des Lettres, Sciences du Langage et Arts
Département des Sciences du Langage : linguistique et phonétique

Analyse linguistique d'un corpus de communication en réseau : Étude morphosyntaxique.

Par Florence Chenu

Thèse de doctorat en Sciences du langage.
Dirigée par Gilbert PUECH
Présentée et soutenue publiquement le 1er septembre 2003

Devant un jury composé de : Claire Blanche-BENVENISTE, Professeur de l'Université de Provence (rapporteur). Catherine KERBRAT-ORECCHIONI, Professeur de l'Université Lumière Lyon 2. Michel MARCOCCIA, Maître de Conférence à l'Université de Technologie de Troyes. Gilbert PUECH, Professeur de l'Université Lumière Lyon 2. Sylviane REMY, Professeur de l'Université Lumière Lyon 2.

Table des matières

REMERCIEMENTS . 1

. . 3

Introduction . . 5

Chapitre 1 Internet et communication . . 9

1.1. Internet et services . . 9

1.1.1. Navigation hypertextuelle . 10

1.1.2. Transfert de fichier et contrôle distant . 13

1.1.3. Courrier électronique . . 13

1.1.4. Groupes de discussion . 13

1.1.2. Salles de discussion . 14

1.2. Les langues et Internet . . 15

1.2.1. Société de l'information . 15

1.2.2. Représentation des langues sur Internet . . 17

Conclusion . 20

Chapitre 2 Communication Interactive . 21

2.1. Communication interactive interpersonnelle . 22

2.1.1. Types de communication synchrone textuelle en réseau. . . 22

2.1.2. Choix pour cette étude . . 26

2.2. Déroulement technique d'une session . . 28

2.2.1. Script général . . 28

2.2.2. Connexion . 29

2.2.3. Propriétés des canaux . . 32

2.2.4. Statuts des participants . 33

2.2.5. Activité du canal . . 34

2.3. Première caractérisation de la communication IRC . . 36

Conclusion . 37

Chapitre 3 Description du corpus brut . 39

3.1. Constitution des corpus . 39

3.2. Présentation générale des corpus . 42

3.3. Format des fichiers bruts . 45

3.3.1. Frontières de sessions . . 46

3.3.2. Activité du canal . . 47

3.4. Exploitation des fichiers bruts. . 50

3.4.1. Présentation quantitative des types de lignes des corpus . . 50

3.4.2. Rythme de l'activité . 53

3.4.3. Présentation quantitative des sous-types de lignes . . 55

3.4.4. Rythme de l'activité conversationnelle et longueur des messages . 59

3.4.5. Mode de contribution des participants à l'activité . 63

3.5. Conclusion . 67

Chapitre 4 Traitement des corpus . 69

4.1. Traitement des interventions dans des codes et langues différentes . 70

4.1.1. Principes du codage . 71

4.1.2. Notes concernant le français dans les corpus . . 74

4.2. Standardisation et lemmatisation . . 77

4.2.1. Standardisation . 78

4.2.2. Lemmatisation . 80

4.3. Anonymisation des corpus . . 81

4.3.1. Justifications . 82

4.3.2. Procédure . 82

4.3.3. Pertes . 85

4.4. Logiciel de traitement des corpus . 86

4.5. Conclusion . 89

Chapitre 5 Le code écrit des IRC . 91

5.1. Code verbal . 93

5.1.1. Orthographe . . 93

5.1.2. Abréviations . . 94

5.1.3. Autres phénomènes de réduction formelle . 111

5.1.4. Phénomènes phonétiques représentés n'impliquant pas de réduction formelle . 117

5.1.5. Signes de ponctuation . . 132

5.2. Groupes de caractères . . 135

5.2.1. Définitions . 135

5.2.2. Emplois des frimousses . . 149

5.2.3. De l'usage des frimousses . . 157

5.3. Conclusion . 158

Chapitre 6 Unités conversationnelles . 161

6.1. Caractérisation du type de comportement verbal, ou type d'interaction. . . 162

6.1.1. Contexte . 167

6.2. Gestion des tours de parole . . 171

6.2.1. Statut du non locuteur, identification du/des destinataire(s) . 173

6.2.2. Changement de tour et successeur . . 176

6.3. Unités conversationnelles . 180

6.3.1. Séquences d'ouverture et de clôture . . 180

6.3.2. Unités conversationnelles thématiques . . 185

6.3. Relation interpersonnelle . 186

6.4. Conclusion . 189

Chapitre 7 Matériel lexical . 191

7.1. Traitement des corpus. . 192

7.1.1. Questionnement théorique . 192

7.1.2. Dilemmes . . 195

7.2. Listes de fréquences . 199

7.2.1. Diversité lexicale . 204

7.2.2. Densité Lexicale . 206

7.3. Spécificités lexicales . 208

7.3.1. Emprunts . 208
7.3.2. Intégration des emprunts . . 214
7.3.3. Morphologie ludique . 216
7.3.4. Français relâché et verlan . 217
7.3.5. Éléments de variétés régionales du français. . 219
7.4. Particularités d'ordre morphologique . 220
7.4.1. Indices pronominaux . 220
7.4.2. Flexions des termes IRC . 221
7.5. Conclusion . 223
Chapitre 8 Structures syntaxiques . 225
8.1. Unité de segmentation. . 226
8.1.1. Problématique . 226
8.1.2. Solution adoptée . 227
8.2. Phrase simple vs complexe . 238
8.2.1. Relatives . 242
8.2.2. Complétives . . 247
8.2.3. Adverbiales . 247
8.2.4. Bilan . . 249
8.3. Types de phrases . . 250
8.3.1. Structures interrogatives . 252
8.3.2. Structures injonctives . . 254
8.3.3. Structures exclamatives . . 254
8.3.4. Structures déclaratives . 255
8.4. Autres phénomènes syntaxiques et morphosyntaxiques . . 261
8.4.1. Absence de marqueur d'enchâssement . 261
8.4.2. Spécificité du marquage du discours rapporté . 262
8.5. Conclusion . 263
Chapitre 9 Temporalité . 265
9.1. Le concept de temps . 265

9.2. Les moyens d'expression de la temporalité dans le corpus . . 271

9.2.1. Morphologie verbale . 271

9.2.2. Autres marques de temporalité . 282

9.3. La référence temporelle dans les corpus . . 288

9.3.1. Valeurs des temps . . 289

9.3.2. Combinaisons principales concernant la référence temporelle . . 300

9.3.3. Tiroir verbal et négation . . 304

9.3.4. Adverbes temporels, tiroir et aspect lexical . . 305

9.3.5. Adverbes temporels déictiques, tiroir et aspect lexical . 307

9.3.6. Noms temporels, prépositions temporelles . 309

9.4. Phénomènes de construction narrative coopérative. . . 316

9.5. Conclusion . 320

Conclusion . 321

Bibliographie . . 325

REMERCIEMENTS

Je tiens à exprimer toute ma reconnaissance à tous les membres du laboratoire Dynamique du Langage où j'ai pu trouver un environnement de travail stimulant et performant.

Tout particulièrement je voudrais dire ma gratitude à Monsieur le professeur Gilbert Puech, qui malgré les fonctions prenantes qu'il assume a su me faire profiter de la qualité de sa réflexion et de ses savoirs.

Mes remerciements vont aussi à Frédérique Gayraud, et Catherine Kerbrat-Orecchioni pour leur relecture attentive.

Que soient ici également remerciés tous ceux qui m'ont techniquement permis de concevoir cet objet : les informaticiens du laboratoire Christian Fressard pour ses conseils avisés, François Pellegrino et Egidio Marsico pour les questions de sauvegarde, le SIR de Lyon2 et la division ERAD pour leur disponibilité, Monsieur Lamande pour les travaux de reprographie.

Enfin, et ce ne sont pas les moins méritants, ceux qui m'ont supportée dans les moments les meilleurs et les plus difficiles, quotidiennement, doivent savoir que je leur suis spécialement reconnaissante : Frédérique, Fanny, Egidio, Laurent, Olivier, Caroline, Christophe, Géraldine, Papa, Maman, Marianne, Mina, Emmanuel, Marie, Corine et son chat, Cristelle, Florence, et tant d'autres…

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Pour Anna.

en vertu de la loi du droit d'auteur.

en vertu de la loi du droit d'auteur.

Introduction

L'avènement d'Internet a fait naître des moyens de communication qui reposent sur des dispositifs nouveaux. Ces derniers mettent en contact des humains proches ou distants, d'une même communauté linguistique ou de communautés linguistiques différentes, d'une même culture ou de cultures différentes…

Avec Internet se sont développés de nombreux champs de recherches, tant concernant la technologie que ses usages. Les *Internet Studies* florissent, et c'est sous ce label que des recherches issues des champs disciplinaires traditionnels se rassemblent. Sociologues, historiens, économistes, anthropologues, didacticiens, linguistes… octroient une part de leur intérêt à ce phénomène né en fin de siècle. Parmi ces disciplines, certaines, qui ont affaire à la langue et à son usage, se reconnaissent plus spécifiquement sous le label d'études en CMO [1] .

Le linguiste, en parallèle avec quelques préoccupations du commun des mortels des sociétés occidentales, s'enquiert de ce qui se joue dans ces dispositifs pour la langue et son usage.

L'étude de l'usage linguistique sur Internet, on ne s'en étonnera pas, se consacre essentiellement à l'anglais. Une bonne synthèse des travaux existants sur les différents dispositifs de communication d'Internet pour l'anglais se trouve dans Crystal (2001).

Les études sur d'autres langues sont peu nombreuses, et naissent souvent de préoccupations de linguistique appliquée à l'apprentissage des langues ou dans des situations de travail. On peut noter que ces perspectives travaillent aussi à la mise en place de situations artificielles sur le réseau pour l'apprentissage des langues et que nombre d'études portent sur les produits linguistiques qui en sont issus [2] .

Pour ce qui concerne le français, les études sur corpus naturel sont peu nombreuses [3] . Elles sont soit très générales [4] , soit très spécifiques. Anis et Luzzatti s'intéressent, par exemple, aux spécificités de l'écrit dans les dispositifs techniquement médiatés (minitel, IRC, etc.).

L'étude d'une langue est une entreprise considérable si on souhaite prendre en compte toutes ses dimensions. Cette recherche se propose d'examiner un certain nombre de paramètres linguistiques tels qu'ils sont mis en œuvre dans l'un de ces nouveaux moyens mis à la disposition des utilisateurs d'ordinateurs ayant accès à Internet : les IRC [5] . Ces paramètres ont été sélectionnés en fonction des contraintes imposées par le dispositif de façon à mesurer la portée

[1] Communication Médiatée par Ordinateur ; CMC dans la terminologie anglaise (Computer Mediated Communication). Cf. Biliographie établie par Marcoccia (2002)

[2] Cf. par ex. revue ALSIC (http://alsic.u-strasbg.fr/) ; Journal of Computer Assisted Language Learning (http://www.szp.swets.nl/szp/journals/ca.htm) ; etc.

[3] On peut toutefois penser que beaucoup d'études sur le français des réseaux restent dans l'ombre, les lieux de publication étant limités (il semble qu'un nombre en évolution de mémoires de maîtrise ou DEA s'y consacrent).

[4] Dejond (2002).

[5] *Internet Relay Chat* – Nous formulerons une description détaillée au Chapitre 2. Considérons pour l'instant qu'il s'agit d'un dispositif de communication sur Internet basé sur l'écrit dactylographié à échange synchrone.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

de ces dernières sur les stratégies de mise en discours.

La conception traditionnelle de l'usage linguistique dégage deux aspects essentiels : sa fonction de communication et sa fonction de représentation. On peut affirmer que l'un ne va pas sans l'autre. Comment en effet communiquer si aucune représentation n'est formée ? Des routes, des couloirs communiquent, c'est-à-dire permettent le passage d'éléments. Il n'y a là nullement besoin de représentation. La notion de communication implique simplement qu'on considère au moins deux entités entre lesquelles circulent des éléments. La communication humaine nécessite quant à elle des représentations. Ce qui circule n'est en effet pas pleinement concret. Les êtres humains doivent être envisagés comme des entités dont le mode d'existence requiert un système de traitement de l'information. La communication humaine est donc affaire de perception et de traitement de l'information. Les éléments qui passent, certes, n'appellent pas nécessairement des représentations conscientes de la part de l'entité source. Il n'y a toutefois pas, dans ce cas, de communication sans traitement par l'entité réceptrice. D'une part ou de l'autre, des représentations doivent être engendrées. La communication humaine, ainsi que nous l'avons entendue jusqu'ici, ne se limite pas à l'usage linguistique, mais en ce qui concerne justement le langage, une grande part des éléments qui "circulent" est intentionnellement mise en œuvre. Il y a donc, pour la communication linguistique, mise en œuvre de représentations complexes de part et d'autre. Notre étude de ce moyen de communication récent ne se souciera donc pas seulement du versant communicationnel impliqué.

L'usage linguistique peut se réaliser à travers différents canaux, et on considère alors une opposition fondamentale entre l'oral et l'écrit. Cette opposition est dite générer des différences linguistiques à de nombreux niveaux. Certes, l'oral est premier ontogénétiquement et phylogénétiquement et cela en fait sans doute un moteur du changement linguistique. L'écrit ne se résume pourtant pas à un simple mode de représentation du langage parlé, il n'est pas de système d'écriture purement phonographique, d'une part, et d'autre part, les différentes situations de communication impliquant l'usage de l'écrit plutôt que l'oral ont des influences sur la construction des messages linguistiques.

Il ne s'agit pas de dire ici qu'il existe une stricte opposition entre oral et écrit. Certes, les réalités physiques impliquées ne sont pas les mêmes, les systèmes de production (appareil phonatoire, membre supérieur) et de réception (système auditif, système visuel) engagés sont différents et impliquent donc des systèmes de traitements de l'information différents et des contraintes temporelles différentes (au niveau de l'encodage et décodage du message linguistique). De même, tandis que le message linguistique oral ne se présente pas de façon pré-segmentée, l'écrit présente différents niveaux de segmentation.

La littérature témoigne souvent de situations prototypiques d'usage de l'oral et de l'écrit pour mieux fonder l'opposition. Ainsi l'oral tendrait vers le versant communicationnel, tandis que l'écrit tendrait vers le versant représentationnel. Cette première dimension d'opposition se base essentiellement sur une conception simplificatrice qui envisage l'oral comme un acte à plusieurs et l'écrit comme un acte individuel. Le dispositif justement envisagé ici permet une situation de communication en temps réel à plusieurs dont le code est l'écrit. Cela permet également de mettre en défaut l'analyse selon laquelle l'oral est spontané et l'écrit planifié.

La littérature relève que dans l'oral prototypique, le fait que la situation de communication soit partagée permet qu'une partie de l'information reste implicite. Il en découle l'observation inverse pour l'écrit. Ainsi l'oral appellerait plus de recours à la prise en compte du contexte par exemple pour identifier un référent, ou construire des liens non explicités entre propositions. Ainsi, l'éloignement en temps et en espace des locuteurs impliqués dans l'échange de messages linguistiques est le paramètre qui fait dire à Chafe (1982) [6] que l'écrit a un caractère détaché

[6] Cette opposition s'apparente à celle entre récit et discours postulée par Benveniste (1974).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

tandis que l'oral présente un caractère impliqué manifestant des différences du point de vue de la construction des messages linguistiques. Cet éloignement implique également pour l'écrit l'impossibilité d'un *feed-back*, imposant alors que les messages délivrés soient autosuffisants, et donc, que le locuteur-scripteur construise la représentation la plus juste possible des connaissances de l'allocutaire. Le lexique ainsi serait affecté par la différence de chenal. Il est commun de dire que la contrainte temporelle influe sur les choix lexicaux (fréquence des termes) et implique redondance et paraphrase pour l'oral, précision et réseaux sémantiques pour l'écrit. La syntaxe de l'oral serait plus souple que celle de l'écrit et manifesterait davantage de parataxe et de structures incomplètes. L'écrit présenterait une syntaxe plus rigoureuse et des phénomènes de subordination, de coordination et davantage d'indices de cohésion. De même, l'utilisation des temps verbaux serait différente à l'oral et à l'écrit. Ainsi, le passé simple n'appartiendrait pas au système de l'oral, qui présenterait essentiellement présent et passé composé. Le système des temps de l'écrit serait plus complexe. Les marques d'énonciation et déictiques sont également dits appartenir au système de l'oral. Enfin, les systèmes paralinguistiques de l'oral et de l'écrit, de natures évidemment différentes, véhiculeraient des informations de qualités différentes (émotions *vs* structure).

Ainsi, il s'agit évidemment d'une opposition théorique qui discrétise à outrance un continuum entre oral et écrit. À l'exception de ce qui concerne les différences liées aux paramètres strictement attachés au canal, toutes les différences relevées le sont pour les situations de communication impliquant l'usage de l'oral plutôt que de l'écrit et inversement. Biber (1998) montre, en effet, que plusieurs dimensions textuelles remplacent avantageusement les dichotomies traditionnelles.

Nous essaierons dans ce travail de caractériser l'usage linguistique dans les IRC, en observant dans un corpus naturel un inventaire de traits linguistiques.

La présentation de ce travail se fera en neuf chapitres. Le premier chapitre se donne pour but de dresser le paysage Internet, d'une part en ce qui concerne les protocoles de communication entre les machines et les dispositifs de communication humaine qu'ils autorisent, et d'autre part en ce qui concerne la connaissance de la situation – ou des situations – linguistique(s) d'Internet. Cette première vue d'ensemble se donne donc pour objectif majeur de situer le dispositif de communication retenu pour cette étude dans l'environnement dans lequel il s'inscrit et dans son contexte contemporain.

Le second chapitre est plus précisément consacré à la présentation de l'un des grands types de communication humaine médiatisée par ordinateur sur Internet : la communication interactive. L'adjectif ici choisi est peu satisfaisant dans la mesure où la communication implique toujours dans une plus ou moins large mesure de l'interaction – il existe des degrés d'interactivité. Ce que nous entendons donc par cet adjectif, ainsi que le font nombre de chercheurs et les vulgarisateurs, c'est une interaction explicite et directe. Parmi les dispositifs présentés au chapitre premier, nous développons, dans le chapitre 2, les spécificités propres à la communication interactive, en mettant en avant la spécificité des IRC. Dans un second temps, nous exposons en détail le dispositif qui génère l'objet de notre étude en relevant quelques-unes de ces caractéristiques techniques. Cela nous permet de montrer les points sur lesquels notre attention se portera aux chapitres suivants.

Les chapitres 3 et 4 ont essentiellement une vocation de présentation de la méthodologie appliquée au recueil et au traitement des corpus. Ainsi, le chapitre 3 présente-t-il la technique de collecte des corpus et son esprit déontologique. Le corpus recueilli présente des spécificités, nous le décrivons donc dans sa matérialité, d'un point de vue relativement technique permettant de mettre en perspective les traitements qui doivent lui être appliqués pour une étude linguistique. Ce chapitre présente l'avantage d'une totale transparence au sujet de l'objet recueilli et permet la

en vertu de la loi du droit d'auteur.

reproduction de la méthodologie. Les premiers traitements décrits dans ce chapitre permettent, en outre, de catégoriser certains éléments du corpus et donc d'en donner une description quantitative. Cette dernière nous permet d'une part de comparer deux sous-ensembles du corpus, et d'autre part, d'établir certaines propriétés de nos corpus.

Le chapitre 4 présente des traitements de niveau linguistique mis en œuvre sur les corpus, notamment du point de vue du codage utilisé pour les différentes langues apparaissant dans le corpus, puis des procédures de standardisation et de lemmatisation pour le français. Ce chapitre est aussi l'occasion de présenter la procédure qui a été mise en œuvre pour l'anonymisation des corpus, le corpus naturel recueilli ne permettant pas d'obtenir l'autorisation des locuteurs. Enfin, une présentation rapide du logiciel retenu pour l'archivage et les traitements ultérieurs est proposée.

Après avoir présenté le dispositif, mis en abyme dans le contexte socioculturel dont il est issu, et les divers traitement qui ont été appliqués pour obtenir un objet formaté pour l'analyse, nous nous attacherons à décrire le code employé et ses spécificités. Le chapitre 5 a l'ambition de relever l'ensemble des phénomènes observés sur le français écrit dactylographié de nos corpus. Il s'agit là de mesurer l'écart à la norme et de tenter de comprendre le fonctionnement et le pourquoi de cet écart. Nous essaierons en effet d'évaluer ce qui se joue pour le code du point de vue de la stricte matérialité linguistique.

Le chapitre 6 se place à l'autre bout de l'échelle et présente les modalités de l'interaction. Les paramètres pris en compte pour cette analyse sont ceux qui sont couramment utilisés pour ce type d'analyse : il s'agit des paramètres identifiés depuis les premiers essais de modélisation de la communication humaine. Nous tentons de dégager les spécificités du dispositif qui impliquent une spécificité de l'interaction.

Les chapitres suivants entrent plus précisément dans l'analyse proprement linguistique. Les moyens lexicaux, puis les moyens syntaxiques mis en œuvre pour la production des messages sont exposés respectivement aux chapitres 7 et 8. En effet, parmi les paramètres linguistiques couramment évoqués pour établir des différences de registre, ou pour caractériser l'opposition écrit/oral, le lexique et la syntaxe sont essentiels. Enfin, le chapitre 9 étudie une dimension également relevée pour différencier des types de textes : la flexion verbale. Ce dernier chapitre sera l'occasion d'examiner les moyens d'expression de la temporalité afin de dresser un tableau des spécificités de nos corpus de ce point de vue. Nous trouverons également là le lieu de présenter un type d'unités conversationnelles particulier de nos corpus.

en vertu de la loi du droit d'auteur.

Chapitre 1 Internet et communication

La présente étude se donne pour but d'examiner du point de vue linguistique un type de communication en réseau dont le support est Internet. Nous nous proposons donc dans ce premier chapitre de formuler de façon succincte et peu technique les connaissances générales nécessaires à la compréhension de ce qu'est ce réseau qui forme un arrière plan à la communication en question.

L'utilisation des ordinateurs à l'origine dans le domaine scientifique a très vite fait naître le besoin de les relier entre eux, puis de permettre la communication humaine médiatisée par ordinateurs. Les premiers réseaux d'ordinateurs, qui n'étaient voués qu'à la communication entre machines pour le partage de ressources ou le transfert de données, se sont vite pourvus de systèmes logiciels pour la communication humaine. Du réseau local à Internet, les services disponibles ne changent, pour ce qui concerne leurs potentialités techniques, que par leur portée géographique.

Nous commencerons, dans ce chapitre, par présenter l'éventail des possibilités grand public que permet le réseau Internet, puis nous proposerons une évaluation de la situation linguistique qu'il manifeste en mettant en avant les difficultés posées à l'analyse.

1.1. Internet et services

Internet – abréviation anglaise de *Interconnected Network* [7] – se définit ainsi que son

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

nom l'indique comme un réseau d'ordinateurs interconnectés appartenant à des ensembles et sous-ensembles c'est-à-dire plus simplement un réseau de réseaux d'ordinateurs. Le projet qui a conduit à l'Internet [8] actuel est issu de l'ARPA (*Advanced Research Project Agency*), organisme de la défense américaine dont le but était d'établir un réseau fiable entre la défense et la recherche militaire et universitaire dans les années soixante. Dans ce contexte, fiable signifie que si une liaison du réseau venait à être inutilisable pour une raison ou une autre, la communication entre les machines ne devait pas être atteinte de façon majeure. Il s'agit donc d'un réseau sans centre. Le principe de conception de ce réseau est que les données qui transitent sur le réseau, en cas de défaillances, sont redirigées de façon automatique pour parvenir à destination [9]. L'Internet actuel s'est progressivement constitué par les rattachements successifs de réseaux. Les connexions physiques qui le constituent sont de différents types (RTC, fibre optique, satellite…) et les machines connectées ne forment pas non plus un ensemble homogène (différents types d'ordinateurs fonctionnant sous différents systèmes d'exploitation).

Au delà des connexions physiques, et des réseaux et sous réseaux qui se sont associés et représentent l'Internet actuel, ce qui fonde Internet, c'est la manière de communiquer : les protocoles de communication entre les machines.Internet est basé sur un ensemble de conventions [10] adoptées par des regroupements de réseaux d'ordinateurs et offre à l'échelle de la planète les possibilités d'un réseau informatique ordinaire : échange de données numériques (texte, son, images, vidéos, programmes...) ; messagerie électronique ; participation à des forums de discussions synchrones ou asynchrones; contrôle de machines distantes... Si Internet permet la communication entre des machines à l'échelle planétaire, la plupart des services Internet sont directement des moyens de communiquer pour les humains. Nous envisagerons les services principaux sur Internet : la navigation hypertextuelle, le transfert de fichier et le contrôle de machines distantes, le courrier électronique, les groupes de discussion et les salles de discussion.

1.1.1. Navigation hypertextuelle

Le service le plus connu d'Internet et qui a grandement participé à rendre Internet d'intérêt

[7] L'essentiel de cette section consacrée à Internet se base sur les éléments d'informations tirés des ouvrages suivants : Dufour (1995), Levine *et al.* (1994), Steyer (1998), et Tannenbaum (1997).

[8] Pour un historique de la constitution de l'Internet actuel on se reportera à Dufour (1995:25-40).

[9] Reroutage dynamique.

[10] Des organismes internationaux tels que l'ISOC http://www.isoc.org/internet/, qui se définit elle-même comme "une organisation globale et internationale destinée à promouvoir l'interconnexion ouverte des systèmes et l'Internet" sont en charge de mettre au point conventions et protocoles. Les principaux comités de l'ISOC sont l'*Internet Architecture Board* (IAB), chargé des évolutions des protocoles de communication ; l'*Internet Assigned Number Authority* (IANA) qui veille au système d'adressage ; l'*Internet Engineering Task Force* (IETF) qui fédère les travaux, spécifications et implémentation des technologies Internet et l'*Internet Research Task Force* (IRTF) qui se consacre à la recherche.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

pour un large public est le dernier-né des services : le *World Wide Web* – plus couramment dénommé le Web, et pour lequel la commission générale de terminologie et de néologie [11] recommande le terme de "toile d'araignée mondiale [12] " – service inventé au CERN [13] en Suisse, développé pour être un outil de collaboration pour la communauté des scientifiques en physique des hautes énergies. Basé sur le principe de l'hypertexte, le Web permet la mise à disposition pour l'ensemble des internautes d'informations consultables en tout point du réseau de façon simple. Le concept d'hypertexte est ancien. En 1945, Vannevar Bush, conseiller du président Roosevelt, avait imaginé dans les pages d'un article – "As We May Think" [14] – un système destiné à améliorer les capacités cognitives. Ce système, MEMEX (Memory Extender), serait composé d'une mémoire dans laquelle seraient stockés les divers éléments d'information et d'un mécanisme permettant d'y accéder. Il repose sur l'indexation associative des éléments d'information, imitant les associations d'idées de l'esprit humain. En 1963, Douglas Engelbart – Université de Stanford – reprend ces idées et initie le système NLS (*oN-Line System*) proposant des ressources de communications synchrones et asynchrones, dont il publie les fondements dans l'article "A Conceptual Framework for the augmentation of Man's Intellect" [15] . Les termes d'hypertexte et d'hypermédia sont inventés par Theodor Holm Nelson en 1965, initiateur du projet Xanadu [16] et reçoivent la définition suivante :

"Un ensemble d'écrits ou d'illustrations interconnectés d'une façon complexe qui ne permet pas de le représenter correctement sur du papier. Il peut contenir des sommaires ou des cartes de son contenu et leurs interrelations ; il peut contenir des annotations, additions et notes de bas de page d'universitaires qui les ont examinés." [17]

Un Hypertexte est un ensemble de données informatives numérisées qui peuvent se consulter et se lire selon des parcours divers [18] . La différence majeure qui l'oppose au

[11] Le programme d'action gouvernemental " Préparer l'entrée de la France dans la société de l'information ", en 1998, s'est allié les services de la commission générale de terminologie et de néologie pour élaborer des recommandations terminologiques à l'attention des différentes administrations –ainsi que l'orthographe académique fut imposée aux fonctionnaires en 1832–, l'argument étant que "L'existence d'une terminologie en français, claire et accessible, facilite l'appropriation des technologies de l'information et de la communication par le plus grand nombre." (Vocabulaire de l'informatique et de l'internet – Journal officiel du 16 mars 1999 http://www.culture.fr/culture/dglf/cogeter/16-03-99-internet.html).

[12] Simple traduction de la métaphore anglosaxone.

[13] Cf. Page du CERN, Organisation Européenne pour la Recherche Nucléaire : http://welcome.cern.ch/welcome/gateway.html, et Page du CERN à propos l'invention du Web : http://public.web.cern.ch/Public/ACHIEVEMENTS/web.html

[14] Vannevar Bush (1945) As We May Think – The Atlantic Monthly. http://www.isg.sfu.ca/~duchier/misc/vbush/vbush.shtml

[15] Engelbart (1963) "A Conceptual Framework for the Augmentation of Man's Intellect." Vistas in Information Handling. Ed. P.D. Howerton and D.C. Weeks. Washington, D.C.: Spartan Books.

[16] Nelson, T.H. (1978) "Dream Machines: New Freedoms Through Computer Screens–A Minority Report." Issued with "Computer Lib." South Bend IN: The Distributors. Et Nelson, T.H. (1990) "On the Xanadu Project", *BYTE Magazine* , Vol. 15, No. 9, Sep. 1990, pp. 298-299.

en vertu de la loi du droit d'auteur.

texte est que le texte propose un parcours fixe tandis que les données d'un hypertexte sont réparties en éléments ou noeuds d'information marqués par des liens sémantiques, qui permettent de passer de l'un à l'autre lorsque l'utilisateur les active. Le terme mathématique d'hyperespace désigne un espace à n dimensions. L'hypertexte figure un espace à n dimensions dans lequel le lecteur se déplace en sélectionnant les liens qui l'intéressent créant sa vision personnelle, son point de vue de l'hypertexte. Les contraintes sémantiques de l'accès aux documents et aux connaissances ne sont pas levées mais l'informatique apporte rapidité, facilité, interactivité, recherche d'information automatisée...

Le principe de base d'un hypertexte réside dans la conception des liens entre les éléments d'information. À son degré de complexité le plus simple, l'hypertexte ne constitue rien d'autre qu'un tourne-page. On distingue en effet plusieurs types de structures d'hypertexte. Linéaire, la structure d'un hypertexte néglige justement l'intérêt de l'hypertexte. Les deux autres types de structure sont la structure hiérarchique (arborescente) et la structure en toile d'araignée. La tendance se dirige vers des systèmes d'information plus intelligents, l'essentiel de l'innovation se situant dans les types de liens. On peut en distinguer trois classes, les deux dernières étant le moteur du progrès en association avec la technique des pages dynamiques [19] :

- les liens explicites : posés manuellement, ancrés dans un lieu précis d'un noeud (un mot, un paragraphe, une image) ;
- les liens calculés sont posés par l'activation d'une ou plusieurs propriétés associées au noeud source et au noeud cible ;
- les liens exécutables : programmes lancés lors de l'activation du lien.

Le Web peut être conçu comme un hypertexte distribué sur des machines géographiquement distantes ou comme un ensemble d'hypertextes locaux entre lesquels existent des liens. D'un point de vue technique, le Web est constitué de l'ensemble des serveurs HTTP [20] à accès ouvert de la planète qui mettent à disposition des pages HTML [21] lisibles avec un navigateur comportant le plus souvent un ou plusieurs liens qui peuvent pointer vers un endroit de la même page (local interne), vers une autre page du même site (local externe), vers une page d'un autre site (externe). Ce sont sans aucun doute l'interface graphique et la possibilité d'accéder rapidement à des informations liées en un

[17] ***"a body of written or pictorial material interconnected in a complex way that it could not be conveniently represented on paper. It may contain summaries or maps of its contents and their interrelations; it may contain annotations, additions and footnotes from scholars who have examined it.". Nelson, Ted. A File Structure for the Complex, The Changing and The Indeterminate, ACM 20th National Conference, 1965.***

[18] Cf. Laufer & Scavetta (1992) et Bruillard (1997).

[19] Dont le contenu est calculé en fonction de paramètres utilisateurs.

[20] HyperText Transfer Protocol

[21] Hypertext Markup Language

en vertu de la loi du droit d'auteur.

seul clic qui ont fait le succès de ce service.

1.1.2. Transfert de fichier et contrôle distant

Les premiers services proposaient des systèmes de commande en ligne tels ceux du DOS [22] qui fait frémir les utilisateurs néophytes. Le FTP (*File Transfer Protocol*) est parmi les premiers outils à avoir été mis à la disposition des utilisateurs, il permet d'échanger des fichiers de tous types (programmes, fichiers de données – son, image, texte…–.) sur Internet. Comme pour le Web, un serveur FTP est un logiciel qui fonctionne sur un ordinateur relié à Internet et rend accessible à la communauté (connexions clients) tout ou partie du contenu de volumes. L'accès peut être restreint à des utilisateurs particuliers ou ouvert à tous.

Telnet, lui aussi, est une application de première génération. Il utilise un protocole permettant de se connecter à une machine distante en tant qu'utilisateur. En France certains services de bases de données qui étaient accessibles par Minitel, le sont aujourd'hui via Telnet (ex. le catalogue de la bibliothèque municipale de la Part-Dieu).

1.1.3. Courrier électronique

Le courrier électronique est l'outil le plus répandu. Il permet d'acheminer des messages. Son avantage sur le téléphone ou sur le fax est qu'il permet de transmettre des informations numériques, donc réutilisables et modifiables, à un correspondant, et qu'il est théoriquement possible de consulter son courrier électronique depuis n'importe quel ordinateur relié à Internet. Le courrier électronique permet des échanges asynchrones entre un locuteur émetteur et *n* locuteur(s) destinataire(s). Des échanges privés entre personnes qui se connaissent de près ou de loin, l'utilisation du courrier électronique a été étendue aux échanges entre des individus ayant un centre d'intérêt commun qu'il soit d'ordre professionnel, politique, culturel ou ludique. Nous référons bien sûr aux listes de diffusions. Ces dernières fonctionnent grâce à des robots logiciels qui gèrent des listes d'adresses électroniques ainsi que la réception et la diffusion de messages. Le principe des listes de diffusion est souvent confondu avec celui des infogroupes sur Usenet.

1.1.4. Groupes de discussion

Usenet est né de la volonté d'ingénieurs systèmes Unix et de quelques chercheurs, de communiquer entre eux. Aux débuts d'ARPANET, les premiers échanges de courriers électroniques laissaient présager de l'intérêt que pourraient avoir certains courriers pour d'autres lecteurs que les correspondants originels. Ce réseau coopératif a connu un succès important auprès de très nombreux utilisateurs non informaticiens renforçant la philosophie d'usage non commercial de ce réseau (Usenet a inventé le principe de la netiquette [23]). La branche française, FNET, s'est constituée de manière informelle au début des années 1980, notamment grâce au CNAM et à l'INRIA ; l'Université Paris 8 via

[22] Disk Operating System

son Département Informatique, fut l'un des premiers sites à s'y connecter.

Les *Newsgroups* (infogroupes) sont des forums de discussion fédérés par thème. Il existe de nombreux forums sur les thèmes les plus divers, dans de nombreuses langues. Tout internaute peut, s'il le désire, consulter les *news* (informations, articles, questions...) disponibles sur un forum et/ou en ajouter. Chacun peut donc mettre à disposition des autres ses connaissances, opinions, etc. ; profiter des connaissances d'autrui ; demander des informations, lancer un débat.

Le principe est en effet d'envoyer un message électronique à un serveur central, qui le met à la disposition de tous. Pendant une durée de temps donnée, tous les courriers reçus par le serveur sont conservés et restent à disposition, ce qui laisse le temps aux internautes de répondre à une intervention. Les messages ne sont pas envoyés à tous les utilisateurs. Ils sont consultés par ceux qui sont intéressés par leur sujet.

De la même manière qu'un courrier électronique, les *news* contiennent du texte auquel on peut greffer des documents tels que des fichiers sons, des images, des programmes... Ces fichiers sont codés et sont placés à l'intérieur même de la nouvelle.

La structuration de l'information dans les infogroupes ne se fait pas seulement par le thème des interventions. Une intervention comprend en effet un titre (*title*), et un corps (*body*). Les réponses à une question ou à un article reprennent le titre de l'intervention à laquelle elles réfèrent précédé des trois caractères "Re:" (pour réponse) et y sont donc subordonnées. Un ensemble Article et Réponse est appelé un *thread*.

Les questions générales souvent répétées sur un sujet donné sont concentrées dans des fichiers FAQ (*Frequently Asked Questions*, Foire aux Questions) qu'on retrouve une fois par mois sur le forum concerné.

En France, on peut accéder à 6 000 groupes de nouvelles environ dont une faible proportion est de langue française. Il est possible de se connecter à des groupes de *news* publics américains qui permettent d'accéder à un plus grand nombre d'infogroupes, mais à une vitesse réduite.

1.1.2. Salles de discussion

Enfin IRC (*Internet Relay Chat* – 1988) et *talk* sont des protocoles qui permettent à des utilisateurs de communiquer de façon synchrone. Talk permet de communiquer à deux, tandis qu'IRC permet de discuter à plusieurs dans des canaux publics ou privés. Nous reviendrons plus spécifiquement sur ce service au chapitre suivant puisque c'est sur celui-là que nous avons choisi de porter notre étude.

Les frontières entre ces différents services s'amenuisent aux yeux des utilisateurs. Dans les premiers temps, on utilisait un logiciel différent selon ce qu'on entendait faire : à chaque service son logiciel. Mais les logiciels qui étaient voués à la consultation d'hypertextes sur Internet intègrent maintenant les modules nécessaires à la mise en œuvre d'autres protocoles : *news*, courrier électronique, transfert de fichiers… D'autre

[23] Ensemble de règles de bonne conduite sur les réseaux.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

part, les serveurs http (hypertext transfer protocol) se sont développés et intègrent le courrier électronique, l'interprétation de requêtes sur des bases de données, la discussion synchrone…

Il faut noter que ce panorama ne fait état que d'une partie des utilisations d'Internet. Il transite sur le réseau quantités d'informations propriétaires qui utilisent des protocoles particuliers dont nous ne traitons pas ici.

1.2. Les langues et Internet

Si Internet permet la communication humaine, il existe un enjeu pour les langues à travers Internet. Nous essaierons de comprendre lequel et tenterons de synthétiser l'information disponible quant au paysage linguistique de ce réseau. Cette présentation a pour objectif de placer notre objet d'étude dans son contexte proche.

1.2.1. Société de l'information

Mac Luhan [24] et le concept de village global, Alvin Toffler [25] et celui de troisième vague avaient dessiné les silhouettes de la situation actuelle. L'information prend valeur d'échange et devient matière première à traiter. Si les réseaux permettent de transporter toutes données numériques, celles-ci n'ont de valeur que si elles sont interprétables. Il ne nous appartient pas de formuler une opinion politique à ce sujet, notre intérêt ne se porte que sur l'étude de pratiques langagières affectées par la circulation sans frontières de l'information et les dispositifs qui la permettent. Le réseau n'a pas changé fondamentalement le moyen de la communication humaine : les langues. On estime en effet à 80% la part d'information électronique de type textuel [26] .

On dénombre environ six milles langues sur la planète et un nombre important de variantes de ces systèmes linguistiques. L'inquiétude que fait naître le concept de société de l'information au plan linguistique est de voir disparaître la richesse que représente la diversité des langues et dialectes au profit de quelques langues voire au profit d'une seule dans les visions les plus pessimistes [27] . L'alternative à cette situation serait de favoriser le plurilinguisme des individus d'une part, et d'autre part de développer la traduction

[24] Mac Luhan (1968)

[25] Toffler (1980)

[26] Kane, O. (2000) "Technologie push : mode de diffusion de l'information", *Euromaster TIC,* juin 2000. http://www.mediaroanne.asso.fr/tic_05/kane/publications/pushmedia/outilstraitementinformation.htm

[27] Cf. entre autres Grenoble & Whaley (1998) ; Hagège (2000) ; Chaudenson (2000) ; Calvet (1987), (1999) et (2002) ; Nettle & Romaine (2000) et sur la toile "déclaration universelle des droits linguistiques" http://www.egt.ie/udhr/udlr-fr.html, "Unesco red book of endangered languages" http://www.tooyoo.l.u-tokyo.ac.jp/Redbook/index.html

en vertu de la loi du droit d'auteur.

systématique entre langues. Les industries de la langue ont sans doute encore de belles années devant elles malgré l'avertissement ressenti à la parution du rapport accablant de ALPAC [28] au sujet de la traduction automatique il y a quelques décennies. La situation aujourd'hui est que les dix langues les plus parlées dans le monde sont les dix langues les plus parlées sur Internet, l'importance de chacune de ces langues n'étant toutefois pas identique dans le monde et sur Internet. Le tableau (1-1) présente le classement des 14 langues les plus parlées dans le monde. Nous verrons à la section suivante ce qu'il en est sur Internet.

Tableau (1-1) – Les langues les plus parlées dans le monde d'après *Ethnologue* (1996) (codes selon le standard ISO 639-2)

1	chinois, mandarin [CHI]	885,000,000
2	espagnol [SPA]	332,000,000
3	anglais [ENG]	322,000,000
4	bengali [BEN]	189,000,000
5	hindi [HIN]	182,000,000
6	portugais [POR]	170,000,000
7	russe [RUS]	170,000,000
8	japonais [JPN]	125,000,000
9	allemand, standard [GER]	98,000,000
10	chinois, wu [CHI]	77,175,000
11	javanais [JAV]	75,500,800
12	coréen [KOR]	75,000,000
13	français [FRE]	72,000,000
14	vietnamien [VIE]	67,662,000

On relèvera aussi les préoccupations de la communauté linguistique francophone et des démiurges du dicible [29] au sujet des usages particuliers du français et notamment de son code écrit au travers des nouvelles technologies de communication.

La langue qui prédomine sur Internet est bien entendu l'anglais et on ne saurait s'en étonner à l'examen des origines et du développement d'Internet [30] . Toutefois des études récentes montrent une progression des autres langues du monde au détriment de

[28] ALPAC (Automatic Language Processing Advisory Council) – le rapport ALPAC de 1966 qui met en garde contre les illusions de la traduction automatique, aura pour conséquence une forte diminution de l'investissement de recherche développement pour la traduction automatique - John R. Pierce and John B. Carroll. (1966) Language and Machines, Computers in Translation and Linguistics, ALPAC report 1966, National Academy of Sciences. Voir aussi Yehoshua Bar-Hillel (1953) Some linguistic problems connected with machine translation Philos. Sci. 20, 217-223. Yehoshua Bar-Hillel (1960) The present status of automatic translation of languages. Franz L. Alt (ed.) *Advances in Computers Volume 1*, Academic Press (New York), pp. 91-163 ; et John Hutchins (1996) ALPAC: The (In)famous Report, *MT News International* 14 (June, 1996), 9-12. Ce rapport eut des conséquences sur les investissements financiers dans ce secteur de recherche pendant plusieurs années.

[29] Formule empruntée à Hagège (1985:252)

[30] Cf. section 1 de ce chapitre

en vertu de la loi du droit d'auteur.

l'anglais.

1.2.2. Représentation des langues sur Internet

La mesure du paysage linguistique d'Internet se heurte à de nombreux problèmes. Le domaine le moins difficile à contrôler est sans doute celui de la répartition des internautes par nationalité. Ainsi, le baromètre publié par Médiamétrie le 12 février 2001 présente une progression du nombre d'internautes français de 47% en 2000; qui suit l'augmentation du parc informatique français.

Dès lors qu'on cherche à compter les locuteurs d'une langue, la question devient plus complexe puisque les différents comptages ne mentionnent pas toujours ce qu'ils comptent : locuteurs natifs ou locuteurs de langue seconde et ce sans compter que la situation de monolinguisme est loin de représenter une norme à travers le monde.

Pour le français, le tableau 1-2 est celui qu'il est possible de dresser d'après diverses études rassemblées par Global Reach© et qui se veulent des estimations de la population de langue maternelle française ayant accès à Internet.

Tableau (1-2) – Population de langue maternelle française ayant accès à Internet – Global Reach ©

France	11,9 Millions	"Baromètres Multimédia" Médiamétrie, Mars 2001. http://www.mediametrie.fr/web/index.html
Canada	2,3 Millions	The Daily Statistics Canada, Mars 2001. http://www.statcan.ca/Daily/English/010326/d010326a.htm
Suisse francophone	0,74 Millions	Nielsen NetRatings marketing research, Mai 2001. http://209.249.142.22/press_releases/PDF/pr_010625.pdf
Belgique francophone	0,8 Millions	Belgian Internet Mapping and InSites, http://www.insites.be/site3/indexjs.htm
Luxembourg	95 000	SIC Gallup Media, Août 2000. http://www.ilres.com/
USA francophones	95 000	Global Reach, http://www.glreach.com/globstats/refs.php3

Évidemment à l'échelle planétaire, la somme paraît ridicule, mais représente 3,44% des internautes.

Il faut bien sûr ne jamais écarter le fait que la situation est loin d'être figée, les enquêtes s'efforcent de dresser un panorama de l'évolution de la situation d'Internet. Ainsi, la croissance du réseau et le développement des sites se feraient au détriment de l'anglais. L'enquête publiée par Global Reach [31] en Mars 2001 estime la proportion des internautes anglophones à 47,5 % ce qui représente une baisse importante par rapport aux années précédentes (en 1999, Global Reach estimait à 60,4% le nombre d'internautes anglophones).

[31] Global Reach http://www.glreach.com/

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

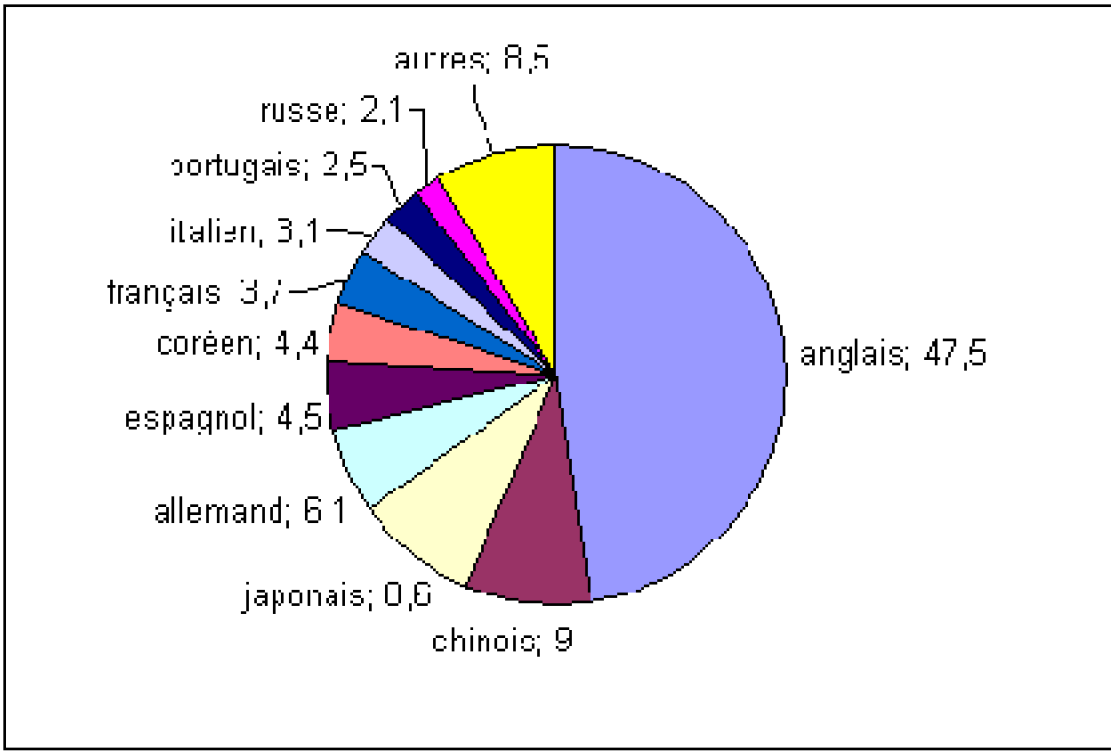

Figure (1-1) – Répartition des Internautes par langue (%) Enquête Global Reach, Online language populations – http://www.glreach.com/globstats/index.php3, Mars2001 (dernière consultation 01/2002)

Évaluer le nombre d'internautes locuteurs d'une langue ne revient pas à évaluer la place des langues sur Internet, la tâche est là plus difficile puisqu'il s'agit d'observer des pratiques et rien ne dit a priori quelle langue un locuteur utilise sur Internet et dans quelles conditions.

L'association Funredes [32] "Réseaux et développement" et l'Agence de la Francophonie produisent des enquêtes sur la présence des langues latines sur le réseau Internet. Ces études ne portent précisément que sur la toile d'araignée mondiale (et Usenet dans l'enquête de 1998) et ne permettent donc pas d'évaluer ce qui se passe du côté des services de communication interactive tels que le courrier électronique et les salles de discussion. La méthodologie de l'enquête peut être consultée sur le site de publication [33] . Nous en résumons toutefois les grands principes : l'étude se base sur les résultats de requêtes soumises à des moteurs de recherche sélectionnés sur des critères précis [34] . Une sélection de 57 termes par langue étudiée constitue le matériel de requête. Ces 57 termes sont choisis de sorte qu'ils aient "une signification et une portée équivalentes dans les langues étudiées" [35] . Les résultats publiés en juin 2001 montrent

[32] Association Réseaux et Développement http://www.funredes.org

[33] Langues et Cultures http://funredes.org/LC/francais/index.html

[34] Notamment sont sélectionnés les moteurs qui ne se servent pas des mêmes index, et qui fournissent un comptage des résultats de la recherche.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

que le français, parmi les langues romanes, apparaît en deuxième position après l'espagnol.

Tableau (1-3) – Proportion de pages Web pour les langues latines – Funredes – 2001

Langue	Proportion de pages Web
Anglais	52,00 %
Espagnol	5,69 %
Français	4,61 %
Italien	3,06 %
Portugais	2,81 %
Roumain	0,17 %
Allemand	6,97 %
Autres	24,96 %

Le tableau ci-après, tiré de la même enquête, est une tentative pour évaluer la productivité des locuteurs.

Tableau (1-4) – Productivité des locuteurs sur le Web – Funredes – 2001

	Pages	Internautes	P/I
ANGLAIS	52,00 %	47,6 %	1,09
ESPAGNOL	5,69 %	4,5 %	1,26
FRANÇAIS	4,61 %	3,7 %	1,25
ITALIEN	3,06 %	3,7 %	1,25
PORTUGAIS	2,81 %	2,5 %	1,12
ROUMAIN	0,17 %	0,13 %	1,31
ALLEMAND	6,97 %	6,1 %	1,14

Selon cette étude "le rapport entre pourcentage de pages et pourcentage d'utilisateurs se situe autour de 1 pour toutes les langues étudiées, ce qui montrerait qu'aujourd'hui la quantité de pages de la Toile produites dans une langue est directement proportionnelle au nombre d'internautes qui pratiquent cette langue. Le résultat de l'anglais surprend : on pouvait s'attendre à une valeur beaucoup plus forte sous l'influence du plurilinguisme [36] . Cela pourrait donc signifier que la productivité des locuteurs de langue anglaise est inférieure à celle des locuteurs dans les autres langues mentionnées, preuve d'une prime à la production des populations de locuteurs d'autres langues occidentales, conscientes des enjeux linguistiques de l'Internet ?" [37]

[35] http://www.funredes.org/LC/francais/L5/L5overview.html

[36] Pour des raisons de stratégie de communication, de nombreux sites d'entreprises et d'institutions proposent l'information qu'ils mettent à disposition dans plusieurs langues (langues officielles, plus anglais et autres éventuellement).

[37] http://funredes.org/LC/francais/L5/L5overview.html

en vertu de la loi du droit d'auteur.

Conclusion

Ce chapitre s'est présenté selon une double perspective en vue de dessiner le cadre des comportements linguistiques que nous souhaitons étudier. Il présente d'une part, d'un point de vue technique, les possibilités d'utilisation du réseau à des fins de communication, et d'autre part une revue des analyses de la situation linguistique sur le réseau Internet.

Il est bien évident que l'analyse de l'utilisation effective des possibilités de communication offertes par Internet, autant que l'analyse de la situation linguistique sont délicates à plusieurs égards.

Le premier point concerne le nombre d'utilisateurs sur Internet. Il est en effet difficile à estimer avec précision puisque derrière chaque adresse IP peut se cacher un ou un ensemble d'ordinateurs, et qu'un même ordinateur peut être utilisé par un nombre variable d'utilisateurs. Il faut d'autre part définir ce qu'on entend par utilisateur. Les critères peuvent en effet être la fréquence, la durée ou le nombre d'utilisations, et encore l'expertise de l'utilisation.

Des questions restent, de plus, au sujet de l'évaluation de l'utilisation relative des différentes possibilités techniques. Dire que le courrier électronique est le service le plus utilisé suppose qu'on se soit donné au préalable des paramètres de mesure et que ceux-ci puissent s'appliquer à tous les autres services. Il est en effet possible de se baser sur les durées d'utilisation, sur le nombre d'utilisations, sur les quantités d'informations véhiculées. Les enquêtes ne donnent évidemment pas les mêmes résultats selon le critère choisi pour la mesure.

D'autre part, sur Internet il existe un degré de stabilité de l'information lié aux différents services. Une information stable peut être plus probablement demandée ou livrée plusieurs fois à un même utilisateur, que les informations volatiles.

Enfin, dès lors qu'on souhaite étudier le comportement d'un groupe socioculturel sur Internet on se heurte à plusieurs paramètres incontrôlables. Un ordinateur peut avoir des utilisateurs multiples, que ce soit dans le cadre privé, professionnel ou institutionnel. De plus, l'assouplissement des frontières et la plus forte mobilité des personnes pour cette décennie n'autorisent pas à considérer de façon homogène l'échange d'informations au plan géographique. On peut aussi supposer d'autre part une inclination plus certaine des personnes expatriées à l'utilisation des nouvelles technologies de communication.

en vertu de la loi du droit d'auteur.

Chapitre 2 Communication Interactive

Le terme de communication implique dans ses conceptions les plus simplistes un émetteur, de l'information et un récepteur et l'on pourrait naïvement représenter cette situation en posant de part et d'autre d'une flèche orientée représentant l'information et sa circulation, l'émetteur et le récepteur. Le terme en effet ne présuppose pas la réciprocité. Tout au plus peut-on arguer que pour qu'il y ait communication, l'information doit être interprétable et interprétée. C'est pourquoi l'adjectif *interactive* qualifie *communication* dans notre titre puisque nous nous intéresserons à présent aux échanges d'informations.

La communication sur Internet, nous l'avons vu, est de deux types principaux : communication entre les machines et communication humaine, la seconde étant relayée par la première. Les services exposés au chapitre précédent permettent d'échanger des données numériques entre ordinateurs selon des protocoles de communication entre machines. La communication entre des ordinateurs signifie que des signaux électroniques empruntent le réseau physique pour être interceptés ou reçus par d'autres machines selon des conventions de communication appelées protocoles, et qui sont définis en couches. Ce qui s'échange entre ordinateurs n'est pas directement interprétable par des humains et n'est pas forcément verbal (programmes, images..). Parmi les services exposés au chapitre précédent, on aura compris que ceux qui sont voués à la communication humaine (verbale ou d'informations directement interprétables par le système cognitif d'un humain) se répartissent en plusieurs classes. Certains services Internet en effet sont basés sur une communication unilatérale : un type de communication dans lequel la réciprocité n'est pas ou très peu assurée – c'est le cas par exemple de la toile d'araignée mondiale – tandis que d'autres sont basés sur l'interaction

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

c'est-à-dire, sur l'échange d'informations. C'est plus précisément ce second ensemble qui nous intéresse dans la section suivante et nous nous attacherons principalement aux outils synchrones (en temps réel).

Nous proposerons une description sommaire des différents dispositifs exploitables dans ce cadre en mettant en avant leurs propriétés spécifiques, ce qui permettra de proposer une typologie basée sur les dispositifs techniques impliqués. À la suite de cela, nous présenterons les choix auxquels nous avons procédés pour cette étude et une première description de notre objet. Cette description, si elle est technique, a l'avantage de donner à voir notre objet d'étude aux lecteurs non avertis de façon assez exhaustive, et permettra d'envisager une caractérisation de ses spécificités.

2.1. Communication interactive interpersonnelle

Parmi les services exposés à la section 1 du chapitre précédent, trois sont consacrés à la communication **interactive** interpersonnelle, ce sont le courrier électronique, le *chat*, et les infogroupes. Ils permettent selon des paramètres participatifs et temporels différents types d'échanges que nous présentons dans le tableau (2-1).

Tableau (2-1) – Services Internet et communication interactive interpersonnelle

	1 émetteur n récepteurs	1 émetteur n destinataires	Communication synchrone
courrier électronique	listes de diffusion	Oui	Non
infogroupes	forums	Non	Non
chat	canaux	Oui (salles privées)	Oui

On notera que ce tableau utilise en têtes de colonne les termes de récepteur et de destinataire pour la raison que la différence entre les deux premières colonnes réside aussi dans la possibilité ou non d'adresser un message seulement à un destinataire identifié et sélectionné par l'émetteur du message.

Tous ces services, à l'origine associés à des protocoles spécifiques requérant l'utilisation de logiciels spécifiques, peuvent recevoir à l'heure actuelle une interface Web. Ainsi, la communication interpersonnelle synchrone sur Internet se joue principalement sur le Web dans les *webchats*, et sur l'IRC. Avant de constituer notre corpus définitif, nous avons observé différents types de *chats* et différents canaux. Le but de ces premières investigations était de sélectionner les lieux de communication présentant un intérêt pour l'analyse linguistique, permettant de constituer un corpus sans enfreindre d'interdictions et selon des modalités informatiques simples. Cette première observation a permis aussi un début de caractérisation de ce type de communication.

2.1.1. Types de communication synchrone textuelle en réseau.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Il existe différents moyens de communiquer sur Internet de façon synchrone. Ainsi, par exemple, la communication orale est-elle possible avec la vidéoconférence, et la téléphonie sur IP. À l'heure actuelle se développent aussi des *chats* audio. Nous ne considérerons pour cette étude que les moyens de communication en réseau qui impliquent le code écrit. Nous décrivons ici les différents types identifiables en fonction de leurs caractéristiques spécifiques.

Les paramètres à prendre en considération pour la caractérisation des différents types de moyens de communication synchrones basés sur le texte écrit sur le net sont le nombre de participants possibles dans l'interaction et le statut public ou privé de l'échange de messages. À un niveau technique, il convient aussi de préciser le moyen logiciel permettant de mettre en œuvre cette communication. Enfin, on distinguera les *chats* en mode texte et les *chats* en mode graphique.

Ces propriétés sont rassemblées dans le tableau (2-2) et permettent de faire ressortir quatre types principaux.

Tableau (2-2) – Types de moyens de communication synchrones basés sur le texte écrit

	nombre de participants possibles	statut de l'échange de messages	moyen logiciel	mode
talk	2	privé	logiciel dédié	texte
webchat	2-x	public ou privé	navigateur	texte et graphique rudimentaire
IRC	2-x	public ou privé	logiciel dédié	texte
Palace®	2-x	public	logiciel dédié	texte et graphique

Ces quatre types appellent quelques commentaires. Notons tout d'abord que parmi les dénominations figurant dans la première colonne du tableau précédent, apparaît un nom déposé. Il s'agit d'un système que nous avons choisi à titre d'exemple pour illustrer un type particulier.

En ce qui concerne la typologie ainsi dressée, la première distinction concerne le statut de l'échange de messages.

Le principe des logiciels de type *talk* est de permettre la communication privée. Ils nécessitent l'installation d'un démon, petit programme résident, chargé de détecter la connexion à Internet des interlocuteurs et les demandes de conversation.

D'un autre côté, les dispositifs de communication en réseau qui permettent l'interaction publique de plus de deux participants – *webchats*, IRC et Palace® – sont en général structurés en canaux. Ces derniers sont, comme les infogroupes, voués à structurer les communications selon leur thème, selon l'identité socioculturelle des participants, ou selon les buts extralinguistiques de la communication.

Les *webchats* sont très répandus. Ils se présentent comme des sites Web, il suffit d'un navigateur pour y accéder, ce qui fait leur succès. Les utilisateurs ont tous le même

en vertu de la loi du droit d'auteur.

statut et peuvent utiliser de petits éléments graphiques en place des frimousses [38] .

Palace® [39] est un système créé en novembre 1994 qui permet de dialoguer en temps réel dans des environnements graphiques. Il a été conçu à l'origine pour être un espace de discussion et de jeu multi-utilisateurs intégré au réseau FSN (Full Service Network), dans le cadre des premiers essais de télévision interactive de Time Warner Interactive® à Orlando en Floride. Un site Palace® comprend un nombre variable d'écrans, appelés pièces (*rooms*), à travers lesquels les utilisateurs peuvent se déplacer comme dans un hypertexte.

Illustration (2-1) – Interface du système Palace®.

[38] *Frimousse* est le terme retenu par la DGLF (Voir note 39 et chapitre 7 section 7.3.1.) pour équivalent de l'anglais *smiley*. On se reportera au chapitre 5, section 5.2.1 pour une présentation de ces unités.

[39] Les illustrations de Palace® proviennent du fichier de documentation fourni avec le client Palace® : Palace® User Software Guide for Windows 95/98/NT Version 3.5 © 1999 Communities.com.

en vertu de la loi du droit d'auteur.

Les utilisateurs sont représentés par des icônes qu'ils composent avec un répertoire d'éléments d'apparence (cheveux, lunettes, vêtements chaussure) et qu'ils peuvent modifier au cours de la connexion. Ils communiquent par bulles de texte.

Illustration (2-2) – Représentation des locuteurs dans le système Palace®

Enfin, IRC est le sigle pour *Internet Relay Chat*. Il s'agit d'un protocole permettant la communication synchrone entre plusieurs locuteurs via Internet. L'IRC est d'invention européenne : il fut écrit entre 1988 et 1992 par le finnois Jarkko Oikarinen [40] qui reçut pour son développement le prix d'excellence en télécommunications Dvorak Award en 1997. Les mots de Jarkko Oikarinen pour définir l'IRC sont :

L'Internet Relay Chat est un système de communication Internet multi utilisateur basé sur le texte, dans lequel les discussions sont organisées en canaux. La popularité d'IRC est basée sur le fait que le réseau IRC se compose de serveurs reliés les uns aux autres. Les utilisateurs peuvent alors se connecter à n'importe quel serveur et rejoindre des canaux de discussion de portée mondiale, envoyer des messages à des utilisateurs sur n'importe quel autre serveur. [41]

Après le mail et le Web, IRC serait le troisième protocole utilisé sur le Net. La communication IRC est basée sur un modèle client-serveur. Un internaute fait tourner sur sa machine un logiciel client qui se connecte à un serveur relié à d'autres serveurs répartis géographiquement. Les liens entre les serveurs forment un réseau IRC qui fait circuler les messages des utilisateurs.

Le terme consacré par la DGLF [42] pour nommer les différents espaces virtuels publics que nous venons d'exposer est *causette*. Nous reproduisons l'article paru au journal officiel pour définir ce mot :

causette n. f. fam. Domaine : Informatique/Internet Définition : Communication informelle entre plusieurs personnes sur l'Internet, par échange de messages affichés sur leurs écrans. Voir aussi : internet. Équivalent étranger :chat. Article

[40] http://www.kumpu.org/jto/

[41] ***"Internet relay Chat is text-based Internet-wide multi-user communication system, where discussions are organized into channels. The popularity of IRC is based on the fact that the IRC networks consist of servers linked to each other. The users then may connect to any server and join world-wide discussion channels send messages to users on any other server." Jarkko Oikarinen. http://www.kumpu.org/jto/research/index.html***

[42] Délégation Générale à la Langue Française

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

DGLF – http://www.culture.gouv.fr/culture/dglf/Soc-inf/soc-inf-accueil.html

Les causettes de langue française sont en petit nombre à côté de celles qui existent pour l'anglais. Les internautes francophones utilisent pour les désigner préférentiellement les termes anglais (*chan, chat*), des traductions de ces termes (*canal*), ou des adaptations orthophonographiques (*tchatches*), les internautes québécois restant les plus inventifs (*bavardoirs*).

2.1.2. Choix pour cette étude

Notre choix s'est porté sur l'IRC plutôt que les *webchats* [43] pour trois raisons essentielles : les *webchats*, à première vue, comportent une grande part d'interventions du type petite annonce et des joutes d'insultes. D'autre part, ils sont beaucoup plus denses en gadgets divers : graphiques, sonores, ainsi que le montre – pour le côté graphique – la copie d'écran suivante :

[43] *Webtchatches* selon la terminologie de ANIS (1998).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

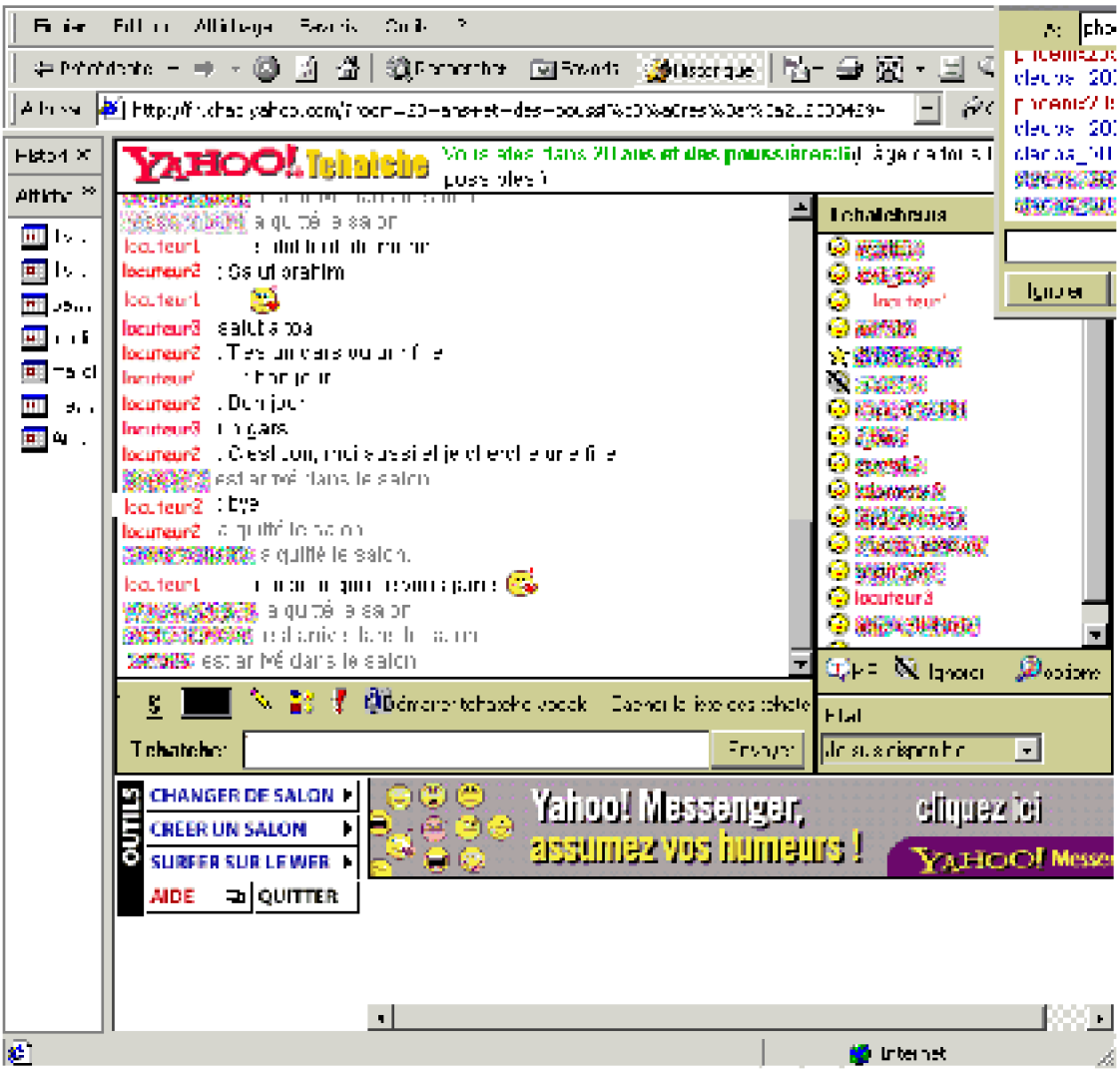

Illustration (2-3) – Interface d'une causette Web.

La communication linguistique semblait ainsi se trouver reléguée à un plan moindre, l'activité se limitant souvent pour les internautes à tester ces gadgets [44] . Il est envisageable que la communauté des Internautes ne se soit pas encore appropriée ces nouvelles techniques et qu'il faille attendre avant qu'elles ne gagnent la maturité que semble davantage manifester la communication IRC. Il nous a donc semblé que, pour un premier abord de ces modes particuliers de communication, il était raisonnable de s'en tenir aux plus simples et stables, non pas que les insultes rituelles [45] ou les processus

[44] L'objet serait intéressant à étudier sur une plus longue période, pour observer les phénomènes de conventionnalisation et autres évolutions.

[45] Au sujet des insultes rituelles, Labov (1978:223-288) a montré qu'il s'agit, dans le vernaculaire noir américain, d'un domaine qui donne "amplement l'occasion de mettre à jour la richesse verbale, la maîtrise syntaxique et la créativité dont savent faire preuve les membres des groupes de pair" – Labov (1978:223).

en vertu de la loi du droit d'auteur.

d'appropriation de l'outil soient inintéressantes mais il faut encore que des bases soient établies.

Nos corpus ont donc été constitués sur les IRC. Nous avons sélectionné deux canaux francophones parmi les quelques-uns qui se présentaient, en fonction de leur fréquentation. Il existe en effet des canaux plus "intimistes" que d'autres, dans lesquels peu de participants interviennent et qui connaissent une activité assez faible. Les canaux que nous avons sélectionnés sont parmi les plus prolixes. On trouvera au chapitre 3 la description précise des corpus recueillis en terme d'activité. Dans un premier temps, nous présentons le dispositif de communication dont nos corpus sont issus, ce qui permettra au lecteur d'appréhender de façon globale le poids du dispositif et comment se joue l'interaction dans les IRC.

2.2. Déroulement technique d'une session

Le titre de cette section invite peu à la lecture. Nous essaierons de présenter ci-après les principes de la communication sur IRC ainsi que le logiciel utilisé. Cette description permettra au lecteur de se représenter précisément l'activité à laquelle nous référerons tout au long de cette étude. Nous avons choisi de présenter de façon exhaustive les grands temps du déroulement de l'activité, en s'arrêtant toutefois aux points 3 et 4 sur des propriétés importantes d'éléments préalablement évoqués (propriétés des canaux et statuts des participants).

2.2.1. Script général

Pour communiquer sur IRC, il faut évidemment un ordinateur connecté à Internet, et un logiciel dédié. Selon la plateforme (Macintosh™, Compatible IBM™…) et le type de système d'exploitation installé (MacOs™, Linux™, Windows™, MS-Dos™, JavaOS™, Unix™…) ce dernier sera différent. On notera qu'il existe de nombreux logiciels clients IRC mais que les plus connus, et donc sans doute les plus utilisés, sont en petit nombre. Ainsi, sur Compatible IBM™ – plateforme la plus répandue à travers le monde –, les vedettes sont MIRC™, et VIRC™. Sur Macintosh™, Ircle™ est l'équivalent. Ces logiciels ne sont pas sont pas tous libres d'utilisation mais leur coût reste modique et il est souvent possible de les tester durant une période de 30 jours sans frais. Leur interface graphique les rend assez simples à utiliser. De plus, ils permettent de créer des scripts afin d'automatiser les commandes courantes. Ces logiciels présentent une grande similarité, tant du point de vue de l'interface que des fonctionnalités, si bien qu'on peut en donner une schématisation :

en vertu de la loi du droit d'auteur.

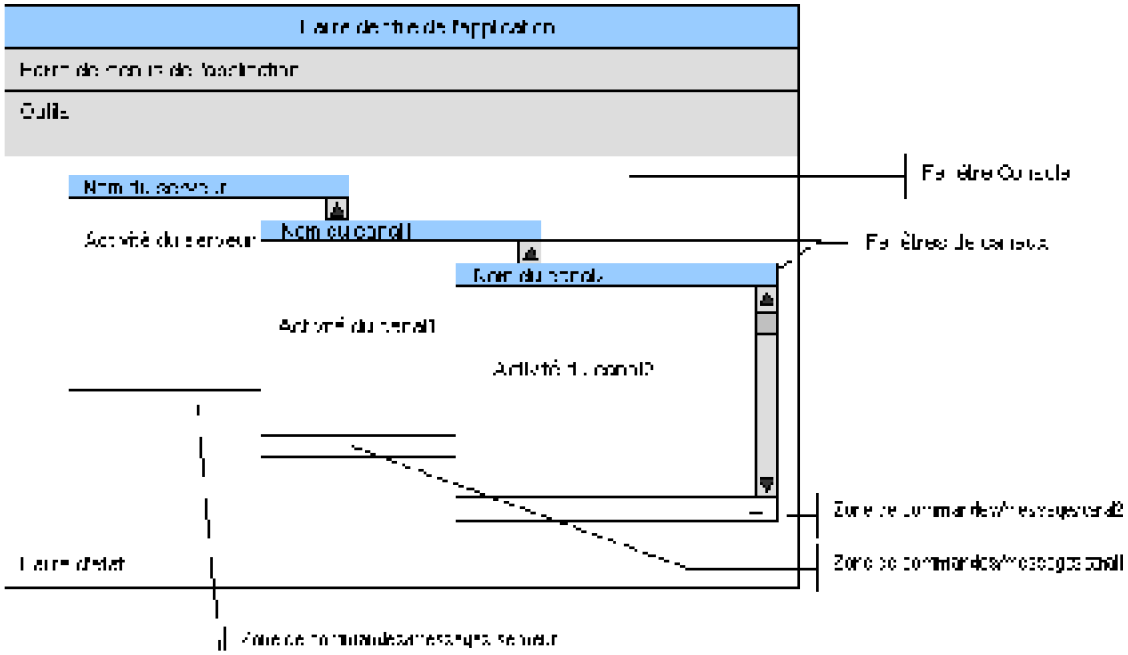

Illustration (2-4) – Interface des clients IRC – état connecté à un serveur, deux canaux rejoints.

Une fois le logiciel lancé, l'utilisateur s'étant muni préalablement d'un surnom (pseudonyme) se connecte à un serveur, lequel lui permet de se connecter à une ou plusieurs salles d'IRC et d'intervenir dans ces salles par le biais de commandes et de messages. Étant connecté à un serveur, l'utilisateur peut quitter ou rejoindre des salles à tout moment. C'est la déconnexion du serveur qui termine l'activité.

Pour une session IRC, au moins deux fenêtres dans lesquelles se produisent des événements en permanence sont présentes sur l'écran : la fenêtre console, qui présente les événements qui se produisent sur le serveur, et au moins une fenêtre de canal dans laquelle se déroule l'activité. À l'affichage de la fenêtre d'un canal est lié l'affichage de fenêtres de canaux privés que les utilisateurs connectés à ce canal créent. Deux fenêtres à gérer pour un utilisateur est donc véritablement un minimum lors d'une session IRC (à moins que l'utilisateur ne se connecte à aucune salle, ce qui limite cependant l'intérêt de la chose). De plus, on sait que certains utilisateurs sont connectés sur plusieurs canaux simultanément. Ces derniers gèrent donc un nombre important de fenêtres lors d'une session IRC.

2.2.2. Connexion

Un utilisateur, pour entrer dans un canal, doit préalablement se connecter à un serveur qui donne accès à un réseau d'IRC par l'intermédiaire du logiciel dédié mentionné ci-dessus. Il existe plusieurs réseaux IRC indépendants (EFnet, UnderNet, DALnet…). Il s'agit pour l'utilisateur de choisir le réseau IRC auquel se connecter en fonction de la proximité géographique des serveurs de ce réseau par rapport au point de connexion de l'utilisateur. Nous donnons ci-après les opérations à effectuer pour se connecter à un canal, de sorte que le lecteur puisse se faire une représentation précise du déroulement

d'une session.

Avant de se connecter à un réseau, l'utilisateur doit remplir un formulaire, que nous reproduisons ci-après illustration (2-5), dans lequel figurent des informations qui permettront de l'identifier. L'information essentielle pour la suite des opérations étant le surnom (*nickname*) par lequel l'utilisateur souhaite se faire reconnaître. Il n'y a bien sûr aucun moyen de forcer un utilisateur à entrer des informations exactes concernant son identité. Les informations insérées dans ce formulaire seront communiquées aux utilisateurs connectés qui le demandent par le biais d'une commande. Nous ne détaillons pas ici l'ensemble des paramètres possibles de la connexion qui ne présentent pas grand intérêt pour notre propos. Pour plus d'information, on se reportera aux manuels d'utilisation des logiciels d'IRC, ou à leur documentation en ligne.

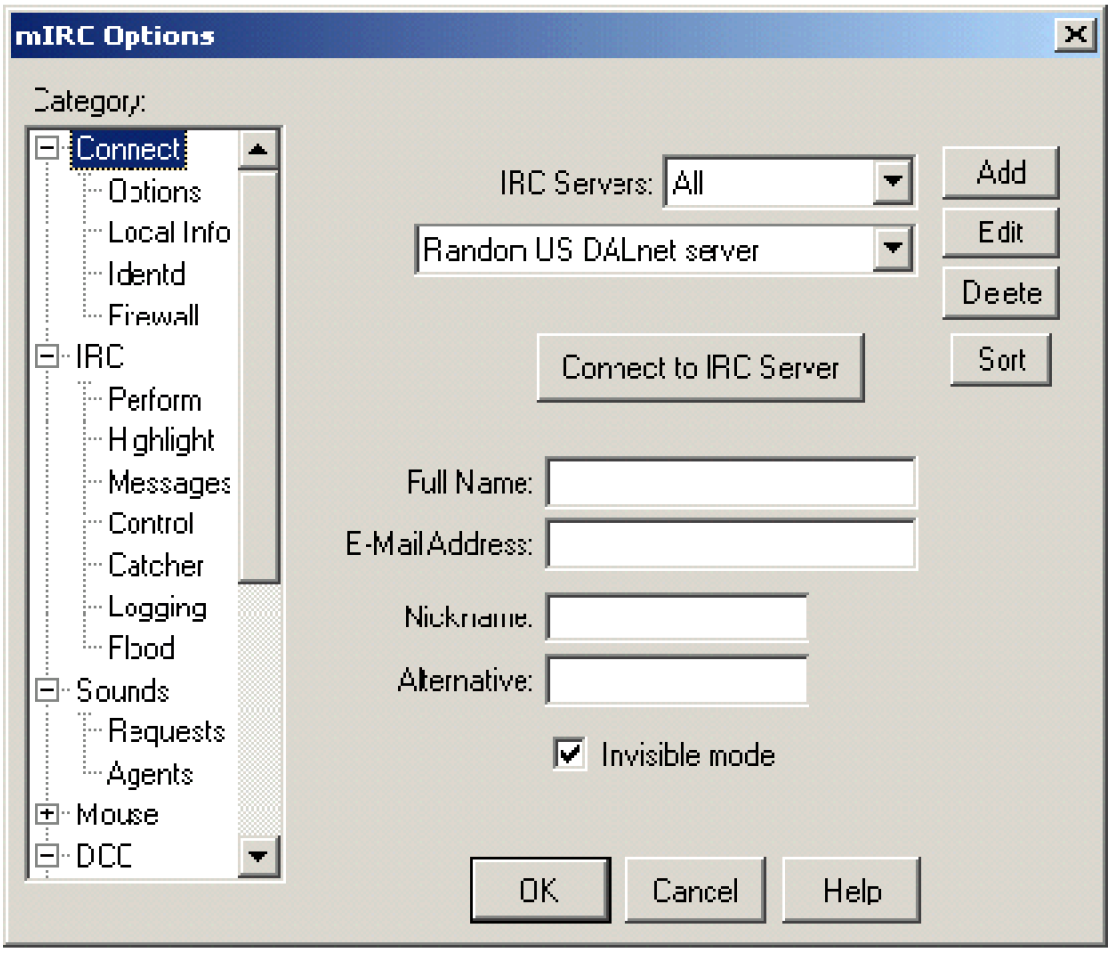

Illustration (2-5) – Formulaire de connexion IRC

Lorsque le contenu du formulaire a été validé en cliquant sur le bouton "ok", la connexion devient possible. Cette dernière est obtenue par un simple click sur l'icône mise en évidence dans la copie d'écran ci-dessous. La connexion engage des requêtes à un serveur IRC dont les réponses s'affichent dans une fenêtre "status" [46].

[46] Ou fenêtre console.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

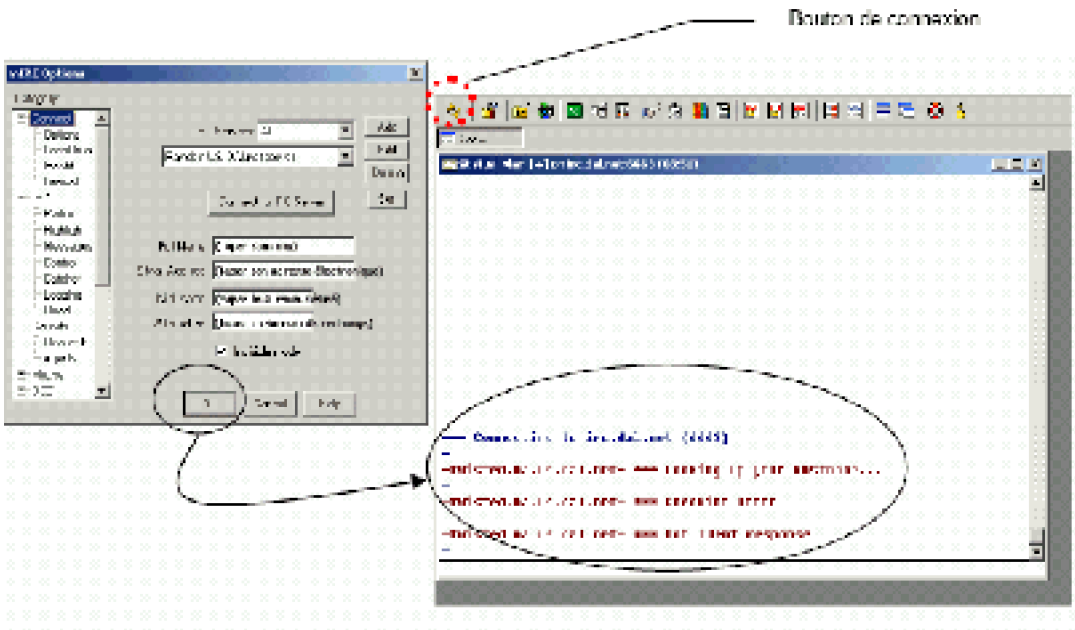

Illustration (2-6) – Connexion IRC

Au terme d'un ensemble d'échanges d'informations entre le client et le serveur et sauf conditions particulières (nombre maximum d'utilisateurs atteint, utilisateur interdit…) la connexion est établie. Une fois entré sur un réseau, l'utilisateur a accès à la liste des canaux disponibles (Illustration 2-7). Ces canaux peuvent présenter des thèmes de discussion comparables à ceux qu'on trouve dans les infogroupes. En principe, cette information se trouve dans ce qu'on appelle le *topic* du canal, qui apparaît dans la colonne de droite de la fenêtre liste des canaux.

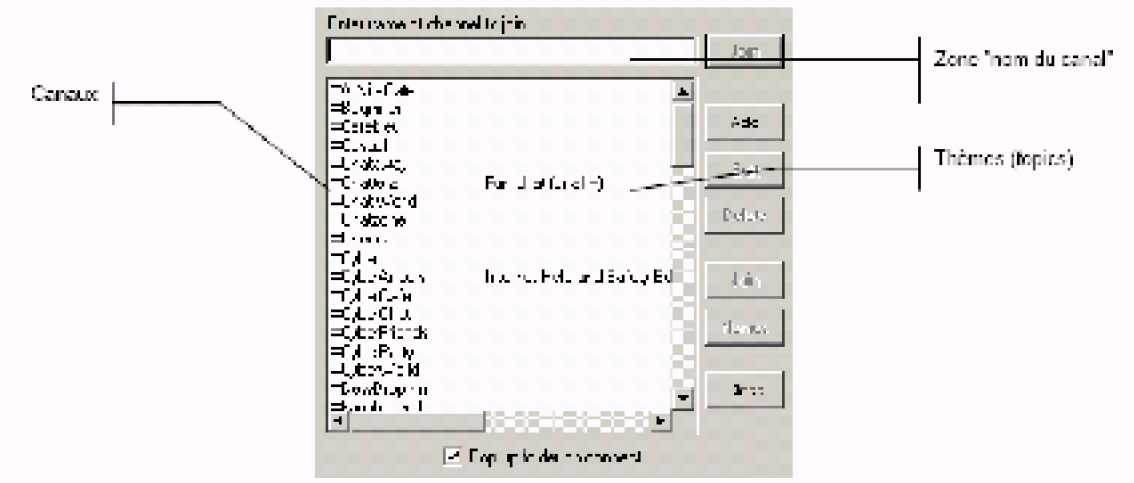

Illustration (2-7) – Liste des canaux

Après avoir sélectionné un canal dans la liste et cliqué sur le bouton "*join*", l'utilisateur est connecté au canal (il peut recommencer l'opération pour se connecter simultanément à d'autres canaux), sauf situations particulières (*nickname* déjà utilisé par un autre internaute ou réservé par un internaute, utilisateur interdit,…). La netiquette recommande

en vertu de la loi du droit d'auteur.

que l'arrivant salue les autres internautes connectés à la salle.

Nous arrêtons à présent la description du déroulement de l'activité pour définir plus précisément ce qu'est un canal et présenter ce qu'on doit savoir des autres utilisateurs.

2.2.3. Propriétés des canaux

Un canal IRC se définit par ses propriétés : nom du canal, thème de conversation, statut public ou privé, mode d'accès au canal, modération du canal, nombre de participants limité ou non. Il faut savoir que ces propriétés sont susceptibles d'être changées à tout moment par les opérateurs du canal, nous les décrivons succinctement ci-après

nom du canal	
thème de conversation	l'attribut "*topic*" d'un canal est une suite de caractères indiquant les buts communicatifs du canal. Il peut être modifié par les opérateurs à tout moment. Dans l'usage, les attributs *topics* sont utilisés pour communiquer des informations à propos de l'activité du canal à ses utilisateurs. Par exemple le *topic* peut inclure un rendez-vous dans le monde réel pour les utilisateurs.
statut	quatre statuts de canal existent. Par défaut un canal est public, c'est-à-dire ouvert à tous les utilisateurs. Un canal secret n'est pas visible pour les autres utilisateurs. Il existe des canaux privés dans lesquels personne ne peut entrer hormis une liste définie d'utilisateurs. Il existe enfin des canaux privés dans lesquels on ne peut entrer que sur invitation.
modération du canal	un canal peut être modéré et dans ce cas, seuls les utilisateurs ayant le statut d'opérateur ou de voice [47] peuvent converser.
nombre de participants	un canal peut être limité en nombre de participants. Dès que le quota est atteint aucune entrée n'est autorisée dans le canal, il faut alors attendre qu'une place se libère.

Il existe des canaux pérennes et des canaux temporaires. Tout utilisateur peut créer un canal mais ce dernier sera détruit automatiquement dès que plus personne n'y sera connecté. Les canaux pérennes sont habités par un robot logiciel qui les maintient ouverts.

La liste des canaux présentée Illustration (2-7) permet de rejoindre un canal existant, elle permet aussi de créer un canal. Si le nom tapé dans la zone "nom du canal" n'est pas un nom de canal existant, un canal sera créé dans lequel le créateur aura le statut d'opérateur et qui sera détruit automatiquement à sa désertion complète. Lorsqu'un utilisateur rejoint un canal, une fenêtre se crée dans laquelle apparaît l'activité du canal et la liste des utilisateurs connectés au canal. Il est possible de se connecter à plusieurs canaux simultanément et dans ce cas une fenêtre par canal est affichée. L'utilisateur a le loisir de gérer ses fenêtres de différentes façons : en mosaïque, en cascade, ou réduites

[47] Les statuts des utilisateurs sont présentés à la section 2.2.4

en vertu de la loi du droit d'auteur.

dans le bas de la fenêtre de l'application, les noms des canaux dans lesquels un message s'inscrit apparaissant en rouge dans le haut de la fenêtre.

Il est assez difficile de concevoir qu'on puisse suivre l'activité de plusieurs canaux simultanément hormis certains canaux ayant une activité très réduite.

2.2.4. Statuts des participants

Les utilisateurs qui se connectent à un canal ont des statuts particuliers. Un même utilisateur peut avoir un statut dans un canal et un autre dans un autre. Ces statuts sont visibles. En effet, lorsqu'un utilisateur arrive dans un canal la fenêtre se présente comme suit [48] :

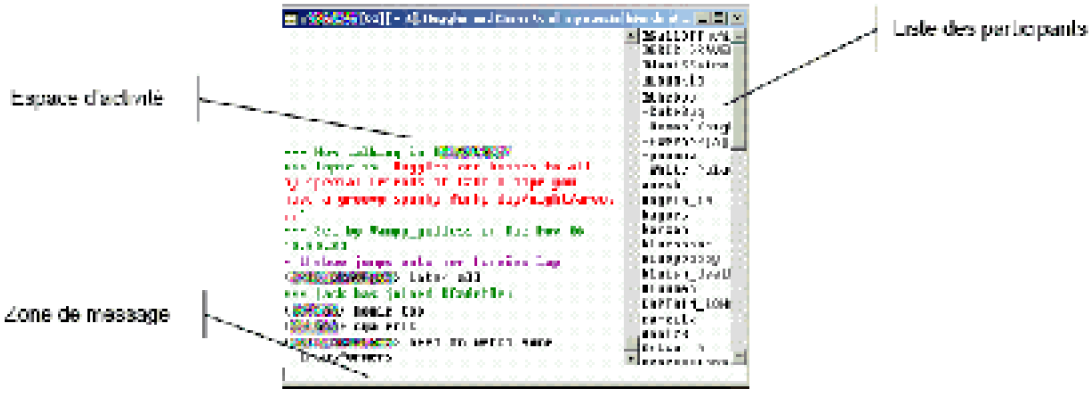

Illustration (2-8) – Statuts des participants

Nous passons en revue les différents statuts possibles des utilisateurs d'un canal.

Les simples utilisateurs se présentent dans la liste des utilisateurs du canal avec seulement leur pseudonyme (*nickname*) ;

Les *voices* peuvent lire les messages que les opérateurs s'envoient, ils peuvent aussi participer aux conversations quand le canal est modéré. Leur surnom apparaît dans la liste accompagné d'un signe "+"

Les opérateurs sont chargés de contrôler les agissements des utilisateurs du canal. Ils se reconnaissent par le caractère "@" qui figure devant leur pseudonyme. Trois sous catégories d'opérateurs existent : *sop, aop, op*. Ce sont respectivement les super-opérateurs, les auto-opérateurs, et les simples opérateurs. Un *aop* reçoit automatiquement le statut d'opérateur à son entrée dans le canal. Il a la possibilité d'exclure temporairement ou définitivement un autre utilisateur, et d'accorder le statut de *voice*. Un *sop* a tous les privilèges d'un *aop* plus celui de pouvoir accorder le statut d'opérateur à un utilisateur temporairement. Les simples opérateurs dépendent donc du bon vouloir des super-opérateurs.

- Les *owners* (ou *founders*) sont les créateurs de canaux. Ils ont tous pouvoirs sur ceux

[48] Les surnoms ont été anonymisés dans les dialogues, ainsi que le nom du canal.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

qu'ils ont créés.

- Les IRC-opérateurs sont chargés de surveiller les serveurs et d'assurer leur bon fonctionnement.
- Les administrateurs enfin, sont au sommet de la hiérarchie.

Ces trois dernières catégories d'utilisateurs ne se reconnaissent pas dans les canaux.

Ayant énuméré les différentes propriétés des canaux et exposé les statuts possibles des participants, nous abordons au point suivant l'étape essentielle : l'échange de messages.

2.2.5. Activité du canal

Un utilisateur connecté à un canal dispose donc d'informations sur le canal lui-même et sur ses participants. En ce qui concerne le canal lui-même, la barre de titre de la fenêtre du canal présente son nom et son *topic*, en ce qui concerne les participants, la zone de droite en présente la liste ainsi que le statut de chacun. L'utilisateur connecté peut toutefois formuler à tout moment de la connexion des requêtes pour connaître les (autres) propriétés d'un canal auquel il est connecté [49] par le biais de commandes textuelles ou de raccourcis (boutons de barre d'outils, commandes de menus…) prévus par le logiciel qu'il utilise. Il en va de même pour obtenir des informations sur les autres utilisateurs. Un utilisateur connecté peut par exemple obtenir la liste des utilisateurs d'un canal, des informations sur l'utilisateur d'un surnom (son nom réel, son adresse IP…), et ce, même si l'utilisateur a quitté le canal.

Ainsi il est également possible d'obtenir des informations sur le serveur. Tout utilisateur peut connaître la liste des serveurs qui y sont reliés et la liste des commandes disponibles sur le serveur.

Ces commandes sont pour la plupart transparentes [50] aux yeux des autres utilisateurs, elles n'ont pas de contrepartie dans la fenêtre d'activité du canal.

Les commandes que nous exposons ci-après, en revanche, ont une incidence sur le canal et sont visibles des autres utilisateurs, c'est ce que nous dénommerons ici "activité du canal".

On peut distinguer dans l'activité du canal les lignes qui apparaissent automatiquement, sans que les utilisateurs puissent les contrôler, des lignes qui apparaissent à la suite d'une intention précise d'un utilisateur. Il faut aussi distinguer les lignes générées par un programme des lignes produites par les utilisateurs eux-mêmes. Et enfin certaines lignes sur lesquelles l'utilisateur n'a pas de contrôle (apparaissant à la suite d'une commande effectuée) n'ont qu'une incidence locale.

[49] Pour la clarté de l'exposé, nous envisageons le fait où un utilisateur est connecté à un seul canal public, nous invitons cependant le lecteur à conserver en mémoire qu'il est possible d'être connecté à plusieurs canaux simultanément, et ce afin de considérer la charge cognitive que peut représenter ce type de communication.

[50] L'adjectif *transparent* dans la terminologie informatique signifie *invisible*. Il ne s'agit donc pas de la *transparence* du politique !

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Lorsqu'un utilisateur rejoint un canal, une ligne s'affiche mentionnant le surnom de l'utilisateur. Il en va de même lorsqu'un utilisateur quitte un canal ou se déconnecte du serveur.

Lorsqu'un utilisateur change de nom, la ligne d'événement associée mentionne l'ancien surnom et le nouveau.

Un utilisateur peut en ignorer un autre et dans ce cas, cet utilisateur ne reçoit plus aucun des messages envoyés par l'utilisateur ignoré.

Un utilisateur peut aussi signaler le fait qu'il s'absente un court instant. Dans le cas des départs, les utilisateurs peuvent afficher en plus du texte généré par le programme un message éventuellement préenregistré.

Nous dénommerons les lignes correspondant à ces commandes dans l'espace d'activité du canal des événements, afin de les distinguer des messages entièrement composés par les utilisateurs.

En ce qui concerne les messages envoyés par les utilisateurs connectés au canal, ils peuvent être publics ou privés. Les messages privés n'ont pas d'incidence dans la fenêtre d'activité du canal, ils ne sont visibles que de leur destinataire. Les opérateurs peuvent envoyer des messages qui ne seront visibles que des autres opérateurs ou *voices*. Ils peuvent aussi restreindre l'activité du canal en n'autorisant que les messages des utilisateurs ayant le statut d'opérateurs ou de *voices*.

En dehors de ces restrictions, les messages publics peuvent avoir deux formes : une forme normale et une forme action. Pour envoyer un message de forme normale, l'utilisateur tape le texte de son message dans la zone de message (voir Illustration 2-4). Ce dernier n'apparaît dans la fenêtre d'activité du canal que lorsque l'utilisateur aura signifié son envoi en appuyant sur la touche "entrée". Après validation, le message ne peut plus être modifié. Le message ainsi envoyé est inscrit dans la l'espace d'activité du canal précédé du surnom de son émetteur.

<surnom_du_participant> texte_du_message

Les messages de forme action apparaissent différemment des messages ordinaires dans la fenêtre d'activité du canal. [51]

** surnom_du_participant texte du message*

Ils sont en quelque sorte un moyen de mettre en relief certains messages. Ils sont composés par les utilisateurs dans la zone de message selon une syntaxe de commande rudimentaire. L'utilisateur tape "/me" à l'initiale de son message de sorte que le message reçoive l'affichage particulier pour les actions, "/me" étant remplacé dans la zone d'activité par le surnom de l'émetteur de la commande. Ce type de message impose donc que le surnom de l'émetteur apparaisse en initiale de message. Les conventions d'utilisation de ce type de message stipulent que l'émetteur doit composer un message à la troisième personne du singulier.

L'espace d'activité du canal est donc le centre où s'enregistrent les interventions

[51] On trouvera aux chapitres 3 et 4 les procédures de traitement des fichiers bruts.

en vertu de la loi du droit d'auteur.

publiques des utilisateurs connectés à un canal ainsi que les événements. Cet enregistrement est séquentiel : les messages et les événements s'inscrivent les uns à la suite des autres dans leur ordre d'arrivée sur le serveur.

Le dispositif porte donc en lui-même des caractéristiques qui ont des incidences sur le déroulement et la gestion de l'interaction.

2.3. Première caractérisation de la communication IRC

Dans un premier temps, nous tenterons de dégager les spécificités que crée le dispositif de communication par rapport aux autres dispositifs connus, ceux-ci nous permettent en première approximation de constituer une caractérisation de la communication linguistique propre aux IRC.

La communication IRC est synchrone, c'est-à-dire que les participants à l'échange linguistique sont engagés dans le même temps ainsi qu'ils le seraient dans une conversation en face à face, mais comme dans les conversations téléphoniques, les participants peuvent se trouver dans des espaces distants et dans des temporalités relatives différentes.

La communication est basée sur l'écrit : les participants utilisent des caractères. Si la conversation téléphonique ampute la situation de communication prototypique des caractéristiques visuelles et de l'espace partagé, dans les IRC, les participants ne peuvent pas non plus bénéficier des informations véhiculées par la prosodie : l'intonation, le rythme, l'accentuation, le débit, la mélodie...

Du point de vue du nombre de participants, les IRC se rapprochent de la conversation en face à face laquelle, contrairement à la conversation téléphonique, engage couramment plus de deux locuteurs. Les interventions des locuteurs n'étant repérables que par la forme écrite de leur surnom qui précède le message interventif en lui-même, et les interventions de divers locuteurs se succédant assez rapidement, la charge cognitive nécessaire pour suivre la conversation semble plus importante. D'autre part un locuteur peut être engagé dans des conversations différentes, qui peuvent se jouer dans le même canal ou dans plusieurs canaux. Le dispositif permet toutefois de consulter facilement la trace des interventions et des événements survenus sur un canal.

Dans les situations de communication courantes, on note la variation que fait surgir le fait que les locuteurs engagés dans un échange linguistique se connaissent ou non et ont des statuts différents. Dans les communications sur IRC, on rencontre plusieurs cas de figure simultanément. Les participants peuvent ne pas se connaître, se connaître virtuellement, se connaître réellement. Hormis ce dernier cas, on ne sait de l'autre que ce qu'il en dit et ce qu'il est dans l'espace virtuel. Ainsi que nous l'avons mentionné, le dispositif lui-même confère un statut aux participants. Des rapports de places peuvent en découler.

La communication linguistique synchrone orale impose une gestion particulière des tours de parole. Ici, les chevauchements, les interruptions revêtent une forme particulière

en vertu de la loi du droit d'auteur.

qui n'a rien de comparable à celle de la conversation en face à face ou au téléphone puisqu'un utilisateur, sauf problème technique ou intervention musclée d'un opérateur, a la pleine maîtrise de la longueur de son intervention. Cette dernière est toutefois contrainte par le fait que tant que le locuteur émetteur n'a pas signifié l'envoi du message, il n'est pas visible des autres utilisateurs. Pour être pertinentes les interventions réactives doivent être temporellement proches de leur source donc d'une longueur limitée. La stratégie employée pour pallier les délais imposés est alors parfois de découper les interventions.

Il découle de cette première caractérisation les points pour lesquels nous pouvons faire l'hypothèse qu'ils se manifesteront de façon spécifique et donneront à observer des stratégies particulières.

Ainsi en premier lieu, le code écrit, plus spécifiquement destiné à l'échange asynchrone, devrait manifester des spécificités. Les caractéristiques lexicales et syntaxiques du corpus devraient également se rapprocher de celles de la conversation en face à face. On devrait également noter des particularités quant aux repérages déictiques et aux formes d'adresses, qui devraient apparaître explicitement plus que dans tout autre type d'échange puisque les indices para- et extra- linguistiques ne sont pas accessibles directement aux interlocuteurs. Enfin, les stratégies de gestion des tours de parole et de construction de la conversation devraient différer de celles de la conversation en face à face.

Conclusion

Après avoir situé le dispositif de communication dont nos corpus sont issus par rapport aux autres types identifiés, nous avons précisément exposé ses caractéristiques, afin que le lecteur puisse se faire une idée claire de ce dont est issu l'objet recueilli.

Les chapitres suivants exposeront les techniques de récupération puis de traitement des corpus.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Chapitre 3 Description du corpus brut

Ainsi que nous le mentionnions au chapitre 2, notre corpus est constitué de plusieurs sessions provenant de deux canaux publics différents, ceci afin d'éviter l'idiosyncrasie qu'aurait présentée l'étude d'un seul canal. Chaque session fait l'objet d'un fichier dont le nom se compose d'une lettre et d'un chiffre, représentant respectivement le canal dont le fichier est issu (P ou F) et le numéro de la session d'enregistrement. Nous détenons d'autres corpus provenant d'autres canaux et d'autres types de communication en réseau synchrone qui ne sont actuellement que partiellement traités, et nous sont principalement utiles pour vérifier qu'il ne se présente pas de différences majeures entre eux. Lorsque nous y faisons référence, nous le mentionnons explicitement. Toutes ces données proviennent évidemment de lieux électroniques publics.

Ce chapitre présente dans un premier temps les principes de la constitution du corpus, puis ses caractéristiques générales. La troisième section contient les spécifications des fichiers collectés, qui permettent d'obtenir des informations quantitatives sur les données exposées à la section suivante. En outre, l'analyse du format des fichiers a permis les traitements automatiques nécessaires à l'exploitation linguistique des données.

3.1. Constitution des corpus

en vertu de la loi du droit d'auteur.

Le principe d'incertitude d'Heisenberg (1927) met en avant que l'observation d'un phénomène est susceptible en elle-même de modifier l'observé et pose au chercheur la question de son influence relativement à son objet d'étude [52] . En linguistique, ainsi, Françoise Gadet (1997:18) [53] rappelle que dans "les contacts inégaux entre un locuteur du dialecte standard (ce qu'est le chercheur) et un locuteur du dialecte non standard, il y a toujours le risque que le standard, socialement dominant, vienne désorganiser la production même des énoncés et en tout cas la réflexion sur le système non standard, dans des proportions difficiles à prédire." [54] Nous avons optimisé les conditions d'observation de sorte qu'elles aient une influence minimale sur l'objet. À cet effet, nous avons choisi des canaux dans lesquels il y avait un nombre de participants important et beaucoup de participants "inactifs". Être inactif dans la communication que nous souhaitions observer était une condition majeure et nécessaire. Pour pouvoir être inactif sans être gênant, il fallait aussi ne pas être une exception.

L'inactivité dans les canaux est un phénomène très répandu, on estime que plus de 70% des connectés à un canal ne sont pas actifs dans le canal mais ce n'est pas à dire que beaucoup de personnes observent l'activité d'un canal et peuvent ainsi être tenus pour des *lurkers* [55] .

"Le lurking c'est rester dans un canal sans parler. Quelques canaux interdisent les lurkers et vous éjecteront du canal. Il ne semble pas que tout le monde apprécie vraiment les lurkers. Après tout vous êtes assis là, à écouter leurs conversations sans y participer du tout. Tout bien considéré, vous devriez constater qu'aussi longtemps que vous répondez lorsqu'on vous appelle ou lorsqu'on vous parle directement, vous devriez être OK." [56]

D'une part la netiquette [57] recommande aux néophytes de commencer par une période d'observation avant de participer à des discussions sur un canal, et d'autre part, l'accès à un canal public donne l'accès à la création de canaux privés, si bien qu'une grande partie des internautes considérés comme présents sur un canal est occupée à des discussions dans des canaux privés. C'est ainsi que notre présence sur le canal ne s'est manifestée que très discrètement [58] puisque nous n'avons eu aucune interaction avec les internautes

[52] Les effets de l'observation sur le comportement linguistique sont identifiés comme un problème majeur dans la littérature.

[53] Voir sur ce point Douglas-Cowie (1978), Trudgill (1983) ; Milroy (1987).

[54] Il ne s'agit bien évidemment pas ici de dire que nos Corpus Présentent un dialecte selon le terme employé par F. Gadet.

[55] Personnes qui se cachent, qui restent tapies dans un coin.

[56] ***"Lurking is hanging in a channel and not talking. Some channels prohibit lurkers and will kick you out of the channel. It doesn't seem like anybody really appreciates lurkers. After all you are sitting there listening to their conversations without partaking in it at all. All in all you should find that as long as you respond when you are called or spoken directly to, you should be OK." http://www.kelseypub.com/irc/terms.shtml***

[57] Nous avons indiqué au chapitre 1 qu'il s'agit d'un ensemble de règles de bonne conduite sur les réseaux. Précisons qu'il existe une netiquette pour chaque service et des netiquettes particulières à certains serveurs, canaux…

en vertu de la loi du droit d'auteur.

des canaux publics et nous sommes contentées de répondre de temps à autre à quelques sollicitations privées ainsi que le requiert la netiquette.

L'enregistrement des sessions est prévu par les logiciels de communication en réseau synchrone. La netiquette précise que les enregistrements de sessions peuvent être une preuve à mettre en avant auprès des propriétaires de canaux lorsque des agissements abusifs ont eu lieu. Nous avons utilisé cette fonctionnalité pour collecter les corpus, donc à d'autres fins que celles prévues.

La collecte de corpus est une activité déontologiquement difficile. Pour obtenir un objet scientifiquement valide, il faut minimiser l'influence de l'observation ou être en mesure de la calculer. L'idéal pour un corpus linguistique est que l'observation ne soit pas connue des locuteurs. Toutefois, enregistrer les conversations de personnes à leur insu est incriminable, quand bien même ces conversations seraient publiques. On peut certes mettre en situation des locuteurs avertis. Cette pratique n'écarte cependant pas la possibilité d'effets de laboratoire et souvent pour les minimiser, on trompe les locuteurs sur l'objet de l'étude au départ. Nous aurions pu mettre en place un canal expérimental indiquant à la connexion que les conversations étaient enregistrées à des fins scientifiques. Mais on ne peut pas espérer qu'un canal vive de lui-même sans une communauté de participants attachée à son existence, nous aurions dû dans ce cas nous impliquer dans une observation participante [59] qui aurait grandement contribué à la forme de l'objet obtenu. Nous avons donc préféré observer l'existant en minimisant notre influence d'observateur ce qui nous a valu un gros travail d'anonymisation des corpus dont nous expliquons la procédure au chapitre 4. Ainsi, en lisant notre étude, on ne saura pas qui a dit quoi à qui ni dans quel canal pour des raisons de préservation de l'anonymat des internautes. On nous reprochera peut-être de prendre de trop grandes précautions puisqu'il s'agit de canaux publics et que les locuteurs sur les canaux publics savent qu'ils sont enregistrés par un robot logiciel en permanence et potentiellement sporadiquement par d'autres personnes, ainsi que nous avons pu le vérifier dans nos corpus.

- Extrait du Corpus F8 [60]

PAK: PDY demande a XXX ses logs [61] ... tu verras %add: PDY

- Extrait du Corpus P7

PBH: tu oublies les log PBB. %add: PBB

- Extrait du Corpus F4

[58] La seule manifestation de notre présence est l'apparition d'un surnom dans la liste des participants, surnom que nous avons créé en fonction d'observations des caractéristiques des surnoms de sorte que le nôtre n'attire pas l'attention. Nous avons créé un surnom de type sans majuscules en nous conformant au système de la langue française. Ce surnom a été crée de sorte qu'il apparaisse en fin de liste mais ne soit probablement jamais le dernier de la liste.

[59] Spradley (1980)

[60] Ces extraits de corpus sont tirés des versions traitées pour l'anonymat des internautes. Nous ne présentons jamais les corpus non-anonymisés. Les suites de lettres en majuscules sont des codes remplaçant les noms propres (surnoms), 'XXX' est le nom d'un robot logiciel. On se reportera au chapitre suivant pour les détails des conventions de transcription. Les lignes débutant par *%add:* présentent le code de l'allocutaire.

[61] ***Un log est un fichier d'enregistrement d'une session.***

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

PAG: tu lis quoi ? %add: PCA *PCA: un log %add: PAG

Il nous a toutefois semblé, par respect pour les internautes qui ont permis cette étude, tout à fait nécessaire et indispensable de ne pas lésiner sur les moyens et le degré de l'anonymisation à mettre en oeuvre.

Avant d'en venir aux traitements qui y sont liés, nous poursuivons ce chapitre par une présentation quantitative des corpus et une description technique des fichiers bruts obtenus.

3.2. Présentation générale des corpus

Il s'agit pour nous maintenant de préciser la méthode de collecte des corpus employée, et de donner les caractéristiques générales de nos corpus. Cette section et les suivantes s'efforcent à la fois de mettre à jour la façon dont nous avons procédé pour éventuellement permettre et faciliter la reproduction de cette expérience, et de comparer nos deux corpus d'un point de vue très général.

Les statistiques de fréquentation des canaux publiées sur le Web nous ont permis de choisir les moments de la journée et les jours de la semaine auxquels effectuer nos enregistrements. Nous avons sélectionné pour chacun des deux canaux – dont les statistiques de fréquentations étaient tout à fait similaires – des moments à fréquentation potentiellement moyenne.

À l'issue de la collecte, il est possible de présenter les propriétés générales de nos corpus, c'est-à-dire les plages d'enregistrement des corpus, ainsi que le volume des données récoltées. On se reportera à la section 3.4, pour des informations relatives à l'activité sur les canaux.

Dans le tableau (3-1), on trouve dans la première colonne, le nom du corpus (rappelons que le premier caractère du nom correspond au canal, soit F ou P et que le second correspond à la session enregistrée), dans la deuxième et la troisième colonne l'heure de début d'enregistrement et l'heure de fin d'enregistrement sont celles de Paris (France). À partir de ces deux dernières informations est calculée la durée d'enregistrement. La dernière cellule du tableau totalise les durées d'enregistrement de toutes les sessions.

Tableau (3-1) – Propriétés temporelles des corpus

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

corpus	Heure de début d'enregistrement	Heure de fin d'enregistrement	Durée
F1	12:23	13:07	0:44
F2	19:11	19:54	0:43
F3	17:35	19:39	2:04
F4	12:21	16:09	3:48
F5	16:16	18:36	2:20
F6	16:08	18:39	2:31
F7	11:33	18:24	6:51
F8	11:55	14:30	2:35
P1	12:19	13:07	0:48
P2	19:10	19:56	0:46
P3	17:35	19:38	2:03
P4	12:21	16:09	3:48
P5	16:07	18:39	2:32
P6	11:32	18:24	6:52
P7	10:43	16:50	6:07
P8	17:51	19:25	1:34
P9	11:54	14:30	2:36
		total	48:42

L'enregistrement des sessions a été effectué simultanément [62] pour les deux canaux à l'exception de 3 corpus : F5, P7 et P8. Le tableau (3-2) présente les correspondances en question.

Tableau (3-2) – Correspondances des enregistrements

	Calendrier des enregistrements									
	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10
Corpus F	F1	F2	F3	F4	F5	F6	F7	-	-	F8
Corpus P	P1	P2	P3	P4	-	P5	P6	P7	P8	P9

L'enregistrement des sessions permet d'obtenir des fichiers qui ne sont pas exploitables tels quels étant donné la taille du corpus, présentée dans le tableau ci-après. Dans ce tableau, la première colonne présente le nom du corpus, il est suivi de sa taille en octets (unité de mesure informatique, un octet correspond en gros à un caractère). La troisième colonne informe sur la part du volume que représente une session dans un canal, tandis que la dernière colonne fournit cette information pour le total du corpus (F1 représente 6% des données enregistrées pour le canal F et 4% du total des données enregistrées).

Tableau (3-3) – Taille des fichiers de logs (en octets)

corpus	taille	part dans	part dans

[62] Des différences de quelques minutes sont toutefois observées.

en vertu de la loi du droit d'auteur.

	(octets)	le canal	le corpus
F1	70 066	6%	4%
F2	53 414	4%	3%
F3	149 097	12%	9%
F4	107 402	9%	6%
F5	135 848	11%	8%
F6	144 564	12%	9%
F7	385 508	31%	23%
F8	182 724	11%	15%
P1	22 866	5%	1%
P2	31 919	7%	2%
P3	63 530	14%	4%
P4	34 976	8%	2%
P5	34 267	8%	2%
P6	107 034	24%	6%
P7	72 363	16%	4%
P8	37 733	8%	2%
P9	46 541	10%	3%

Ces premières données mettent en évidence, ainsi que cela est plus clairement présenté au tableau suivant, le fait que la durée d'enregistrement n'entretient pas un rapport simple avec la taille d'un corpus.

Tableau (3-4) – Mise en correspondance tailles/durées

corpus	octets/min.	car./min	rang	rang durée	rang taille
F1	1592,41	1312,29545	1	7	7
F2	1242,19	1035,09302	2	8	8
F3	1202,4	983,612903	3	6	3
F4	471,061	387,20614	8	2	6
F5	970,343	785,714286	5	5	5
F6	957,377	784,993377	6	4	4
F7	937,976	772,545012	7	1	1
F8	1178,86	975,483871	4	3	2
P1	476,375	389,458333	3	8	9
P2	693,891	568,108696	1	9	8
P3	516,504	420,934959	2	6	3
P4	153,404	124,947368	9	3	6
P5	225,441	184,802632	7	5	7
P6	259,791	214,053398	6	1	1
P7	197,174	163,444142	8	2	2
P8	401,415	331,765957	4	7	5
P9	298,34	247,141026	5	4	4

Les fichiers du Corpus F renferment en majorité une plus grande densité

en vertu de la loi du droit d'auteur.

d'information que les fichiers du Corpus P, il appartiendra à des analyses ultérieures d'expliciter ce fait.

Nous donnons ci-après la description technique des fichiers collectés qui a autorisé la mise en place de procédures de traitements automatiques. Ces procédures permettent d'une part d'obtenir une première caractérisation des corpus, et d'autre part, sont un préalable au formatage des données pour un traitement linguistique.

3.3. Format des fichiers bruts

Nous avons présenté au chapitre 2 les principes de l'activité d'un canal. Nous exposons ci-après les informations qui ont permis la récupération et le traitement des corpus et notamment une analyse des fichiers bruts en vue de traitements automatiques les rendant exploitables d'un point de vue linguistique.

Le logiciel client IRC que nous avons utilisé permet l'enregistrement des *logs* séparément pour chaque canal. Il suffit de configurer le logiciel pour l'enregistrement lors d'une session, celui-ci crée alors un fichier au nom du canal dans lequel l'activité du canal est enregistrée séquentiellement. Puis à chaque nouvelle session sauf indication contraire de l'utilisateur, l'activité des canaux auxquels l'utilisateur est connecté est enregistrée à la suite dans les fichiers créés lors du premier enregistrement. L'illustration suivante (3-1) présente le fichier de *log* issu d'une session qui a été réalisée à cette seule fin.

Exemple de fichier

```
Session Start: Tue Oct 16 17:04:27 2001
[17:31] *** Now talking in #test
[17:32] <limo> hello
[17:32] *** limo is now known as prune
[17:33] <prune> salut
Session Close: Tue Oct 16 17:05:18 2002
```

Illustration (3-1) – Session IRC

Cet exemple très court permet de se faire une idée globale du type de fichiers que nous avons traités. On peut en effet observer qu'il existe différents types de lignes. Notamment, le début et la fin du fichier sont marqués de façon particulière par les lignes suivantes :

Session Start: Tue Oct 16 17:04:27 2001 … Session Close: Tue Oct 16 17:05:18 2002

Certaines lignes représentent les interventions des participants :

[17:32] <limo> hello [17:33] <prune> salut

Tandis que d'autres sont générées automatiquement par le système à la suite de certains événements :

[17:31] *** Now talking in #test [17:32] *** limo is now known as prune

Dans les sous-sections suivantes, nous tentons d'exposer de façon exhaustive les

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

spécifications techniques des fichiers de corpus que nous avons collectés. Commençant par les frontières de session, nous exposons ensuite les types de lignes qui forment le contenu des sessions. L'analyse de ces spécifications permet ensuite d'appliquer des traitements automatiques distinguant l'activité linguistique spontanée de l'activité générée par le système en contrepartie d'événements causés par les participants.

Pour la présentation technique des différents types de lignes, nous utilisons les conventions suivantes.

Le texte invariable apparaît en caractères droits, tandis que les parties changeantes sont représentées par des noms de variables en caractères italiques. Les espaces sont représentées par le caractère ␣ en décalage par rapport à la ligne de base, tandis que les retours ligne sont représentés par le caractère␤. Ainsi, les lignes

<limo> hello <prune> salut

sont représentées par :

<surnom_du_participant> ␣ *message_du_participant*␤

3.3.1. Frontières de sessions

Pour un fichier de *logs*, à chaque session apparaît une ligne mentionnant l'ouverture de la session, suivie d'une ligne mentionnant le nom du canal, puis d'une ligne concernant la propriété "*topic*" du canal. Ces lignes se présentent comme suit, les variables sont en italiques, le reste est le texte fixe apparaissant à chaque ligne du même type.

Premières lignes d'une session : Session Start: Jjj Mmm cc hh:mn:ss YYYY [63] ***#***
****** Now talking in #nom du canal#***

*** Topic is 'texte du sujet de conversation'␤

Les lignes précédées de '***' apparaissent à l'utilisateur dans l'espace d'activité de la fenêtre du canal lorsqu'il se connecte à ce canal. Ce sont des lignes locales auxquelles correspond une unique ligne serveur signalant aux internautes connectés à la salle l'arrivée du participant. À ces lignes que nous dénommerons en-têtes de *logs*, répond une ligne locale de fin de session. Cette dernière correspond à une ligne serveur signalant aux internautes connectés à la salle le départ du participant.

Ligne de fin de session :

Session Close: *Jjj Mmm cc hh*:*mn*:*ss YYYY* ␤

Le repérage automatique de ces lignes a donc permis la segmentation du fichier de

[63] ***Les variables représentant des informations temporelles se présentent comme suit. Trois caractères dont le premier est en majuscule codent le jour de la semaine (Jjj) selon une liste finie (Mon pour lundi, Tue pour mardi, Wed pour mercredi, Thu pour jeudi, Fri pour vendredi, Sat pour samedi, et Sun pour dimanche). L'information suivante représente le mois (Mmm) selon le même principe. Suit, codé sur 2 caractères le numéro du jour (cc) dans le mois. Puis, les heures, les minutes et les secondes sont représentés par des groupe de deux chiffres séparés par des deux points (hh:mn:ss). Enfin, l'année est représentée par quatre chiffres (YYYY)***

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

logs et la récupération des propriétés temporelles de chaque session.

3.3.2. Activité du canal

Entre l'en-tête de *log* et la ligne de fin de session se trouve l'activité du canal qui est enregistrée telle qu'elle est présentée à l'utilisateur au cours de la connexion. Les lignes d'activité du canal sont formellement de trois types : les lignes de message, les lignes d'actions, et les lignes système. Les logiciels clients IRC permettent d'assigner une couleur particulière à chacun de ces types de lignes de sorte que la lecture soit facilitée pour l'utilisateur.

La syntaxe de ces types de lignes est assez simple à mettre en évidence. L'analyse a permis des traitements automatiques pour l'exploitation des corpus en vue d'une analyse linguistique. Nous la présentons ci-après.

3.3.2.1. Activité message et action

Les lignes de message ne connaissent pas de limite en taille risquant d'entraver la communication, elles peuvent s'étaler sur plusieurs lignes écran [64] . La ligne la plus longue de notre corpus présente 307 caractères (soit une soixantaine de mots).

Une ligne de message se présente comme suit (les variables apparaissent en italiques) :

Ligne de message

<surnom_du_participant> ⇧message_du_participant↵

La procédure de reconnaissance de ces lignes se base donc sur un début de ligne en "<", le surnom du participant est ensuite isolé en récupérant la chaîne de caractères entre chevrons et en remplaçant le premier caractère qui suit le chevron fermant par un caractère spécial non ambigu. L'intervention du participant est reconnue comme la chaîne de caractères suivants jusqu'au retour chariot, signe de fin d'intervention.

Ligne d'action

** ⇧message_du_participant↵*

Les lignes d'action sont identifiées sur le début de ligne : une astérisque suivie d'une espace. L'intervention du participant est reconnue comme précédemment. Il faut noter que le premier mot qui apparaît sur une ligne d'action est le surnom de l'auteur de cette ligne.

3.3.2.2. Activité système

[64] Rappelons que la notion de ligne au sens traditionnel n'a pas de pertinence pour nous, puisque un message ne prend pas le même nombre de lignes sur l'écran en fonction de différents paramètres (taille de l'écran, taille de la fenêtre, définition, taille des caractères…). Sans le spécifier, lorsque nous employons le terme de ligne, c'est pour référer à une suite de caractères terminée par un caractère spécial qui est un retour chariot.

en vertu de la loi du droit d'auteur.

Les lignes systèmes apparaissent en fonction des événements causés par les utilisateurs : arrivées, départs, changements de surnoms, changements de statuts des utilisateurs, changements des propriétés du canal. Leur forme commune est la suivante (ce qui apparaît entre crochets est optionnel).

Ligne système

*** *message_généré_par_le_système*[*(message_du_participant)*]

C'est le contenu de la variable *message_généré_par_le_système* qui varie suivant le type d'événement qui a causé le message système. Nous détaillons ci-après les différents cas.

POUR LES ENTRÉES/SORTIES DU CANAL

Une ligne apparaît dans la fenêtre d'activité du canal chaque fois qu'un nouveau participant arrive ou quitte le canal, permettant aux autres utilisateurs de réagir en fonction. Ces lignes sont dans une certaine mesure redondantes avec la liste des participants (cf. chapitre 2 – 2.2.4), elles ont toutefois l'avantage de présenter l'information au moment où l'événement se produit dans la zone où est le plus souvent concentrée l'attention des participants.

Lorsqu'un nouvel utilisateur se connecte au canal (entrées) la ligne système a la forme suivante :

Ligne système : entrées

*** *surnom_du_participant* has joined *#nom_du_canal*

Quant aux sorties de canal, il en existe deux types – nous ne considérons ici que les sorties volontaires du canal, les sorties par la force sont traitées au point suivant – l'utilisateur peut en effet se déconnecter seulement du canal dont nous observons la fenêtre d'activité, et rester connecté au serveur et à d'autres canaux ; ou bien il peut se déconnecter du serveur, et du même coup, de tous les canaux auxquels il était connecté. L'information présentée par la ligne système correspondante n'est alors pas la même suivant le cas, informant les autres utilisateurs s'il est toujours possible de contacter l'utilisateur qui vient de quitter le canal sur un autre canal ou non.

Lorsqu'un utilisateur se déconnecte du canal seulement la ligne système a la forme suivante :

Ligne système : départs canal

*** *surnom du participant* has left *#nom du canal*[*(message du participant)*]

Et, lorsqu'un utilisateur se déconnecte du serveur IRC la ligne système a la forme suivante :

Ligne système : départs serveur

*** *surnom du participant* has quit IRC[*(message du participant)*]

Ainsi que nous l'avons annoncé, d'autres événements peuvent se produire et

en vertu de la loi du droit d'auteur.

engendrer des lignes systèmes, nous les présentons ci-dessous.

POUR LES CHANGEMENTS DE PROPRIÉTÉS DU CANAL, ET DES PARTICIPANTS :

Nous avons vu au chapitre 2 que les canaux et les participants sont définis par des propriétés. Ainsi un canal est caractérisé par son nom, un thème de conversation, un statut (public, secret, privé, sur invitation), un état (modéré ou non) et le nombre de ses participants ; tandis qu'un utilisateur est caractérisé par son surnom, son statut, et l'ensemble d'informations qu'il a fournies lors de sa connexion.

Pour ce qui concerne les canaux, seuls les utilisateurs ayant le statut d'opérateur ont le loisir d'en changer les propriétés, et cela se limite au changement de l'attribut thème de conversation (*'topic'*), et de l'attribut état (permettant de spécifier quels statuts d'utilisateurs autorisent à intervenir sur le canal).

Pour ce qui concerne les utilisateurs, les attributs surnom et statut peuvent changer au cours de la connexion.

Ainsi, lorsqu'un utilisateur change de surnom, la ligne système se compose comme suit :

Changement de surnom :

*** *surnom_du_participant* is now known as *nouveau_surnom_du_participant*

Notons que seul le participant lui-même peut changer son surnom pour un autre en cours de session.

En revanche, une déconnexion du canal peut avoir été imposée à un utilisateur, et ce seulement par un utilisateur ayant le statut d'opérateur. Dans ce cas, la ligne apparaissant dans la fenêtre d'activité du canal se présente comme suit :

Départs forcés :

*** *surnom_participant* was kicked by *surnom_opérateur* [(*message_opérateur*)]

Enfin, les commandes des opérateurs, qui permettent d'agir sur le statut des utilisateurs et sur les propriétés du canal ont les contreparties listées ci-dessous :

Interventions relatives au statut des participants :

*** *surnom_de_l'opérateur* sets mode *mode* *ip_de_l'utilisateur*

*** *surnom_de_l'opérateur* sets mode *mode* *surnom_de_l'utilisateur*

Changement de l'attribut 'topic' :

*** *surnom_de_l'opérateur* changes topic to *'nouveau_texte_du_topic'*

Changement de l'état de modération :

*** *surnom_de_l'opérateur* sets mode *mode*

Étant donné ces indications, nous avons pu établir une procédure de reconnaissance automatique des types de lignes des corpus assignant des balises à chaque ligne en

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

fonction de son type, et nous sommes en mesure de donner la proportion de chaque type de ligne dans les corpus.

3.4. Exploitation des fichiers bruts.

Avant d'en venir à l'exploitation propre à l'analyse linguistique de nos corpus, cette première série de traitements présentée permet d'obtenir des informations quantitatives sur les données, permettant de futures comparaisons.

Nous analysons ci-après la répartition des grands types de lignes dégagés dans les corpus au chapitre précédent, puis nous examinons le rythme de l'activité des canaux. Une présentation détaillée des différents types de lignes permet ensuite d'envisager l'activité conversationnelle et quelques propriétés des messages. On analyse enfin le mode de contribution des participants à l'activité.

3.4.1. Présentation quantitative des types de lignes des corpus

Les corpus recueillis sont de tailles différentes ainsi que nous l'avons exposé à la section 3.2. Dans le tableau (3-4), nous présentons les informations quantitatives disponibles pour nos corpus quant aux différents types de lignes. La colonne LS "lignes système" comptabilise le nombre de lignes qui apparaissent en contrepartie des événements causés par les utilisateurs ainsi que cela a été détaillé à la section précédente. Les colonnes "lignes de message" (LM) et "lignes d'action" (LA) selon la distinction faite à ladite section ont été totalisées dans la colonne "interventions humaines" (IH). La dernière colonne représente le nombre de lignes total pour le corpus tous types confondus (TLC). Nous présentons dans les deux dernières lignes du tableau les totaux sur l'ensemble des corpus ainsi que les moyennes qui permettent de se faire une idée globale de leur volume.

Tableau (3-4) – Types de lignes : données quantitatives pour l'ensemble des corpus

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Corpus	LS	LM	LA	Totaux IH	Totaux TLC
P1	447	121	1	122	569
F1	1206	342	11	353	1559
P2	276	451	22	473	749
F2	918	289	4	293	1211
P3	696	821	67	888	1584
F3	2501	1040	87	1127	3628
P4	534	344	22	366	900
F4	1849	668	34	702	2551
F5	2563	706	27	733	3296
P5	630	223	5	228	858
F6	2640	798	25	823	3463
P6	1696	853	20	873	2569
F7	6412	2572	77	2649	9061
P7	1204	576	12	588	1792
P8	494	368	10	378	872
P9	686	417	12	429	1115
F8	2630	1150	23	1173	3803
totaux	27382	11739	459	12198	39580
moyennes	1610,706	690,529	27,000	717,529	2328,235

On notera la part importante des lignes système comparativement aux lignes de messages et lignes d'actions, ainsi que le synthétise la Figure (3-1).

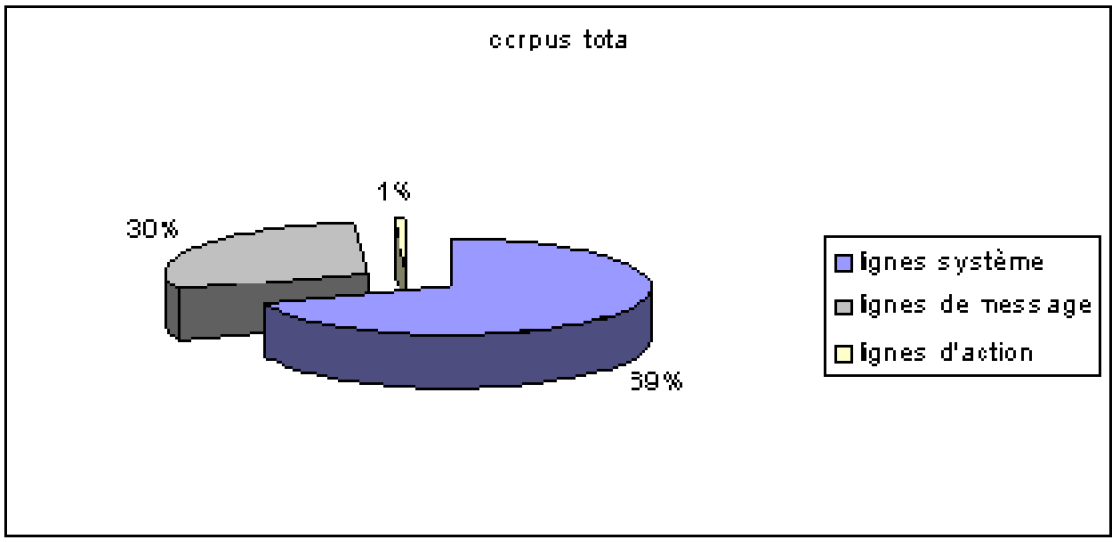

Figure (3-1) – Répartition des lignes par type sur l'ensemble du corpus

Cependant, si on détaille les données et compare les résultats pour chacun des canaux étudiés, on peut noter une différence de contenu.

en vertu de la loi du droit d'auteur.

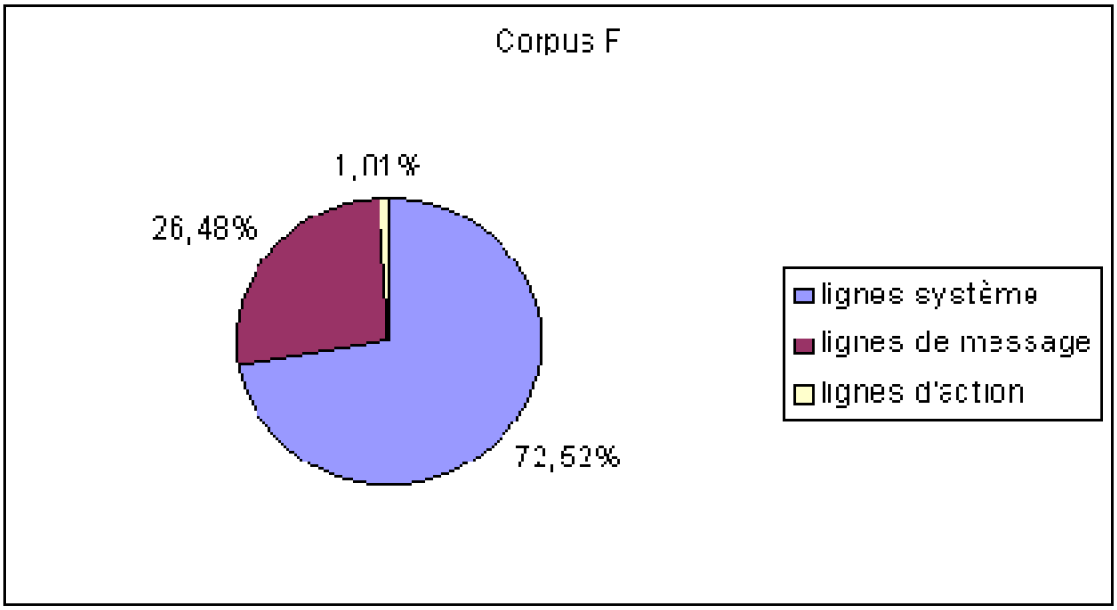

Figure (3-2) – Répartition des lignes par type – Corpus F

En effet, pour le Corpus F, ainsi qu'on peut le constater sur la Figure (3-2), les lignes système sont en proportion très majoritaire. 26,48% seulement des lignes sont d'intervention linguistique humaine. Le Corpus P contient proportionnellement plus de lignes appartenant à l'activité de conversation, ainsi que le représente la figure (3-3).

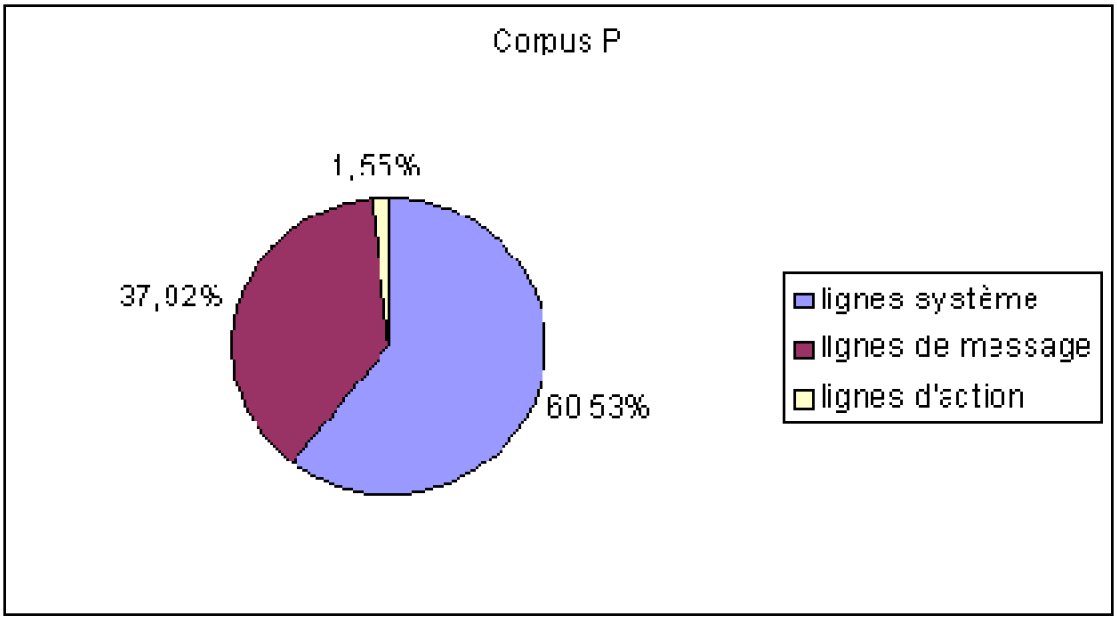

Figure (3-3) – Répartition des lignes par type – Corpus P

Cette différence ne peut s'expliquer a priori et requiert qu'on mène l'analyse plus profondément. Ainsi, l'examen des mêmes propriétés sur chacune des sessions révèle des variations ainsi que le montre le tableau, présenté ci-dessous, détaillant les proportions pour chaque corpus.

Tableau (3-5) – Types de lignes : données proportionnelles pour l'ensemble des corpus

en vertu de la loi du droit d'auteur.

	lignes système (%)	lignes de message (%)	lignes d'action (%)
P1	78,56	21,27	0,18
F1	77,36	21,94	0,71
P2	36,85	60,21	2,94
F2	75,81	23,86	0,33
P3	43,94	51,83	4,23
F3	68,94	28,67	2,40
P4	59,33	38,22	2,44
F4	72,48	26,19	1,33
F5	77,76	21,42	0,82
P5	73,43	25,99	0,58
F6	76,23	23,04	0,72
P6	66,02	33,20	0,78
F7	70,76	28,39	0,85
P7	67,19	32,14	0,67
P8	56,65	42,20	1,15
P9	61,52	37,40	1,08
F8	69,16	30,24	0,60

On remarque, il est vrai, une certaine constance à travers l'ensemble des corpus en ce qui concerne les traits de répartition des types de lignes. Cependant, on notera que certains des corpus manifestent des résultats inverses à la tendance générale (P3 et P2), comportant un nombre de lignes de messages et d'actions supérieur au nombre de lignes systèmes. L'influence de ces corpus sur les résultats du canal P n'explique qu'en partie la différence notée. Au sujet de cette dernière, à ce stade de l'examen des données, on est seulement en mesure de faire des hypothèses sur ses causes. On peut en effet supposer qu'un type de comportement générant des lignes systèmes, plus fréquent dans un canal que dans l'autre, génère ces résultats, ou bien que l'ensemble des comportements générant des lignes systèmes est en cause, ou encore qu'il s'agit d'une différence liée au type d'interaction verbale.

3.4.2. Rythme de l'activité

Ces informations peuvent être mises en correspondance avec les informations temporelles de façon à mesurer le rythme de l'activité pour chacun des corpus, donnant de premières indications sur la quantité d'unités à traiter par les utilisateurs. Nous avons défini à la section 3.3 une unité de segmentation des corpus : il s'agit de la ligne, définie comme chaîne de caractères terminée par un retour chariot – notre définition n'a donc aucun trait spatial ou visuel. Cette unité de segmentation convient dans la mesure où elle correspond à des unités d'information homogènes identifiées. Pour chaque corpus, le tableau ci-dessous présente le nombre d'interventions humaines (H/min.), le nombre d'événements (E/min.) et le nombre de lignes tous types confondus (total/min.) par

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

minute.

Tableau (3-6) – Nombre de lignes par minute pour l'ensemble des corpus

	H/min	E/min	total/min
F1	8,02	27,41	35,43
F2	6,81	21,35	28,16
F3	9,09	20,17	29,26
F4	3,08	8,11	11,19
F5	5,24	18,31	23,54
F6	5,45	17,48	22,93
F7	6,45	15,60	22,05
F8	7,57	16,97	24,54
P1	2,54	9,31	11,85
P2	10,28	6,00	16,28
P3	7,22	5,66	12,88
P4	1,61	2,34	3,95
P5	1,50	4,14	5,64
P6	2,12	4,12	6,24
P7	1,60	3,28	4,88
P8	4,02	5,26	9,28
P9	2,75	4,40	7,15

Dans le tableau (3-6), plus le score obtenu dans la dernière colonne est important, plus l'activité globale du canal en question est dense. La mesure de l'activité conversationnelle est représentée par la première colonne. On note une variation importante du nombre de lignes par minute tous types confondus se produisant sur un canal. L'activité la plus importante se comptabilise sur le Corpus F1, ce qui est fortement contribué par les lignes d'événements. En ce qui concerne les interventions humaines (messages et actions), c'est le Corpus P2 qui obtient le meilleur score. On remarquera de nouveau que le Corpus P présente globalement des rapports inférieurs c'est-à-dire qu'il manifeste une activité globale moins dense, tant pour les lignes systèmes que pour les lignes de message et d'action.

Tableau (3-7) – Nombre de lignes par minute Corpus F et P globalement

	H/min	E/min	total/min
corpus F	6,06	15,99	22,05
corpus P	2,67	4,10	6,77
total	4,17	9,37	13,55

Les calculs pour les différents paramètres envisagés manifestent des rapports équivalents à ceux envisagés plus haut. Ainsi qu'on avait pu l'inférer des résultats présentés à la section 3.2, le rythme de l'activité globale est plus important sur F que sur P. L'importance des lignes systèmes dans les corpus nous amène à considérer leur contenu et à nous interroger sur leur rapport à l'activité de conversation. Dans la section

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

suivante nous tentons de déterminer la nature de ces lignes systèmes et présentons nos observations sur leur impact dans le dispositif de communication étudié.

3.4.3. Présentation quantitative des sous-types de lignes

3.4.3.1. Types de lignes système

Nous avons caractérisé à la section 3.3.2.2. les différents types de lignes systèmes correspondant à des opérations réalisées par les utilisateurs des canaux. Nous avons distingué les opérations d'entrée et sortie, des opérations liées à des changements de propriétés du canal ou de ses participants. Les deux figures suivantes présentent les proportions des différents types de lignes système identifiés pour chacun des canaux. L'objectif est ici de déterminer si des comportements particuliers s'observent.

Nous rappelons, pour une lecture efficace des tableaux et figures de cette sous-section, que les lignes systèmes arrivées, départs canal, départs serveur et changement de nom sont les contreparties d'opérations volontaires des participants, tandis que les expulsions temporaires, et les expulsions définitives sont des opérations de force réalisées par des participants ayant le statut d'opérateurs. Les lignes systèmes "statut opérateur" sont les contreparties des événements d'attribution du statut d'opérateur à un utilisateur par un opérateur ou par un robot logiciel. Enfin, la catégorie "autres" représente les opérations de changement de l'attribut "*topic*", de l'état de modération du canal ou autres opérations de gestion du canal.

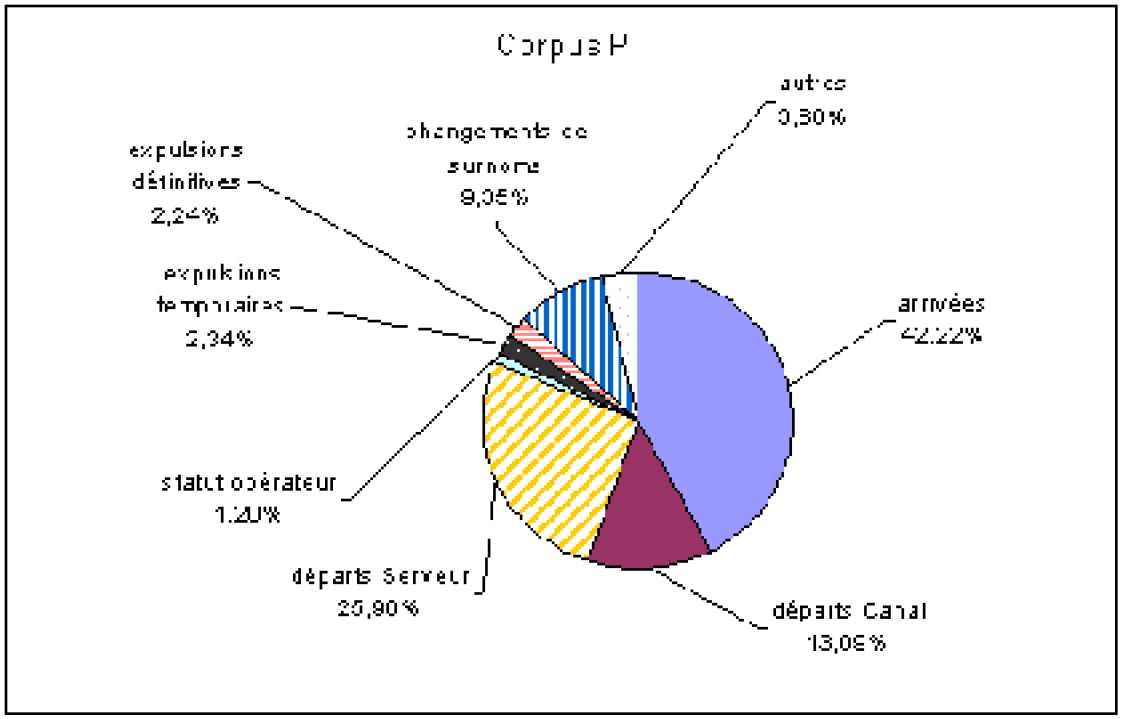

Figure (3-4) – Proportion des types de lignes système – Corpus P

en vertu de la loi du droit d'auteur.

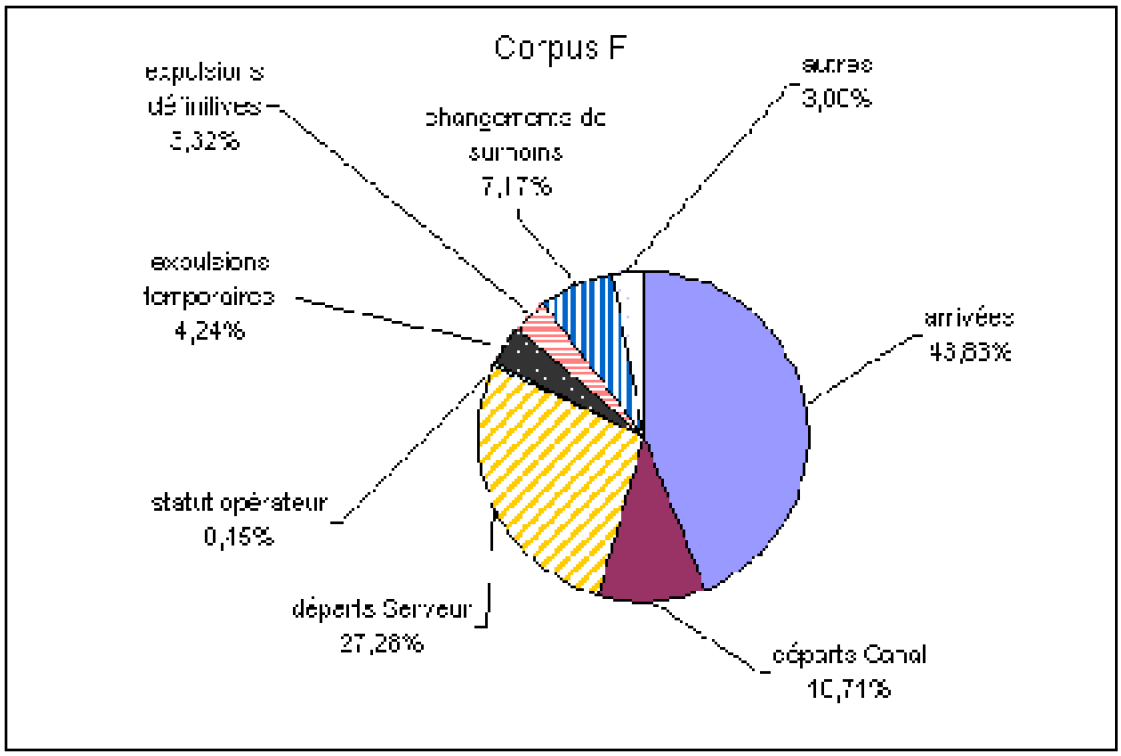

Figure (3-5) – Proportion des types de lignes système – Corpus F

Cette présentation permet de constater que la très grande partie des événements d'un canal concerne des arrivées et des départs. L'accord du statut d'opérateur ou de *voice* à d'autres utilisateurs est assez sporadique. Les coups de force des opérateurs (expulsions temporaires et définitives) ainsi que l'activité de changement de nom sont assez importants.

Si on observe la situation d'un point de vue global, on ne constate ainsi pas de différence majeure de comportement entre les deux canaux. Si de petites différences s'observent, on ne peut toutefois les incriminer. Il s'agit par exemple, de la proportion de lignes se rapportant à l'accord du statut d'opérateur, plus forte sur P que sur F, de la proportion d'expulsions plus importante sur F que sur P.

En ce qui concerne la proportion des types de lignes pour chaque canal, on peut dresser les tableaux suivants qui présentent pour chaque corpus la part de chaque type de ligne repéré à la section 3.3.2.2 :

en vertu de la loi du droit d'auteur.

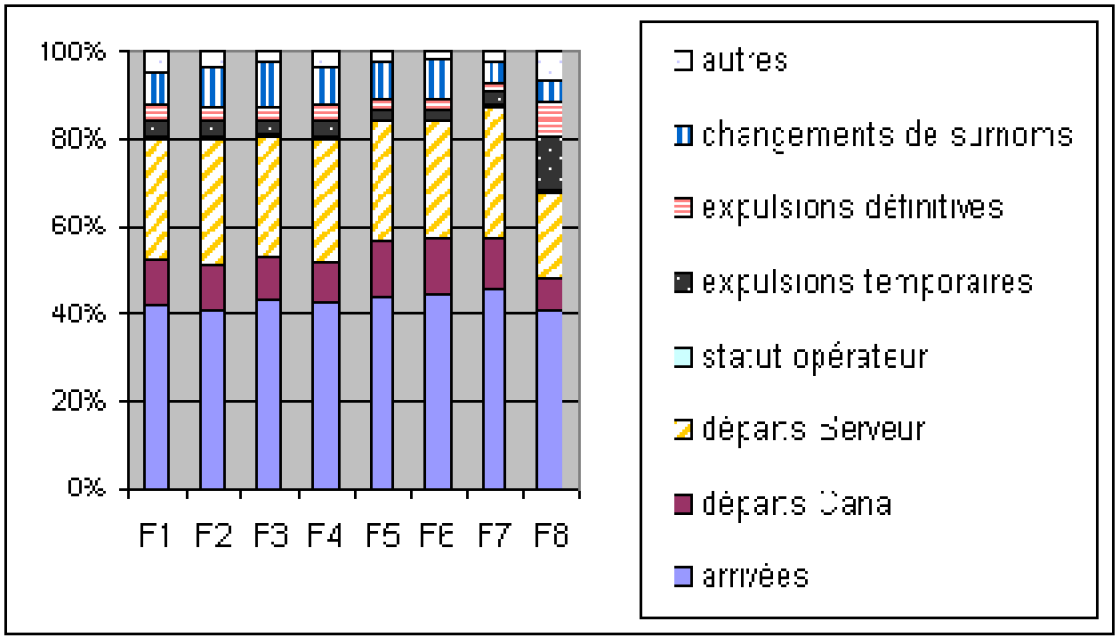

Figure (3-6) – Proportion des types de lignes système – Corpus F

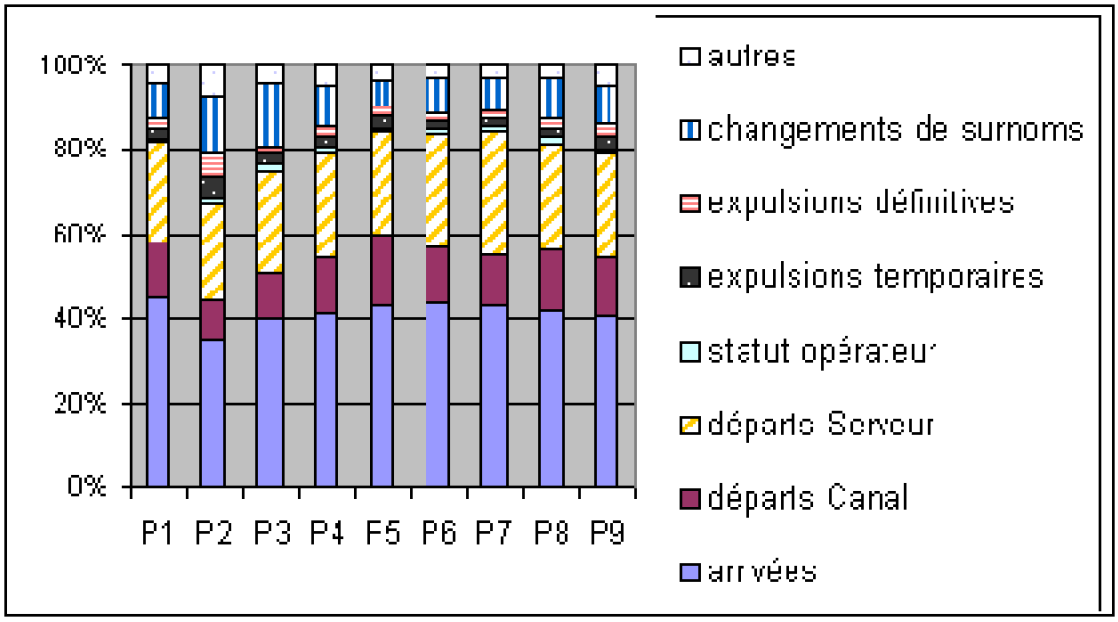

Figure (3-7) – Proportion des types de lignes système – Corpus P

Pour ces propriétés, les deux canaux ne diffèrent pas. Les proportions globales présentées figures (3-6) et (3-7) sont équivalentes. Il arrive toutefois que certains des corpus (F8 et P2 notamment) manifestent des événements de force dans des proportions supérieures à la moyenne.

Concernant la nature des lignes système, il n'y a donc pas de différence majeure entre les deux canaux étudiés. Les entrées et les sorties sont en nombre équivalent, ce qui suggère un nombre de participants assez constant dans les canaux. Elles représentent plus de 82% des lignes systèmes. Rappelons qu'il faut compter comme

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

sortie du canal les déconnexions IRC, les déconnexions du canal, les *kicks* et *bans* qui sont des départs forcés.

Ce qui nous intéressera dans la suite de ce travail est bien entendu essentiellement l'information linguistique spontanée présente dans les sessions. Les lignes d'événements sont des messages linguistiques à structure figée et se répètent à travers le corpus. Nous avons d'emblée rejeté leur étude au plan linguistique, cela va de soi, en les qualifiant de *lignes système*. Cependant, à la fin de certaines de ces lignes, un utilisateur peut ajouter un message. Après examen de ces éléments, nous avons convenu de les écarter de l'analyse dans la mesure où ils n'entretenaient aucun rapport avec l'activité de conversation en cours dans le canal [65] .

Les lignes système qui influent en propre sur l'activité de conversation sont en petit nombre, et il s'agit essentiellement des lignes signifiant l'arrivée d'un nouvel utilisateur dans un canal, mais tous les types de ligne système peuvent appeler à commentaire de la part des participants. Lorsque cela s'est trouvé être le cas dans les corpus, nous l'avons marqué.

3.4.3.2. Lignes de message et d'action des corpus

Nous avons distingué à la section précédente entre deux types de lignes de messages. La distinction se fait en fonction de la forme de ces interventions, qui apparaissent comme émanant d'un participant dans un cas, et pas dans l'autre. Sur l'ensemble des corpus, les lignes messages représentent 96,24% des interventions humaines. Il n'y a pas de différence notable entre les deux canaux de ce point de vue, puisque le Corpus P comptabilise 96,06% de lignes messages tandis que le Corpus F en comptabilise 96,33%. Les proportions des ces différents types de messages pour chaque canal apparaissent dans les deux figures suivantes pour chacun des canaux et chacune des sessions.

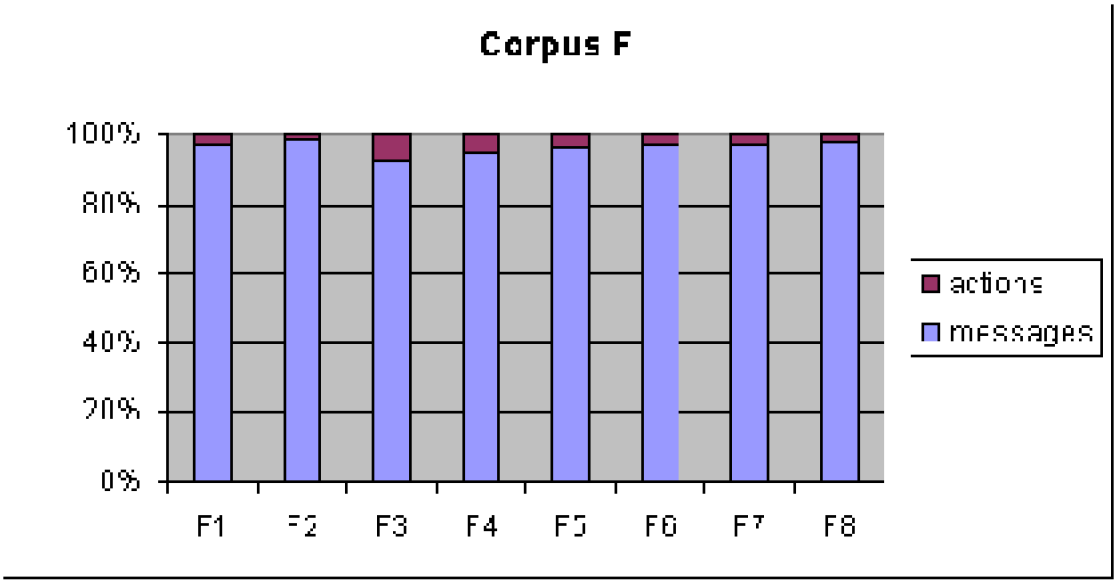

Figure (3-8) – Répartition des lignes messages et action – Corpus P

[65] On en trouvera la liste en annexe

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

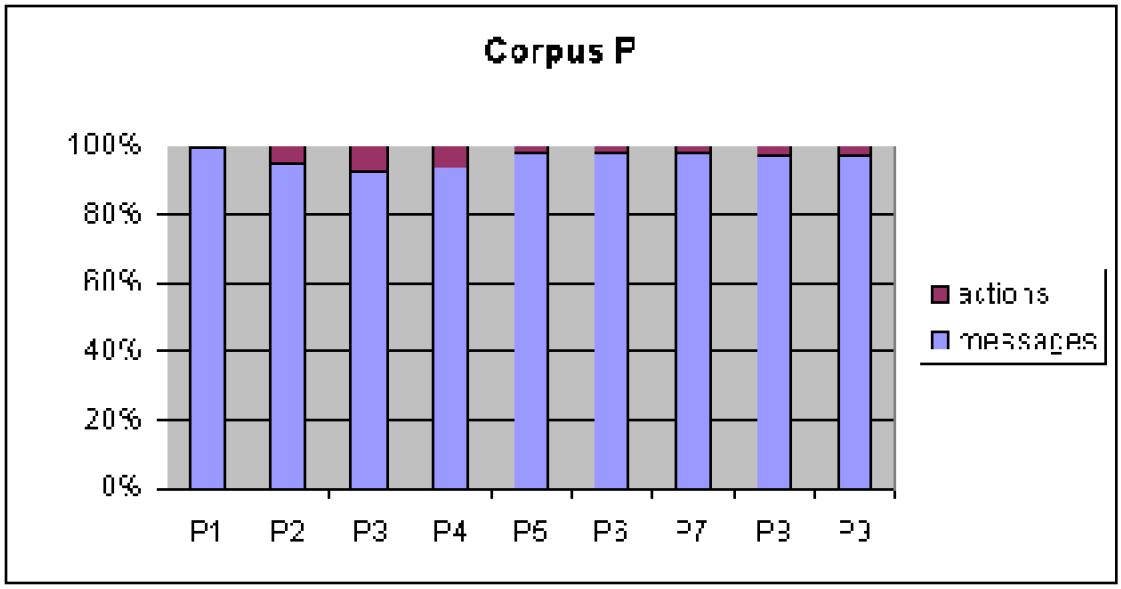

Figure (3-9) – Répartition des lignes messages et action – Corpus P

Quelques sessions se distinguent : F3, P3 et P4 présentent une utilisation proportionnellement plus importante des lignes d'action (respectivement 7,72%, 7,54% et 6,01%), tandis que P1 et F2 comptabilisent les plus faibles proportions sur ce point (respectivement 1,36% et 0,81%).

Rappelons que tout utilisateur peut produire une ligne action, à condition qu'il conforme son message à une syntaxe particulière. Ce fait explique sans doute en partie la faible utilisation de ce type de messages, puisqu'il faut à la fois avoir un commentaire à faire sur soi-même, et connaître la syntaxe de la commande.

Quant à expliquer les différences d'utilisation au travers des corpus sur ce point, il nous faudra avancer dans l'analyse linguistique. Nous pouvons signaler ici cependant, que les locuteurs les plus prolixes sont aussi ceux qui utilisent les lignes d'actions. À l'inverse il apparaît des locuteurs qui ne se manifestent pas autrement que par ce type d'intervention.

3.4.4. Rythme de l'activité conversationnelle et longueur des messages

Comme première caractérisation des éléments linguistiques de nos corpus, nous donnons des informations quantitatives relatives au rythme moyen de l'activité conversationnelle et à la longueur observée des messages dans les corpus.

La première observation mêle lignes d'actions et lignes messages pour obtenir le rythme des interventions.

en vertu de la loi du droit d'auteur.

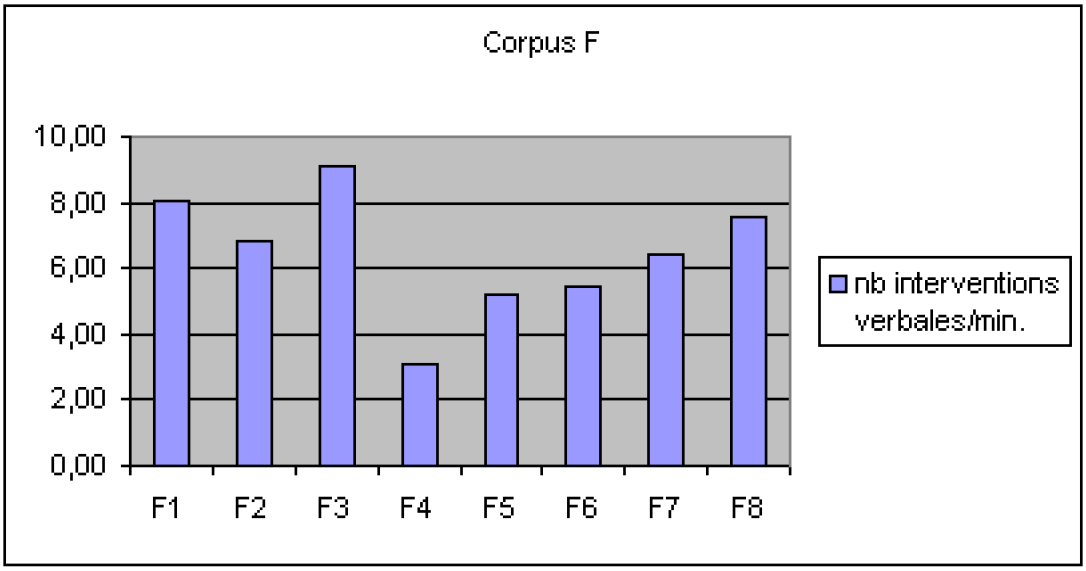

Figure (3-10) – Nombre d'interventions verbales par minute – Corpus F

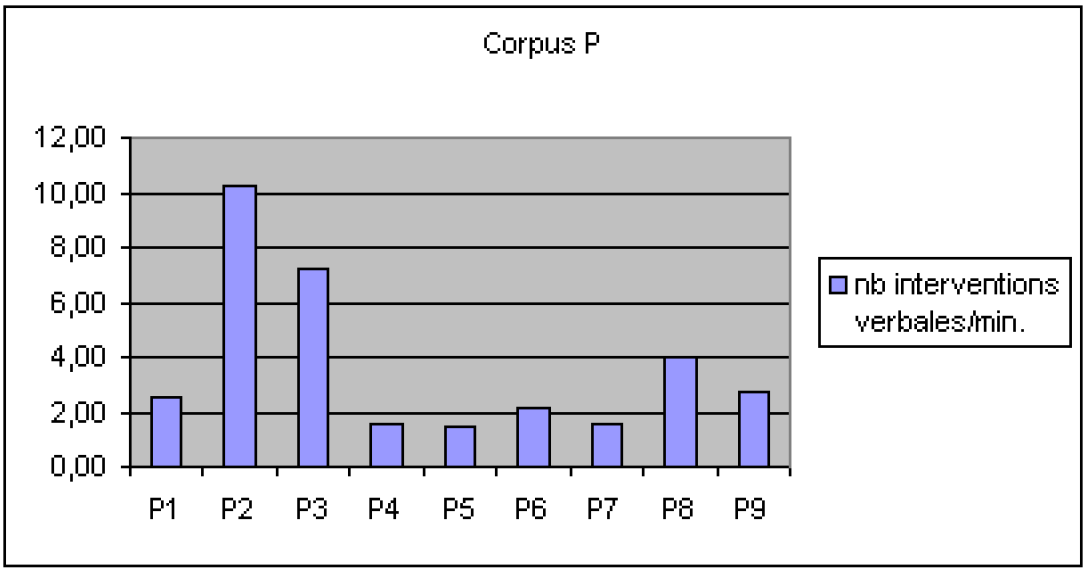

Figure (3-11) – Nombre d'interventions verbales par minute – Corpus P

On note une variation importante de cette mesure au sein de chacun des corpus, le Corpus P présentant globalement un rythme des interventions inférieur à celui du Corpus F.

Mais cette mesure ne dit rien de la quantité d'information véhiculée, puisqu'une ligne message peut varier en longueur. Il convient donc de mesurer le nombre moyen de mots [66] constituant une ligne, et le nombre moyen de mots produits par minute.

[66] À ce stade, un mot est défini comme une suite de caractères délimitée par des espaces, signes de ponctuation ou retour ligne

en vertu de la loi du droit d'auteur.

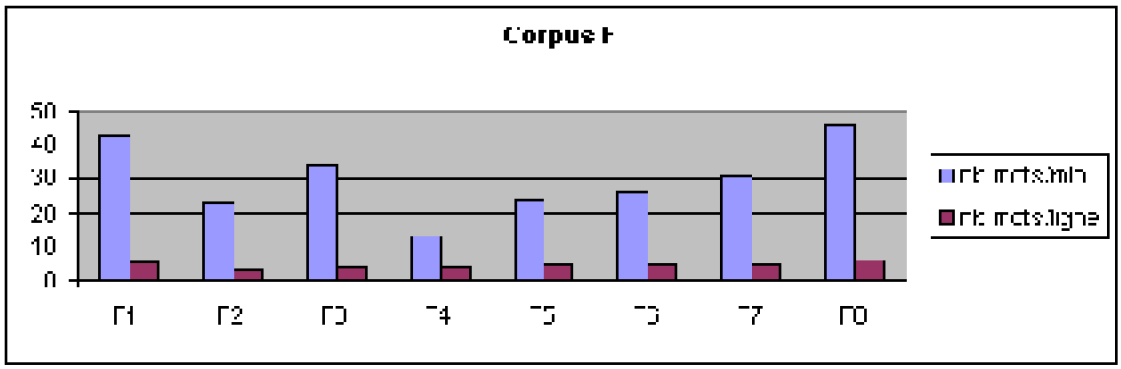

Figure (3-12) – Activité conversationnelle : rythme et longueur des messages – Corpus F

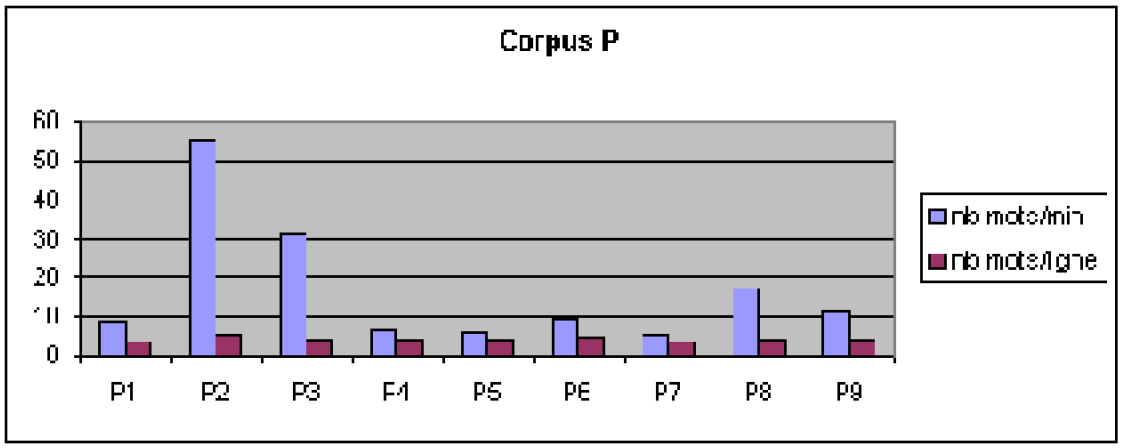

Figure (3-13) – Activité conversationnelle : rythme et longueur des messages – Corpus P

Les figures (3-12) et (3-13) présentent de grandes variations quant au rythme de l'activité (nombre de mots par minute) et une certaine stabilité à travers les corpus de la longueur des messages (nombre de mots par ligne). Le rythme de l'activité ne semble pas influencer la longueur des messages. Nous supposons que la variation du rythme est due au nombre de participants.

Le nombre de participants actifs est corrélé à la durée d'enregistrement ainsi qu'on peut l'observer sur les figures (3-14) et (3-15), ce qui est attendu puisque des participants entrent dans le canal tout au long des sessions.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

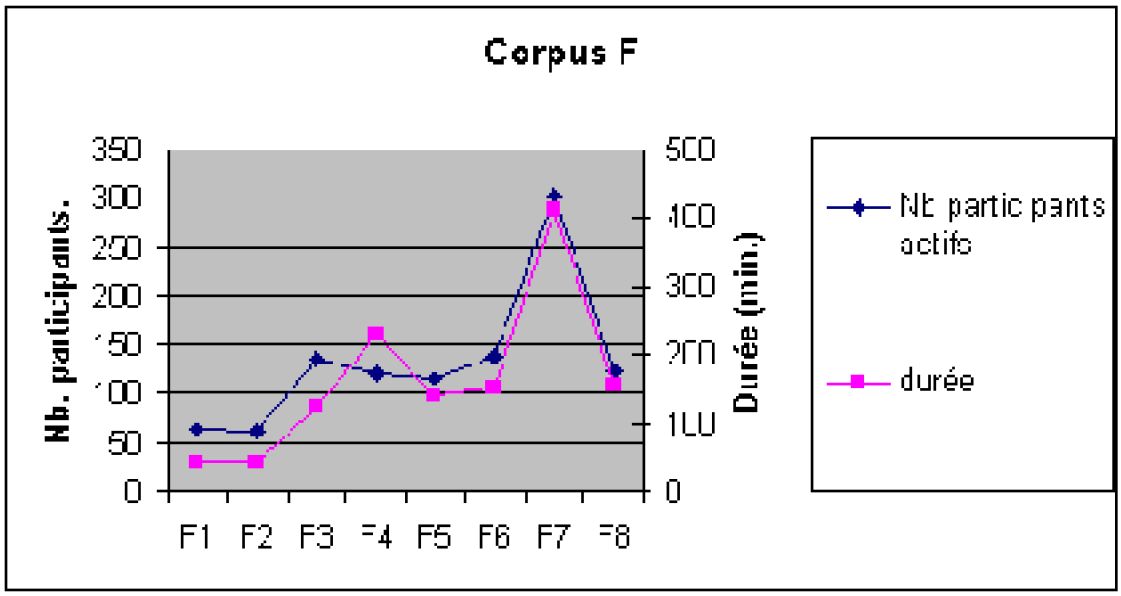

Figure (3-14) – Durée d'enregistrement/nombre de participants – Corpus F

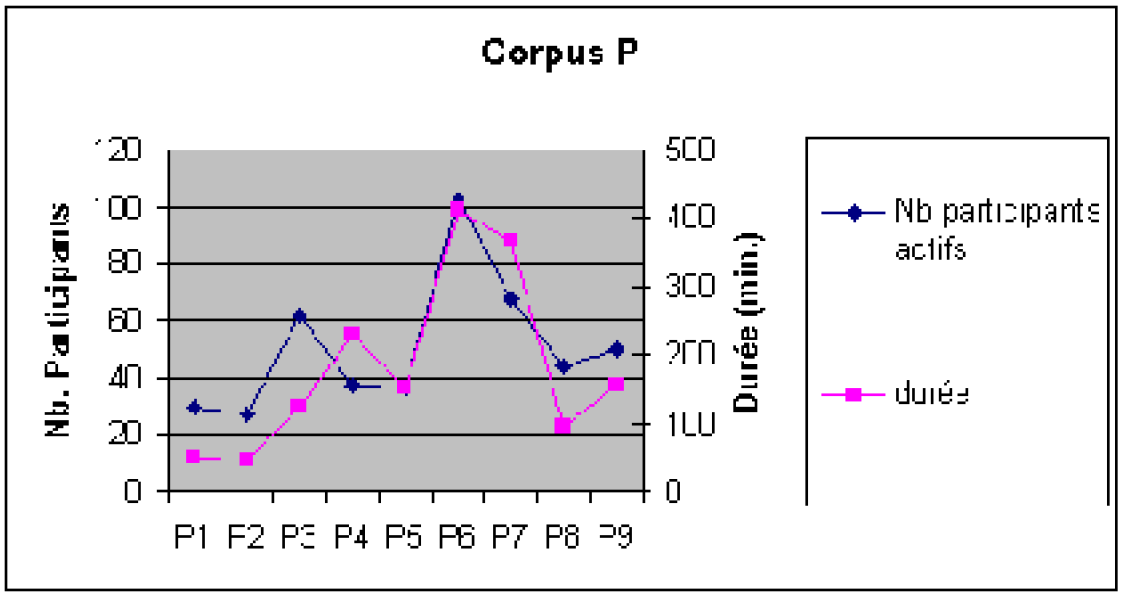

Figure (3-15) – Durée d'enregistrement/nombre de participants – Corpus P

On note que les Corpus P4 et F4 présentent un comportement particulier, qui dissocie la mesure de durée de la mesure du nombre de participants. L'activité globale pour F4 et pour P4 est la moins dense respectivement pour tout le Corpus F et P à la fois en nombre de lignes par minute (Cf. Tableau 3-6) et en nombre de mots par minute. En revanche la longueur des messages ne présente pas de différence caractéristique pour ces corpus par rapport à l'ensemble. Il ne s'agirait donc pas d'une différence due à des facteurs linguistiques.

Le nombre de participants n'influerait pas sur la longueur des messages.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

3.4.5. Mode de contribution des participants à l'activité

Il est important de rapporter le rythme de l'activité à ses générateurs, c'est-à-dire aux participants. On observe le nombre moyen d'interventions par participant pour chaque corpus et pour chaque session dans les deux figures suivantes.

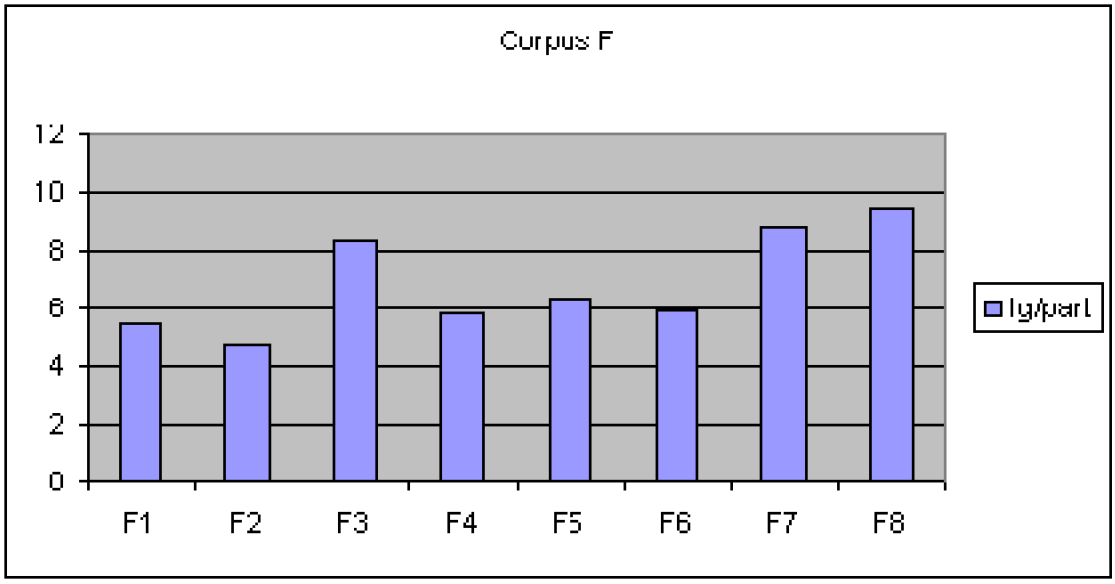

Figure (3-16) – Nombre moyen d'interventions par participant – Corpus F

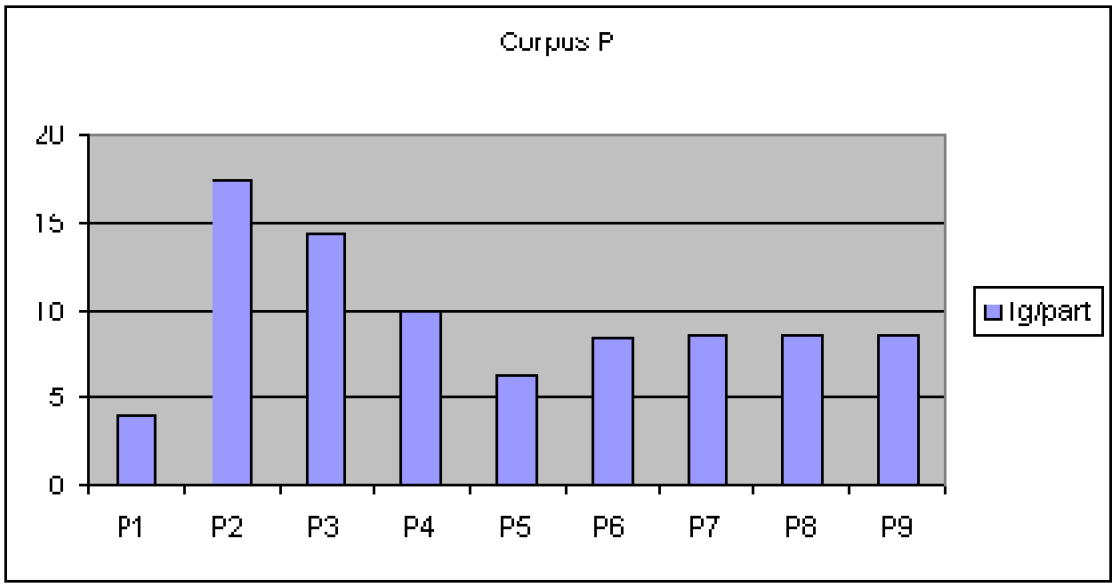

Figure (3-17) – Nombre moyen d'interventions par participant – Corpus P

On repère des variations importantes tant entre les deux canaux qu'entre les sessions de chaque canal. La figure suivante rassemble les sessions simultanées des deux canaux (1: F1 et P1 ; 2 : F2 et P2 ; 3 : F3 et P3; 4 : F4 et P4 ; 5 : F5 ; 6 : F6 et P5 ; 7 : F7 et P6 ; 8 : P7 ; 9 : P8; 10 : F8 et P9).

en vertu de la loi du droit d'auteur.

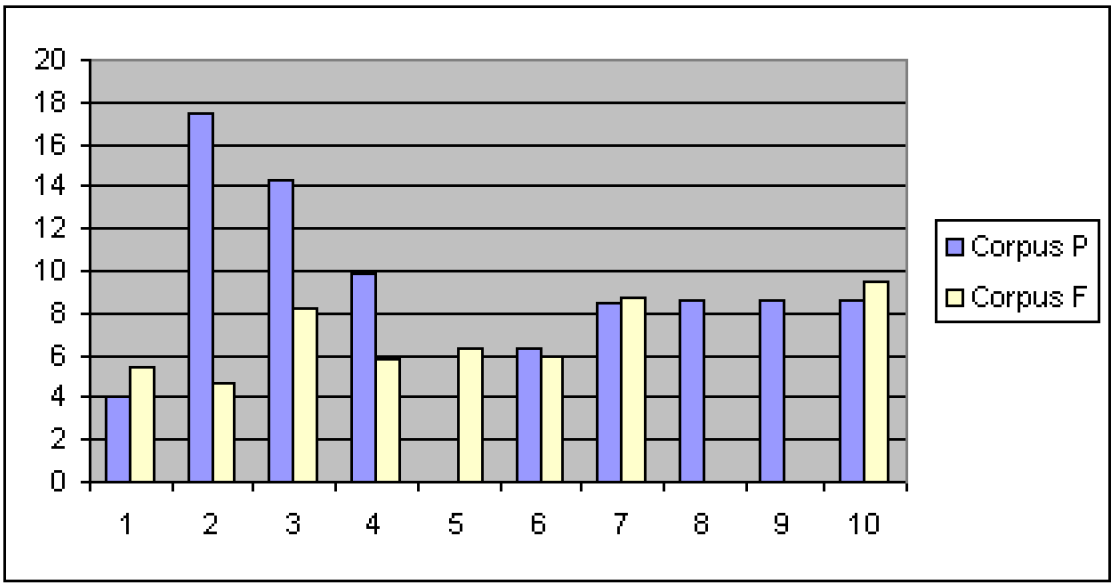

Figure (3-18) – Nombre moyen d'interventions par participant – Corpus P et F

Il ne semble donc pas que la différence soit liée aux différents moments sélectionnés pour cette étude.

Un autre facteur global d'importance est la participation relative des connectés à un canal. Dans une situation idéale, si chaque locuteur participait à égale mesure, on pourrait produire une figure présentant une droite. Ce n'est pas le cas pour nos corpus, la courbe obtenue pour chacune des sessions de chacun des canaux présente la silhouette de la série 2 sur la figure suivante, tandis que la série 1 représente la situation idéale évoquée. En abscisse se trouve le nombre de locuteurs tandis qu'en ordonnée se trouve la part de contribution de chaque locuteur.

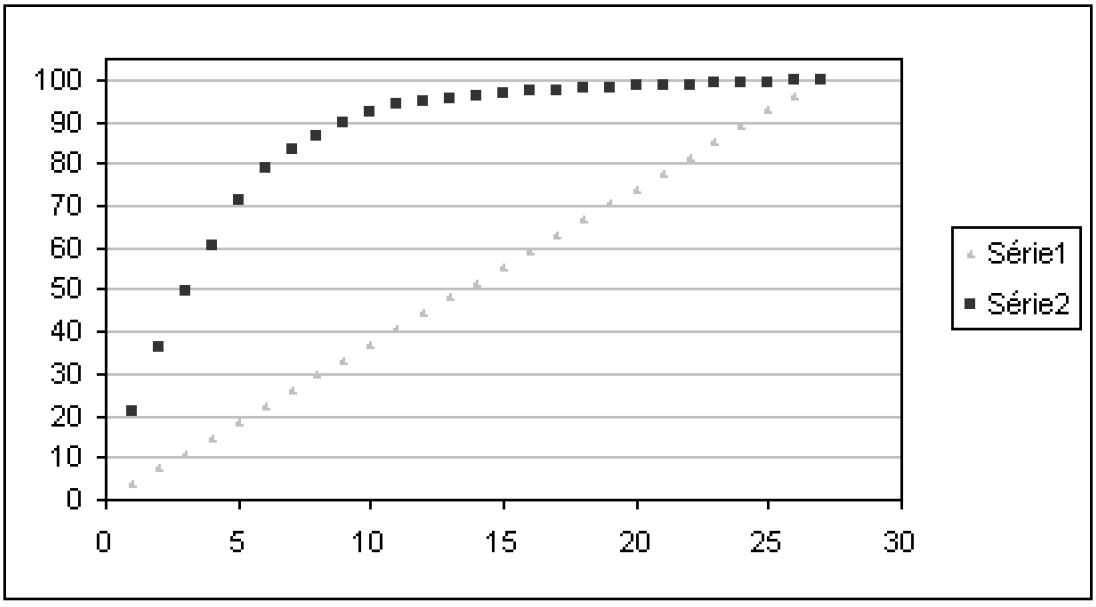

Figure (3-19) – Participation relative des connectés

en vertu de la loi du droit d'auteur.

On peut constater qu'un petit nombre de locuteurs constitue l'essentiel de l'activité d'une session. D'autre part, un grand nombre de participants ne produisent qu'une seule, ou quelques interventions. Les figures suivantes présentent cet effet. Nous avons regroupé dans le groupe 1 les locuteurs qui n'ont produit qu'une seule intervention, le groupe 2 est celui des locuteurs qui ont produit de 2 à 10 interventions, et le groupe 3 est celui des locuteurs ayant produit plus de 10 interventions.

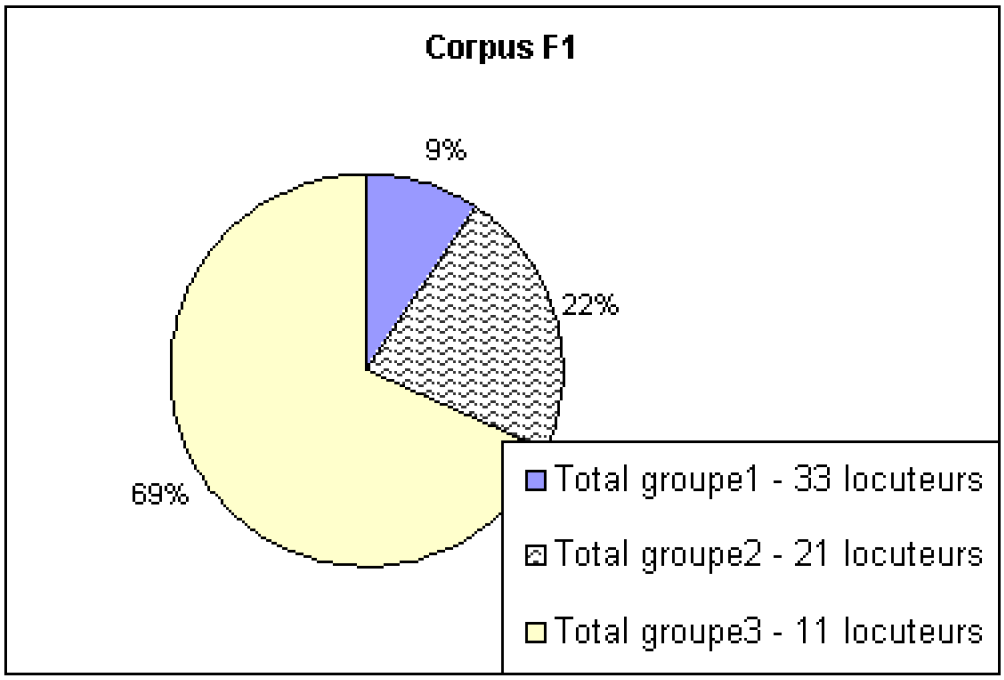

Figure (3-20) – Participation relative des connectés – Corpus F1

Cette figure (et les suivantes) se lit comme suit : 9% des interventions ont étés produites par 33 locuteurs qui n'ont produit qu'une intervention chacun. 22% des interventions ont étés produits par 21 locuteurs qui ont produit entre 2 et 10 interventions chacun. 69% des interventions ont étés produits par 11 locuteurs qui ont produit plus de 10 interventions chacun.

Nous épargnons au lecteur le détail des sessions, en signalant que la très grande majorité des corpus présente une répartition similaire. Nous nous contenterons de présenter les résultats pour chacun des corpus sur les deux figures suivantes.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

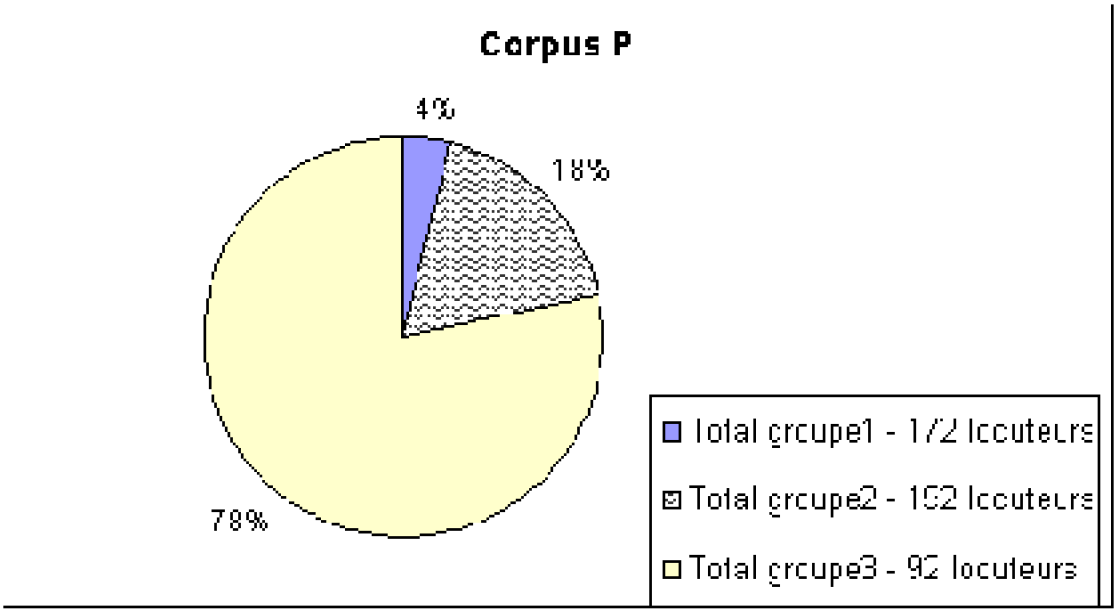

Figure (3-21) – Participation relative des connectés – Corpus P

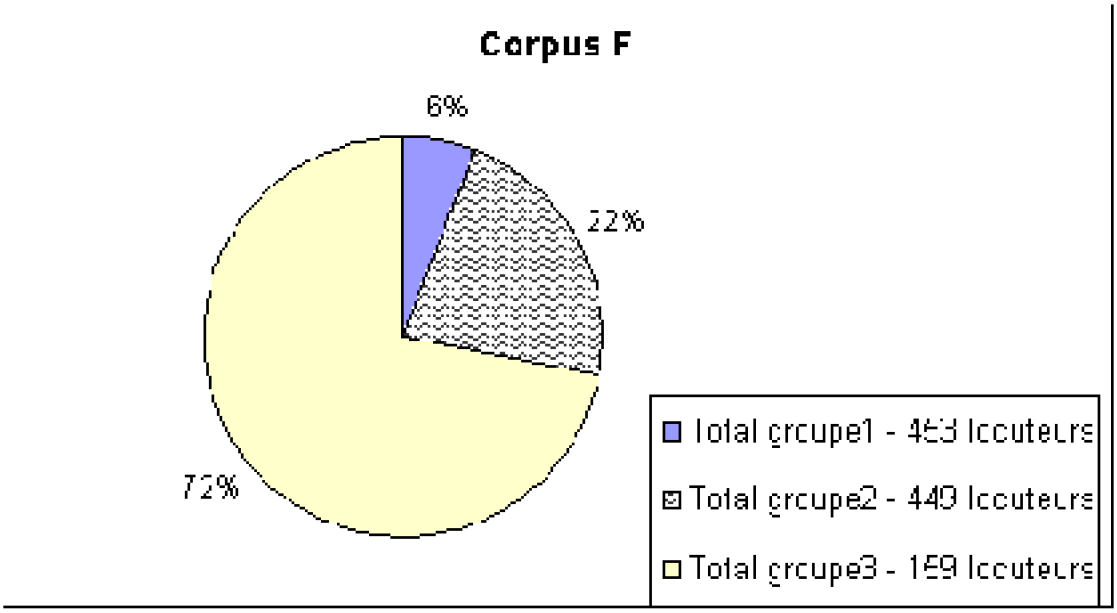

Figure (3-22) – Participation relative des connectés – Corpus F

La tendance est donc qu'une proportion restreinte de locuteurs produit l'essentiel de l'activité messages. Dans le tableau suivant, nous avons comptabilisé le pourcentage de locuteurs nécessaires à la production de 80% de l'activité messages pour chaque session de chaque corpus.

Tableau (3-9) – Pourcentage de locuteurs contribuant à 80% de l'activité messages

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

corpus	pourcentages de locuteurs ayant produit 80% de l'activité message
F1	27,69
P1	43,33
F2	40,32
P2	25,92
F3	19,12
P3	17,74
F4	25,21
P4	27,03
F5	31,03
-	
F6	26,08
P5	30,55
F7	19,53
P6	21,57
-	
P7	23,53
-	
P8	31
F8	17,89
P9	22

3.5. Conclusion

Dans ce chapitre nous avons pu mettre en évidence un certain nombre des propriétés de nos corpus. Les premières observations font état de la vitesse moyenne à laquelle se construit l'activité d'un canal. Cela nous amène à nous interroger sur l'activité de lecture des utilisateurs. La vitesse de lecture sur écran est en moyenne inférieure à la vitesse de lecture sur papier. Ainsi, un lecteur moyen lit environ 240 mots par minute sur papier, tandis qu'il lit 200 mots par minute (1294 signes par minute, espaces comprises) sur écran (25% plus lente en moyenne). Ainsi, pour certains corpus, on peut supposer qu'une grande partie des utilisateurs n'est pas en mesure de lire toute l'information qui se présente à lui et ce, d'autant plus que le rythme de l'activité n'est pas linéaire, pour plusieurs raisons. Il faut rappeler en effet que l'activité se constitue en temps réel et que c'est là, évidemment, un caractère qui distingue nos corpus des corpus écrits habituels, la vitesse de décodage dépendant alors de la vitesse de production. D'autre part, les interventions humaines ainsi que les lignes systèmes sont présentées d'un seul bloc et non pas caractère par caractère. Le nombre de participants engagés dans une unité conversationnelle ainsi que le nombre d'unités conversationnelles menées en parallèle peut localement augmenter le rythme de l'activité. Il est alors probable que les utilisateurs

en vertu de la loi du droit d'auteur.

ne lisent qu'une partie de l'activité, cela pouvant éventuellement entraîner des dysfonctionnements dans les échanges.

Nous avons montré également en observant le format des fichiers collectés que toutes les lignes n'ont pas le même statut et que chaque type doit donc faire l'objet d'un traitement particulier. Ainsi, nous avons pu observer que les lignes systèmes sont, pour la plupart des corpus, en proportion très supérieure à celle des lignes de messages, et que leur nature est particulière. Subséquemment, si les utilisateurs ne lisent qu'une partie de l'activité, il est probable qu'ils concentrent leur lecture sur les lignes de messages. D'autre part, nous avons vu que les lignes systèmes sont fortement redondantes, il est donc probable que leur décodage soit le fait d'un traitement global particulier.

Les deux possibilités d'intervenir spontanément et verbalement sur un canal ne sont pas également utilisées. Il conviendra d'observer leur fonction au moment des analyses linguistiques.

Enfin, nous avons observé qu'une petite proportion des participants actifs d'un canal constitue l'essentiel de l'activité.

en vertu de la loi du droit d'auteur.

Chapitre 4 Traitement des corpus

On l'aura compris, la présente étude se base sur un corpus de pratiques linguistiques spontanées. On attribue partiellement à Chomsky le désintérêt, manifesté par une partie de la communauté scientifique des linguistes, pour le recours aux corpus, des années soixante à quatre-vingts. Chomsky en effet formule un rejet de principe de l'utilisation des corpus au profit du questionnement de l'intuition du locuteur, principe adopté par une large part de la communauté. Sag & Wasow (1999:3), entre autres, affirment ainsi que :

"Tout locuteur normal de toute langue naturelle a acquis un ensemble immensément riche et systématique de connaissances inconscientes qui peut être exploré en consultant les jugements intuitifs des locuteurs" [67] ***.***

Or les locuteurs d'une même communauté linguistique (ou un même locuteur dans des situations différentes) n'ont pas toujours les mêmes usages. Le locuteur idéal de Chomsky n'a pas de réalité.

"Les langues manifestent de la variation et du changement et le constat de l'hétérogène est coextensif à la notion de langue." (Gadet, F. 1989:3).

Les linguistiques de corpus ou sur corpus cherchent à décrire les usages effectifs de la langue. Habert (1997) propose de distinguer deux attitudes pour la constitution de corpus et l'analyse qui en découle [68] . D'une part, la position qui consiste à tenter d'effacer les

[67] ***"Every normal speaker of any natural language has acquired an immensely rich and systematic body of unconscious knowledge which can be investigated by consulting speakers intuitive judgments"***

[68] Pour ce paragraphe nous nous inspirons de Habert *et al*.(1997) pp.143-159.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

variations en constituant de gros corpus est celle des chercheurs qui souhaitent établir des descriptions de référence, l'idée étant que si on ne parvient pas à déterminer les caractéristiques de l'ensemble des productions langagières, le fait de récolter le maximum de données aboutira à faire émerger la part représentative. D'autre part, la position est de constituer des groupes de données selon des paramètres précis concernant leurs conditions de production et de réception. Cette démarche se constitue comme "linguistique de la variation" à l'exemple de Biber (1988). Quant à l'analyse, elle est fonction des objectifs de recherche, et souvent se contente pour les domaines les plus établis de la linguistique d'appliquer les catégories et concepts descriptifs usuels [69] .

Dans notre cas, il semblait bien évidemment difficile de procéder autrement qu'en récoltant des données, d'une part parce que les locuteurs qu'il aurait fallu interroger ne se déclarent pas toujours ouvertement comme utilisateurs de *chat*, et d'autre part le dispositif de communication étant complexe, il n'est pas dit que la représentation que ses utilisateurs en ont soit parfaitement en adéquation avec la réalité.

Nous avons exposé au chapitre précédent les traitements qui ont été appliqués aux corpus bruts. Ces derniers permettent d'une part de distinguer l'activité linguistique spontanée (lignes messages et actions) de l'activité générée par le système en contrepartie d'événements causés par les participants (entrées, sorties, changements de surnoms…).

Ce chapitre présente les traitements de base qui ont été appliqués à nos corpus. Ces derniers ont été réalisés pour répondre à plusieurs objectifs. Le premier est de séparer du reste ce qui dans les corpus est d'intérêt pour nos projets d'analyse, et de rendre le corpus lisible, afin d'étudier sa matérialité. Il s'agit également d'annoter le corpus avec les informations linguistiques utiles à nos analyses (morphologie, syntaxe et perspective discursive).

Dans un premier temps, nous exposerons la façon dont nous avons traité les corpus pour sélectionner ce qui fait l'objet de notre étude, c'est-à-dire les énoncés des locuteurs francophones. Cette section fait l'occasion de caractériser les corpus quant aux codes utilisés. Nous présentons ensuite de façon globale les caractéristiques des corpus qui ont rendu nécessaires des procédures de normalisation, dont nous exposons les principes – pour permettre la lemmatisation. Les procédures d'anonymisation qui ont été mises en place et leur justification sont ensuite expliquées. Enfin, il est procédé à une présentation rapide du logiciel de traitement de corpus retenu pour cette étude.

4.1. Traitement des interventions dans des codes et langues différentes

Les utilisateurs d'IRC ne sont pas contrôlables du point de vue de leur identité sociale. Nous l'avons évoqué précédemment, il s'agit même d'une caractéristique reconnue de ce

[69] Mondada (2001).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

moyen de communication que de permettre à un locuteur de se faire passer pour autre qu'il n'est, ainsi que le suggère le message d'accueil d'un serveur IRC :

"Bienvenue sur IRC – où les hommes sont des hommes, les femmes sont des hommes et les garçons sont des agents du FBI. Mais quelques-uns des hommes sont vraiment des femmes. Amusez-vous" [70]

En ce qui concerne la langue cependant, il n'est pas possible de tromper complètement. S'il est assez aisé de se faire passer pour plus jeune ou plus vieux, d'inventer les circonstances de sa vie, en revanche il est impossible de se faire passer pour locuteur d'une langue qu'on ne parle pas et difficile de donner l'illusion d'être locuteur natif d'une langue.

La netiquette des canaux que nous avons sélectionnés pour constituer le corpus prévoit que les utilisateurs s'expriment en français. Elle ne restreint cependant pas la parole aux seuls francophones de France métropolitaine, bien que les buts explicites des canaux en question soient de rassembler des locuteurs géographiquement situés en France. Si les opérateurs des canaux sont chargés de veiller au respect de la netiquette, il est évident qu'avant une intervention de leur part justifiée par le non respect du code spécifié par la netiquette du canal, il faut qu'il y ait manquement à la règle. De plus, les opérateurs ne sont pas aussi intransigeants que ne le laisse entendre la netiquette ainsi que le montre l'exemple suivant dans lequel un locuteur intervient explicitement suite à un manquement à la netiquette :

Extrait du Corpus P1 *PBC: si ca ne derrange personne je vais pas faire la police

Nos corpus, tels qu'ils ont été recueillis et avant tout traitement ne sont donc pas homogènes du point de vue de la langue. Il faut noter cependant que, concernant le code, la netiquette est très largement respectée : la plupart des interventions sont effectivement en français, comme on le constatera à la section suivante.

4.1.1. Principes du codage

Un premier codage a donc dû être opéré sur les corpus qui permet de classer les interventions en différentes catégories, selon le code utilisé. Au premier abord, cette opération peut sembler triviale, nous verrons que ce n'est pas tout à fait le cas.

Nous avons classé les interventions en cinq catégories pour ce qui concerne la langue. Ces catégories ne se justifient que pour les traitements que nous nous étions proposés, et nous en présentons maintenant les principes.

En examinant de façon globale les corpus, nous avons remarqué une présence importante de l'anglais par rapport aux autres langues – hormis le français. Il nous a semblé donc important pour caractériser le corpus de marquer spécifiquement les interventions en anglais. Cette position se justifie aussi par le fait que les locuteurs francophones ont montré dans nos corpus une tendance à communiquer avec les participants qui utilisent l'anglais pour leurs interventions.

[70] ***Welcome to IRC – where the men are men, the women are men, and the boys are FBI agents. But some of the men are really women. – Enjoy! (en anglais dans le texte – notre traduction).***

en vertu de la loi du droit d'auteur.

Il est aisé de distinguer l'anglais du français mais beaucoup moins de déterminer s'il s'agit d'un anglais langue seconde ou d'un français langue seconde. Les études linguistiques mentionnent des *patterns* de production typiques de locuteurs de langue seconde pour une langue maternelle particulière aux plans phonologique, morphosyntaxique et sémantico-lexical et pragmatique, des maladresses, des "brutalités", mais ce n'est que sur un faisceau d'indices à occurrences régulières qu'on sera à même de décider si un locuteur utilise une langue qui n'est pas sa langue maternelle [71] . Pour l'anglais, nous n'avons fait la distinction que lorsque le locuteur s'exprime en français tout au long du corpus à l'exception de quelques interventions qui sont, par exemple, des réponses qu'il adresse à des locuteurs utilisant l'anglais. Pour le français [72] , nous avons repéré des locuteurs probablement français langue seconde : en observant toutes les interventions d'un locuteur particulier des indices linguistiques récurrents ont été repérés tels que les erreurs sur le genre des noms, le choix des pronoms, le choix lexical et des structures syntaxiques non conventionnelles.

Du point de vue inverse, aucune utilisation de *foreigner talk* [73] n'a été repérée dans le corpus de la part des locuteurs utilisant le français.

Les interventions qui sont mixtes du point de vue du code ont été classées dans la catégorie correspondant à la langue dominante pour un locuteur, les éléments du xénolecte étant marqués comme tels à l'intérieur même de l'intervention ainsi que l'illustre l'exemple suivant :

Extrait du Corpus F1 *PAD: t cvontent je vais avoir mlein de prive now@gb lol

Nous avons rangé dans une classe particulière toutes les interventions dans d'autres langues et qui ont été ignorées des utilisateurs francophones ou incomprises. Ainsi se trouvent dans cette classe quelques interventions en grec, en allemand,… et également un certain nombre d'interventions qui se sont présentées comme indécidables du point de vue de la langue (ratés de la communication). Dans d'autres cas les suites de caractères semblaient ne pouvoir appartenir à aucune langue (Cf. exemple ci-dessous).

Le rassemblement d'éléments aussi divers se justifie par le fait qu'ils sont ignorés des locuteurs francophones ou qu'ils ne suggèrent que des commentaires métalinguistiques tels que ceux des exemples ci-dessous.

Extrait du Corpus P3 *PAT:
'¼ŽÐƒT#[ƒrƒX,Æ,Íˆá,¢‰f'œ,Í,·,×,Ä–³#C#³,Å,·,Ì,Å#A#¶,Å#—,ÌŽq,Ì,·,×,Ä,ðŒ©,é,±,Æ,ª#o—^,Ü,·!!,à,

[71] Cf. Mayfield Tomokiyo (2000); (2001).

[72] Il s'agit de *code switching*. Le *code mixing* fait l'objet d'un traitement particulier.

[73] Terme usuel pour désigner le discours des locuteurs natifs tel qu'il est modifié en interaction avec des locuteurs non natifs. Les traits répertoriés comme caractéristiques sont divers et concernent tous les niveaux linguistiques (phonétique : débit plus lent, articulation plus marquée, plus grande utilisation des pauses et accents, utilisation d'un plus large éventail fréquentiel, production des formes complètes, évitement des contractions…; morphologie et syntaxe : production d'énoncés bien formés peu complexes et courts, morphologie régulière et réduite, marquage explicite des relations grammaticales ; lexique : utilisation d'items fréquents et non idiomatiques…; pragmatique : plus de vérifications de compréhension, plus de répétitions, plus de requêtes de confirmation…) cf. Long (1983).

en vertu de la loi du droit d'auteur.

%add: INDET *PAT: —cŽ™ƒvƒŒƒC#A#—Žq#Z#¶ƒvƒŒƒC#A#…'…ƒvƒŒƒC#AŠÅŒì•wƒvƒŒƒC"™,¢,ë,¢,ëÈ ƒRƒXƒ`ƒ…#[ƒ€,ð—pˆÓ,µ,Ä#A,©,í,¢,¢#—,ÌŽq,ª–ˆ"ú, ,È,½,ð,¨'Ò,¿,µ,Ä,¨,è,Ü,·#B %add: INDET *PAB: alors qu'es que ca? %add: ALL [74] *PAB: un op au primier qui arrive a casser le code ! %add: ALL *PAJ: J'laiEuMonCode PAB . %add: PAB *PAB: PAJ rien a voir ! %add: PAJ *PAS: c un flood texte ;p %add: PAB Extrait du Corpus F4 [75] *PCF: svara . %add: INDET *PCF: nu . %add: INDET *PCF: lite respekt om man få be . %add: INDET (6) *PCA: c'est pas le suedois la langue officielle de ce chan mais le français %add: PCF

Nous avons enfin marqué les interventions strictement spécifiques à l'IRC. En principe, les commandes des utilisateurs n'ont pas de contrepartie dans les lignes messages, toutefois il est possible de paramétrer les logiciels pour qu'ils le permettent ; la commande apparaît alors comme une ligne message et est immédiatement suivie de la ligne système correspondante. Ce fait donne naissance à des comportements ludiques particuliers, ainsi, sont rassemblées dans cette catégorie des interventions qui sont soit des commandes à un robot, soit des imitations de commandes, soit un code particulier utilisé pour s'adresser à l'ensemble des utilisateurs toutes langues confondues.

À l'issue de ce codage, qui nous permettra de nous intéresser spécifiquement au français, il nous est possible de donner un aperçu de la représentation de chaque code dans les corpus. Dans l'histogramme ci-dessous, le groupe 1 représente le nombre d'interventions indécidables du point de vue de la langue et/ou qui ont été ignorées des autres utilisateurs. Les groupes 2 et 3 rassemblent respectivement les interventions en anglais et en français. Le groupe 4 est constitué des interventions dans d'autres langues que le français ou l'anglais qui ont été intégrées à l'interaction. Enfin le groupe 5 rassemble les interventions qui sont des commandes ou pseudo-commandes informatiques.

[74] La ligne %add contient le code correspondant au destinataire du message lorsque ce dernier a pu être identifié. Les messages adressés à la cantonade sont marqués "ALL", tandis que ceux pour lequel aucun destinataire n'a pu être identifié sont marqués "INDET". Enfin, les destinataires secondaires sont notés entre parenthèses.

[75] Les chiffres entre parenthèses dans les extraits de corpus représentent le nombre de messages omis pour la présentation.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

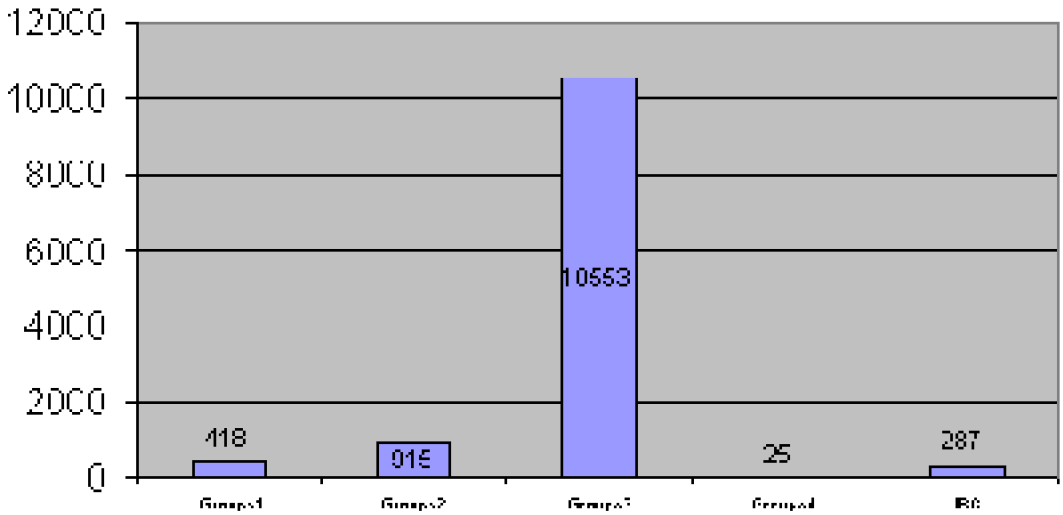

Figure (4-1) – Répartition des lignes messages et actions du corpus selon le code

On remarquera sans peine que les interventions autres que francophones sont peu importantes. Pour l'anglais, 20% des interventions proviennent de locuteurs qui utilisent par ailleurs le français, et 12% sont des interventions que nous qualifions de non originales puisqu'il s'agit de moules d'énoncés préconçus, dans lesquels des variables sont remplies automatiquement ou manuellement dont nous donnons ci-après quelques exemples, les lettres majuscules représentant les parties d'énoncé variables :

Énoncés préconçus X is back. gone N min Y s [76] ***X slaps Y around a bit with a large trout*** [77]

Ces "moules" sont disponibles dans certains logiciels clients et se répètent par endroits tout au long du corpus.

Pour le français, 1% des interventions a été marqué comme du français langue seconde selon les critères évoqués plus haut.

Quant à la forme du codage dont nous venons d'exposer les principes, elle associe à chaque intervention la catégorie qui lui correspond. Le codage des locuteurs et celui de la langue sont donc indépendants.

4.1.2. Notes concernant le français dans les corpus

Le codage étant effectué il nous permet de sélectionner pour notre étude les lignes messages qui nous intéressent, c'est-à-dire celles qui sont en français, sans occulter complètement les autres lignes et en autorisant le cas échéant la mise en relation de lignes de types différents du point de vue du code.

Depuis le début de ce chapitre, nous parlons du français comme s'il s'agissait d'une

[76] ***X est de retour. Parti N minutes, Y secondes.***

[77] ***X frappe Y avec une grosse truite.***

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

évidence théorique. Le français que nous trouvons dans nos corpus est particulier à plusieurs points de vue.

D'après le SIL [78] , 77 millions de personnes parlent le français comme langue maternelle à travers le monde, dont 51 millions en France où une douzaine de parlers régionaux sont dénombrés [79] , les quelques 26 millions de locuteurs restants étant répartis dans 53 autres pays. Des représentants de ces communautés distantes peuvent bien entendu probablement se trouver parmi les participants de nos corpus puisque Internet ne connaît pas de véritables frontières géographiques. Les barrières [80] que connaît Internet sont économiques, culturelles ou techniques. Nous avons vu au chapitre premier les difficultés que représente l'évaluation du paysage linguistique sur Internet. La répartition des locuteurs francophones (de langue maternelle) sur Internet d'après les chiffres de Global Reach [81] se fait comme suit.

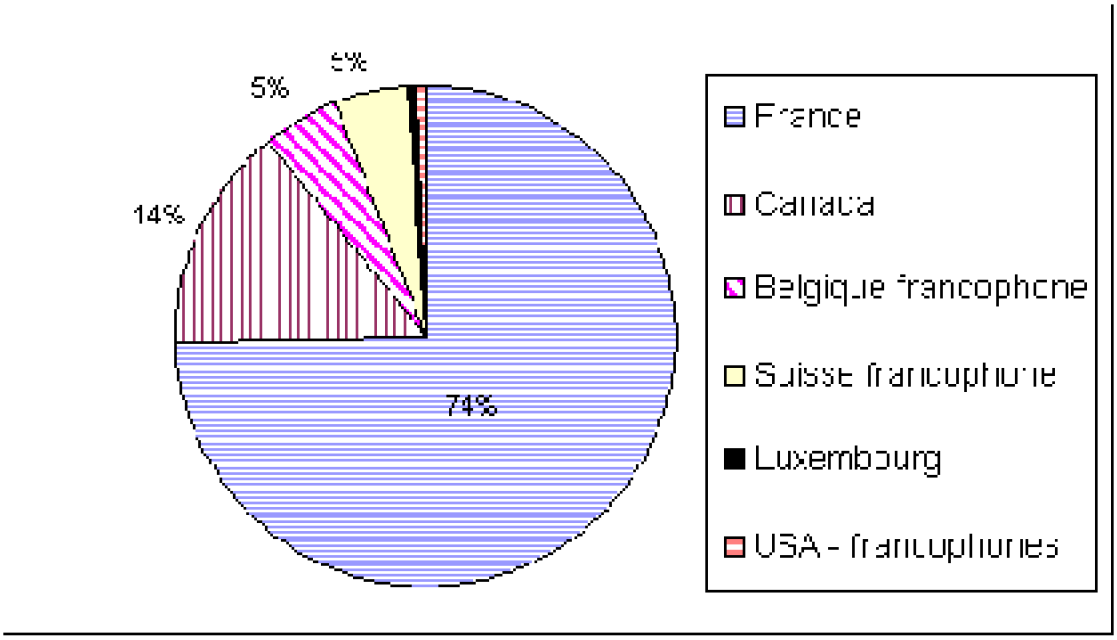

Figure (4-2) – Nombre d'internautes francophones

La répartition est, dans les grandes lignes, similaire à la celle des locuteurs francophones par pays que donne le SIL International [82] :

78 Ethnologue : The languages of the world http://www.ethnologue.com/web.asp

79 Carton, Rossi, Ausseterre & Léon (1983).

80 On parle aujourd'hui de fracture numérique.

81 Global Reach http://www.glreach.com/globstats/

82 Ethnologue : The languages of the world http://www.ethnologue.com/web.asp

en vertu de la loi du droit d'auteur.

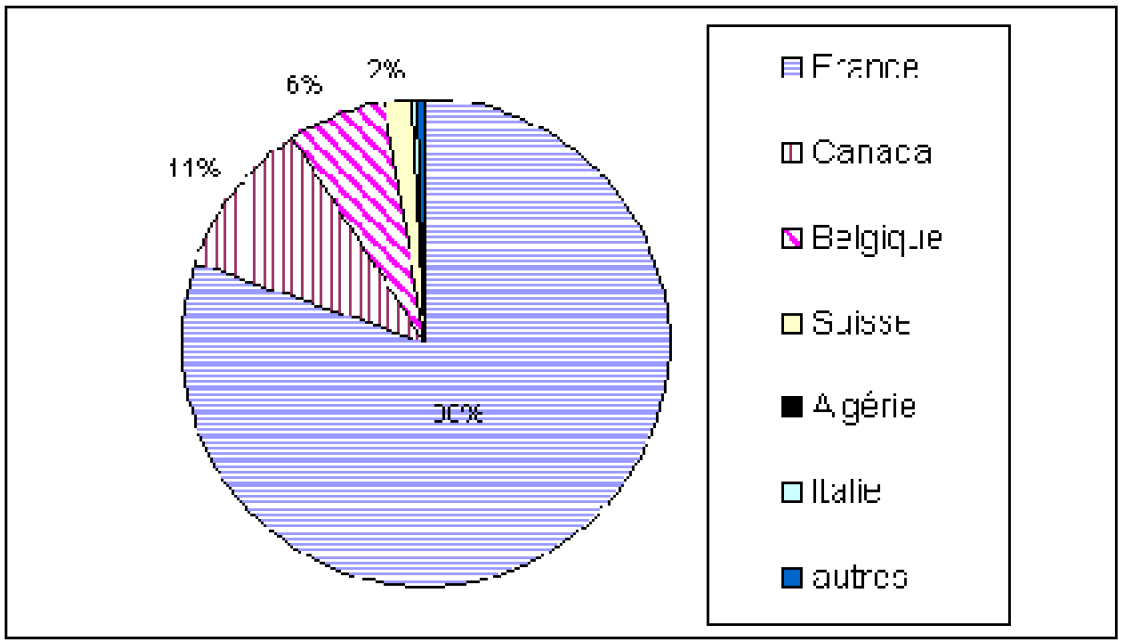

Figure (4-3) – Nombre de locuteurs francophones

Les buts exprimés des canaux étudiés sont de rassembler des personnes francophones physiquement en France cependant, la relative faible part de canaux IRC francophones laisse supposer que des locuteurs francophones hors de France, voire des locuteurs d'une langue peu parlée dans le monde ayant le français en langue seconde utilisent ces canaux.

On peut aussi supposer que les locuteurs présents dans les IRC, d'une part sont sélectionnés parmi la population selon des critères socioéconomiques et socioculturels et d'autre part ont un certain intérêt pour la chose informatique et une certaine culture en ce domaine, il reste qu'ils appartiennent nécessairement aussi à des communautés linguistiques géographiquement définies. Nous nous attendons donc à trouver dans nos corpus des indices de variations diatopiques, diastratiques et diaphasiques [83] . Les variables sociales [84] traditionnellement prises en compte dans les études de la variation ne peuvent être établies. Les locuteurs donnent, certes, des indications à propos d'eux-mêmes au cours des conversations, qui peuvent être source de diversité. On doit évidemment se garder d'accorder trop de crédit à ces dernières, et ce, en raison de la spécificité de la communication IRC mentionnée au début de cette section. Néanmoins l'âge avoué des locuteurs oscille entre 16 et 39 ans, quelques-uns seulement se déclarant âgés de plus de 30 ans. Et ceci est corroboré par le fait que les conversations font assez souvent référence à des situations scolaires (examens, devoirs, cours…). La position géographique semble être pour la majorité la France Métropolitaine, quelques locuteurs se déclarant être des Québécois en France. La situation inverse est aussi affirmée ou peut-être inférée de la conversation. Certains encore se déclarent français hors de France Métropolitaine, étant enfants d'expatriés ou en stage à l'étranger. Enfin, un certain nombre

[83] Variations régionales, sociales, ou en fonction de situations discursives.

[84] Milroy (1987)

en vertu de la loi du droit d'auteur.

de locuteurs peuvent être identifiés comme Québécois au Québec. Les autres situations sont très rarement revendiquées par les locuteurs.

Les éventuels facteurs de variations esquissés ici pour nos corpus sont sans doute affaiblis par le medium, qui d'autre part engendre un certain nombre de difficultés de traitement, ainsi que nous le montrons dans la section suivante.

4.2. Standardisation et lemmatisation

Ainsi que nous le verrons en particulier au chapitre 5, les corpus que nous avons recueillis présentent un écart important à la norme de l'écrit. Ils comprennent des graphies, unités et découpages d'unités particulières ; et des utilisations de la casse des caractères, des signes de ponctuation et de groupes de caractères, très éloignées de la norme. Les quelques exemples rassemblés ci-dessous permettent de se faire une première idée de ces faits. Notons que la plupart des interventions présente au moins un de ces phénomènes. Dans les exemples ci-après, les points d'intérêt sont mis en gras.

Extrait du Corpus P6 *PC6: le vendredi midi c autorisé:) Extrait du Corpus F3 *PCO: il parait que les joueurs ont eu moins d egueulle apres le match contre l'Espagne Extrait du Corpus F7 *PKV: §å£µt £ë§ Fî££ë§ Extrait du Corpus P6 *PI7: ca y est t'est defouler.?? Extrait du Corpus P7 *PBA: PBB hey hohein, chacun son tour Extrait du Corpus F5 *PCZ: aaaaaaaaaattttttchoum Extrait du Corpus F5 *PBI: PARLEZZZ MOIIIIII PLZ Extrait du Corpus P7 *PAG: j suis dja en vac moa Extrait du Corpus P8 *PAN: j ai les 2 morceuax en surimpression lol

Afin de rendre les corpus plus lisibles, de faciliter les traitements automatiques et notamment afin de permettre la lemmatisation, il nous a fallu procéder à la normalisation des corpus. Cette procédure permet également de mesurer l'écart à la norme et de l'analyser qualitativement. Ce travail de normalisation ne peut être réalisé par des procédures automatiques le code n'ayant pas été étudié systématiquement auparavant et présentant des irrégularités importantes. Certes, les logiciels de traitement automatique prennent en compte une partie des usages écartés de la norme rencontrés dans les corpus écrits telles que les hésitations sur les tirets (ex. *plateforme, plate forme* ou *plate-forme*), les variantes orthographiques et morphologiques (accents, majuscules, cédilles, pluriels des formes en **-** *als -aux*…), les erreurs de frappe fréquentes (inversion de lettres, confusion de lettres proches sur le clavier…) [85] . Mais la prise en compte de ces phénomènes doit être augmentée des phénomènes spécifiques à un type de document numérique. Ainsi, les corpus issus de la numérisation, par exemple, et traités par reconnaissance optique de caractères présentent les phénomènes liés à l'activité humaine mentionnés ci-dessus, ainsi que des phénomènes spécifiques liés à la procédure technique imparfaite de reconnaissance optique de caractères.

Nous avons donc effectué manuellement sur l'ensemble des corpus ce travail de mise en conformité et si nous avons généralisé certains traitements de façon

[85] Fluhr *et al.* (1994), Fluhr (1997), Chenon *et al.* (1998)

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

automatique, ce fut toujours accompagné d'une vérification.

4.2.1. Standardisation

Lorsqu'il s'agit de traiter un corpus oral, le linguiste se rend nécessairement à l'activité de transcription, c'est-à-dire la reproduction transformée des données acoustiques en données écrites. La transcription permet de passer d'un mode de représentation à un autre qui ne comporte pas les mêmes dimensions. L'avantage de la scripturalité est de rendre l'information globalement accessible, mais elle implique des choix, et plusieurs types de transcription peuvent être envisagés. Quel que soit le type retenu (orthographique, phonétique…), le chercheur se trouve devant les difficultés, liées à l'objet lui-même, que sont ses multiples dimensions. Selon les objectifs de l'étude, le système de transcription prendra en compte une sélection de ces dimensions. L'utilisation de la transcription phonétique est entreprise essentiellement pour les études qui se consacrent à des niveaux d'analyse concernant la matière phonique mais la plupart des études sur de grands corpus oraux lorsque les objectifs sont lexicaux, syntaxiques, discursifs, etc. se reportent à une transcription orthographique qui reçoit différents aménagements selon les besoins de l'étude. Le modèle de transcription [86] de l'équipe du GARS [87] à l'université de Provence est adapté aux recherches sur la syntaxe et à l'analyse discursive. Il utilise une transcription orthographique quand les données le permettent [88] .

Cette transcription est réalisée en orthographe standard, sans les "trucages orthographiques" qu'utilisent les écrivains [89] pour produire un effet d'oral.

Les grands dictionnaires servent de référence pour les mots de la langue, les noms propres, les interjections et les onomatopées. DELIC – Corpus du français parlé Conventions de transcription. [90]

Dans les cas que le contexte ne permet pas d'éclaircir, certains éléments sont mis entre parenthèses: *il(s) répare(nt) leur(s) maison(s).* Lorsque deux transcriptions sont également plausibles, les deux sont conservées avec des marques particulières. Les notations phonétiques quand elles interviennent se produisent en insérant une note de bas de page. La ponctuation et les majuscules de début de phrases ne sont pas ajoutées.

Nos corpus sont de type écrit, cependant ils s'écartent de la norme écrite et présentent des phénomènes fréquemment considérés comme typiques de l'oral. Ils ne

[86] Cf. site web du groupe DEscription Linguistique Informatisée sur Corpus http://www.up.univ-mrs.fr/delic/corpus/index.html

[87] Groupe Aixois de Recherches en Syntaxe

[88] Le parler des jeunes enfants, de locuteurs affectés par des troubles du langage, de locuteurs maîtrisant mal le français reçoit préférentiellement des transcriptions phonétiques. Cf. Blanche-Benveniste (1997:28-29).

[89] Blanche-Benveniste (1991).

[90] ***Cf. site Web du groupe DEscription Linguistique Informatisée sur Corpus http://www.up.univ-mrs.fr/delic/corpus/index.html***

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

pouvaient donc être traités d'une façon conventionnelle pour les raisons mentionnées en introduction de ce chapitre. Notre but étant, d'une part de qualifier l'écart à la norme, et d'autre part d'étudier les moyens lexicaux et syntaxiques, il nous a fallu trouver un système de transcription qui permettent de mettre en parallèle les formes normées aux formes déviantes (sur le plan graphique seulement). Nous avons adapté les principes de transcription du GARS à nos corpus. Ce qui nous intéresse dans ce corpus, ce sont ses propriétés lexicales, syntaxiques, morphosyntaxiques et discursives. Ainsi que le préconise le GARS, notre normalisation – puisqu'il ne s'agit pas d'une transcription au sens strict du terme – consiste en une simple standardisation orthographique, sans toucher à la ponctuation, en utilisant les grands dictionnaires comme référence ainsi qu'ils sont utilisés par les chercheurs du GARS. Toutefois nos corpus présentent des néologismes, des emprunts, des noms propres, des interjections [91] , des sigles et des onomatopées qui ne se trouvent pas dans les dictionnaires, ainsi que les quelques exemples ci-dessous l'illustrent sommairement.

Extrait du Corpus F1 *PCJ: aaaaaaaarf PAD. change de nick où je te vire mdr...
Extrait du Corpus F5 *PAC: hummmmmmmmmmmmmmmmmmmmmm qui parle de sirop d'érable là :P j'y vais dans quelques jours miam miam menoummmmmmmmmmm :P

Il nous fallait pouvoir, dans ces cas aussi, repérer les différentes réalisations d'une même unité. Nous avons, pour les noms propres de personnes (et surnoms), rétabli la forme par laquelle les personnes désignées se dénomment elles-mêmes, et lorsque cela n'était pas possible, nous avons adopté la forme la plus usitée ou la forme la plus simple, et en dernier recours la première forme rencontrée [92] . Pour ce qui concerne les néologismes, emprunts, régionalismes, idiosyncrasies, nous avons standardisé à la forme la plus fréquente s'il s'en dégageait une, sinon à celle répondant le mieux à la norme orthographique. Les interjections et onomatopées en très grand nombre dans notre corpus qui n'apparaissent pas dans les dictionnaires ont subi le même traitement.

Ainsi, pour les exemples ci-dessus, on trouve *nic*, *nick* et nous avons standardisé *nick* puisqu'il s'agit de la forme la plus fréquente ; *mdr*, *mdrrr*, *mdrr…* ont été standardisés en *mdr* ; *menoummmmmmmmmmm* a été standardisé *menoum* puisque nous avons pu dégager un phénomène de redoublement de caractères également sur des unités standards (cf. chapitre 5).

Ce codage a été effectué en sorte que l'énoncé effectif soit toujours déductible de la forme codée. De cette façon, les corpus à la suite de ces opérations contiennent à la fois l'énoncé effectif et sa normalisation.

À l'issue de ce traitement, nous avons récupéré dans une table toutes les opérations de normalisation et les avons analysées de façon à obtenir une information sur l'écart à la norme orthographique que présente le corpus et dont nous discuterons au chapitre 5.

[91] Il ne s'agit pas seulement de "particularités", puisque certaines interjections fréquentes ne se trouvent pas dans les dictionnaires.

[92] Notons que cette information n'est pas présentée dans le corps de la thèse pour les noms propres de participants (surnoms), puisqu'il a été procédé à l'anonymisation des corpus – cf. section 3 de ce chapitre.

en vertu de la loi du droit d'auteur.

4.2.2. Lemmatisation

La lemmatisation d'un corpus doit permettre un certain nombre d'analyses linguistiques. Elle autorise essentiellement le repérage de catégories morphosyntaxiques qui sont utiles à la fois pour les analyses du lexique (chapitre 7) et de la syntaxe (chapitre 8).

Le terme technique de lemmatisation désigne, selon Brunet (2000:1)

"L'opération de regroupement qui rassemble les formes différentes appartenant au même vocable – ce que font précisément les auteurs de dictionnaires quand ils établissent leur nomenclature".

Cette méthode utilisée notamment en statistique lexicale [93] nécessite qu'on s'interroge sur les unités en question. On distingue en effet différents rangs d'analyse en fonction des unités traitées. La lemmatisation est un procédé utilisé en lexicométrie pour des besoins d'analyse sémantique. Elle ne s'occupe ni de la description des phrases, ni des unités de rang supérieur. Se présentent donc comme candidats au statut d'unité, le mot et le morphème.

Le mot a longtemps été considéré comme un donné, une évidence, si on examine la chaîne parlée [94] , les mots sont difficilement délimitables. Le mot est avant tout une unité graphique : chaîne de caractères séparée des autres par des espaces ou ponctuations. Mais la graphie elle-même pêche, les mots composés ne répondant pas à la définition graphique qu'on pourrait naïvement donner du mot. On propose parfois une distinction terminologique ramenant le mot à son statut d'unité graphique, et créant de nouveaux termes tels que "lexie", "synapsie", "synthème",... pour désigner l'unité lexicale [95] . Le morphème en tant que plus petite unité linguistique à deux faces, se présente comme un meilleur candidat au statut d'unité lexicale que le mot, unité complexe, unité significative non minimale, analysable en unités significatives plus petites. L'objection formulée à l'encontre de l'utilisation du morphème comme unité lexicale de base, outre qu'il s'accorde bien moins à l'intuition du locuteur, est que la composition que forme un mot n'est pas toujours le résultat de la simple addition des sèmes des morphèmes composants. Charles Muller (1992:3) rappelle que "le terme de mots, en linguistique, n'a pas de définition satisfaisante, et cette entité, assez claire pour le sens commun, est scientifiquement suspecte". Il précise que "rien n'interdit de prendre d'autres bases que le mot ou le vocable, si l'on en trouve de moins suspectes, mais que toute option de ce genre comporte des difficultés spécifiques sur le plan pratique" [96] . Il convient donc de proposer une définition pour le mot qui fixe clairement l'objet envisagé et que Muller donne comme "unité graphique, séparée des unités voisines par un blanc ou une ponctuation ; elle se

[93] Et dans nombre de domaines de Traitement Automatique du Langage Naturel (TALN), cf. Carré *et al.* (1991).

[94] La forme linéaire du langage dérive de son caractère vocal puisque les énoncés s'inscrivent dans le temps, les sons se succèdent.

[95] Cf. Rey (1977:15).

[96] Muller (1992:11)

en vertu de la loi du droit d'auteur.

retrouve certes dans le langage oral dont la graphie n'est qu'une représentation seconde ; mais la tradition graphique, pour les idiomes qui en ont une fait apparaître en général un découpage de la phrase (la "première articulation") qui dans le langage oral est moins conscient." [97] Cette première approximation est immédiatement remise en cause par des phénomènes indiscutables qui sont qu'une unité graphique peut correspondre à plusieurs unités de lexique et inversement qu'une unité de lexique peut se composer de plusieurs unités graphiques. Un ensemble de règles doit être établi permettant de décider de la délimitation des unités : la norme lexicologique. Ce qui importe selon Muller [98] , c'est moins la qualité scientifique des décisions prises que la cohérence et l'application systématique des règles. L'ensemble des questions qui doivent être posées pour établir la norme lexicologique – formes contractes, formes composées, cas de soudure graphique ou trait d'union, sigles et abréviations – doit donc tenir compte de ces deux principes.

La lemmatisation convient en lexicométrie dans la mesure où les accidents de la morphologie flexionnelle ne modifient pas de façon fondamentale le sens des unités dans le texte sauf rares exceptions de formes telles que *lettres, forces, règles*, *bagues*,… En revanche, si l'objet d'étude est la syntaxe ou la morphologie, il est nécessaire de disposer de la forme occurrente du mot, ou d'une lemmatisation augmentée d'un codage de la flexion. Cette option permet non seulement d'avoir accès dans des corpus à des possibilités d'étude de la fréquence d'unités lexicales mais aussi, ramenant à la même forme des flexions différentes parce que conditionnées par la classe d'un mot, de s'intéresser à la syntaxe et à la morphologie.

La lemmatisation au sens strict du terme, si elle prend en compte les homographes du type *il **est** à l' **est** de Paris ; Sophie a **été** fatiguée cet **été*** ; ne résout pas le problème des lemmes homographes du type *la pie **vole ;** ce sont des **jumelles*** .

Selon ce qu'on se propose d'étudier, les exigences de lemmatisation sont différentes. La forme la plus simple consiste à ramener la forme occurrente à sa forme de base. Pour notre part, nous l'avons accompagnée d'un codage de la flexion et de l'étiquette de catégorie grammaticale. Cette opération a été réalisée par des procédures semi-automatiques avec contrôle, selon le système proposé pour CHILDES [99] et les spécifications de Parisse & Lenormand (2000) auxquelles nous avons apporté quelques aménagements. Elle permet d'obtenir des données pour l'analyse du lexique, de la morphologie et de la syntaxe. Nous préciserons les choix auxquels nous avons opéré aux chapitres correspondants aux analyses mentionnées, soit les chapitres 7 et 8.

4.3. Anonymisation des corpus

97 Muller (1992: 4-5).

98 Muller (1992).

99 Cf. MacWhinney (1995).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Un gros travail de traitement des données a été nécessaire pour réaliser l'anonymisation des corpus. On peut se demander à quel titre ces opérations sur les corpus étaient indispensables puisqu'il s'agit de conversations dans des canaux publics et que le dispositif technique lui-même permet l'enregistrement des conversations [100] . On néglige trop souvent les questions juridiques associées à la constitution de corpus, qui concernent le droit d'accès et de rectification des données concernant la personne (loi Informatique et Liberté), le droit moral, le respect de la propriété intellectuelle [101] . Sans même considérer les conséquences dramatiques du manquement à ces règles, nous n'aurions su mener notre étude sans assurer la plus grande considération à ceux qui l'ont permise.

4.3.1. Justifications

Tout d'abord, le fait que quelque chose soit enregistrable, ou réplicable ne dit rien sur la légalité de l'opération. Le droit de la propriété intellectuelle et le droit au respect des personnes et de la vie privée sont en constante évolution depuis l'avènement de l'informatique et des nouvelles technologies de l'information et de la communication, ils comportent cependant un certain nombre de principes auxquels les chercheurs doivent prêter attention [102] .

Le caractère public des canaux est en principe connu des utilisateurs, toutefois l'usage de surnoms plutôt que de noms réels semble indiquer une réticence à la reconnaissance de l'identité des utilisateurs. Les surnoms ne garantissent pourtant pas l'anonymat puisque pendant son temps de connexion un utilisateur peut obtenir les adresses IP [103] des autres utilisateurs connectés. Enfin, certains surnoms sont "déposés" par les utilisateurs : un serveur permet d'enregistrer un surnom et d'empêcher les autres utilisateurs d'en faire usage.

Il nous a donc semblé déontologiquement indispensable de procéder à l'anonymisation des corpus.

4.3.2. Procédure

Une procédure d'anonymisation consiste à rendre impossible l'identification des individus. Dans le cas de données audio ou audiovisuelles, on gomme les caractéristiques individuelles visuelles, sonores ou linguistiques susceptibles d'activer la reconnaissance d'un individu. Dans le cas de l'écrit dactylographié, on gomme des informations linguistiques. Malheureusement, il ne s'agit pas dans notre cas d'associer simplement un

[100] Cf. Chapitre 2.

[101] Cf. Pujol (1993) et le site Web de la Commission Nationale de l'Informatique et des Libertés http://www.cnil.fr.

[102] Cf. polémiques au sujet du programme ECHELON.

[103] Une adresse IP permet d'identifier une machine connectée au réseau de façon plus ou moins stable. Cf. Steyer (1998), Tanenbaum (1997:438-440).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

code à un surnom puis de remplacer dans les corpus le surnom par le code. Si à un moment T il ne peut correspondre qu'un utilisateur à un surnom, sur une session un utilisateur peut utiliser plusieurs surnoms. En effet, au cours de la connexion, les utilisateurs ont la possibilité de changer de surnom et ils utilisent cette possibilité fréquemment ainsi que le montre le tableau suivant (4-1). En moyenne, il intervient près d'un changement de nom par minute sur l'ensemble du corpus.

Tableau (4-1) – Changements de surnoms, propriétés générales [104]

Corpus	Nb. c.d.n.	c.d.n./min
F1	92	2,09
F2	81	1,88
F3	269	2,17
F4	163	0,71
F5	213	1,52
F6	231	1,53
F7	308	0,75
F8	129	0,83
P1	39	0,81
P2	38	0,83
P3	106	0,86
P4	51	0,22
P5	39	0,26
P6	141	0,34
P7	96	0,26
P8	49	0,52
P9	64	0,41

Nous avons établi une procédure automatique prenant en compte ce phénomène et associant un code à chaque série de surnoms associée à un participant actif. Il ne nous était en effet pas nécessaire de coder les surnoms des utilisateurs inactifs. À titre d'appréciation quantitative, nous donnons le tableau résumé des changements de noms survenus dans chaque corpus, ce qui permet de se rendre compte de leur importance. Cette procédure permet en même temps d'évaluer le nombre de participants actifs différents pour chaque session.

Tableau (4-2) – Changements de surnoms, propriétés détaillées

[104] **c.d.n. : changements de surnoms**

en vertu de la loi du droit d'auteur.

Corpus	Nb. de participants actifs	Nb. de c.d.n. des participants actifs	Nb. de participants ayant changé de surnom	Nb. de c.d.n. maximum par participant actif
F1	65	15	9	4
F2	62	12	10	2
F3	136	43	26	5
F4	120	27	21	2
F5	116	30	19	7
F6	139	56	35	6
F7	302	71	44	7
F8	123	27	20	4
P1	30	5	3	3
P2	27	7	5	2
P3	62	33	16	5
P4	37	16	9	4
P5	36	12	7	3
P6	103	47	20	9
P7	68	24	14	5
P8	44	11	7	4
P9	50	27	13	8

Nous donnons l'algorithme pour cette procédure en annexe. Il faut toutefois signaler un phénomène que nous ne prenons pas en compte dans ce traitement automatique pour la raison qu'il n'a pas de trace récupérable dans les *logs* et qu'il semble très peu fréquent : il s'agit du phénomène de clonage c'est-à-dire lorsqu'un même utilisateur se connecte sous des surnoms différents simultanément. Cette pratique est réprouvée par la netiquette et souvent détectée par les opérateurs des canaux qui excluent alors ces utilisateurs indisciplinés.

Extrait du Corpus P8

*PAW: PBI les pubbeurs miteux que j'ai dégagé c tes copains

%add: PBI

(1)

*PBI: Il faut bien que j'en ai, non ?

%add: PAW

*PAG: ptdr clonage powa

%add: PBI, PAW,

(1)

*PAG: ses poto ses ses clone!

%add: PAW

en vertu de la loi du droit d'auteur.

Au terme de ce traitement, tous les surnoms des participants apparaissent donc dans le corpus sous la forme d'un code de 3 caractères.

4.3.3. Pertes

L'anonymisation des corpus engendre une perte d'information. En effet, les surnoms sont souvent utilisés pour signifier et n'ont pas une simple fonction de dénomination des individus – qui par le fait que les individus choisissent eux-mêmes leur surnom porte déjà de l'information [105] .

Par exemple un utilisateur nommé X se renommera X-DEJEUNER, X-bouffe, ou Xfume, Xboulot…

Les surnoms permettent donc aux utilisateurs, par le biais de changement de surnom au cours de la connexion, de communiquer des informations à propos de leur activité, de signaler le fait qu'ils sont éloignés de leur écran pour un instant… Dans des cas plus rares, le changement de surnom intervient en relation directe avec la thématique de conversation. Ces cas ont été marqués dans le corpus. Nous donnons un exemple ci-après, dans lequel des utilisateurs simulent une vente aux enchères.

Extrait du Corpus P3 *** XXXXX0 is now known as une *PZZ: une fois *PAB: trop tard :P *** XXXXX1 has joined #YYYYY *PAS: j'ai des verres anti reflet avec verre affine *** une is now known as deux *PAB: on peut pas revenir en arrieur *PZZ: deux fois *PAS: et une monture en titane... *PAB: 800 une foix (1) *** deux is now known as Trois *** XXXXX2 has quit IRC (Quit: ¤) *PZZ: Trois fois, adjugé vendu

Enfin, il arrive que le surnom lui-même fasse le prétexte de la discussion, ce que nous avons aussi marqué dans les corpus.

Extrait du Corpus P3 *PAK: PCC, quel drôle 2 nick *PBZ: ces koi se nick chelou ? *PBZ: grave (1) *PAS: ouhais franchement yen a ont la cheville qui explose (2) *PAK: Ouai, C comme s'appeller PAS, ça ne se fait pas

La procédure automatique produit donc pour chaque corpus une table de correspondance entre les surnoms réels et les codes calculés de sorte qu'il n'y ait pas de doublons et que les surnoms correspondant à un utilisateur unique reçoivent le même code. Un code particulier a été attribué pour les lignes d'action de façon à être en mesure de les repérer toujours. Pour chaque corpus, nous avons procédé au remplacement des surnoms réels par les codes calculés. Pour que les fichiers ainsi conçus soient compatibles avec le format CHILDES [106] , la contrainte était que les codes soient de trois caractères, ce qui autorisait 46 656 codes différents. Toutefois pour des raisons techniques, nous avons dû réserver le premier caractère ce qui n'autorise plus que 1 296 codes ce qui est en dessous du nombre de participants répertorié dans nos corpus [107] . Les codes ne sont

[105] Certains surnoms précisent le sexe du locuteur, son âge, son statut civil, des caractéristiques physiques… D'autres surnoms sont des énoncés, des titres de film, des noms de divinités, de personnages célèbres, de héros…

[106] Format de fichier choisi pour la présente étude

[107] Les corpus utilisés effectivement pour les présentes analyses comptabilisent 1517 participants.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

donc uniques qu'à l'intérieur d'un corpus donné.

L'anonymisation des corpus a été réalisée, ainsi que nous venons de le spécifier, de façon à ce qu'elle soit compatible avec le format de fichier imposé par le logiciel de traitement de corpus que nous avions sélectionné : CHILDES. La section suivante propose les raisons de notre choix et esquisse les caractéristiques de ce programme.

4.4. Logiciel de traitement des corpus

La forte croissance des industries de la langue a engendré de nombreux programmes de traitement automatique des langues naturelles. Ces derniers ont de nombreuses applications [108] : reconnaissance et synthèse vocale, constitution de thésaurus et extraction de terminologie, interrogation de bases de données multilingues, traduction automatique, synthèse automatique, dialogue homme machine, identification des langues… Ces applications nécessitent des analyses à tous les niveaux : acoustique, phonétique, phonologique, morphologique, syntaxique, sémantique et pragmatique. Tous ces outils, certes imparfaits, auquel le linguiste contribue, sont en grande partie inutilisables pour lui, et ce, parce que les objectifs de ces applications sont grandement différents des objectifs du linguiste. À ce jour, quantité de procédures, même les plus élémentaires du traitement de corpus, ne sont pas disponibles pour le linguiste qui ne travaille pas sur l'anglais. Nous avons pour notre part développé un certain nombre de procédures permettant le traitement de quelques caractéristiques de nos corpus ainsi que nous l'avons mentionné par exemple pour l'anonymisation, ou pour le repérage des différents types de lignes. La mise au point de ces procédures nécessite un investissement temporel important qui se justifie par la réutilisabilité des produits, nous les mettons donc à la disposition de la communauté.

À la lumière de la présentation de nos données au chapitre précédent, on se persuadera que le corpus dont nous disposons n'a la forme d'aucun corpus habituel.

Des logiciels non dédiés à l'analyse linguistique de corpus, tels que les logiciels de traitement de texte, ou plus simplement des éditeurs de texte permettent d'effectuer un grand nombre d'opérations sur des corpus écrits ou transcrits (recherches d'items, comptages…). La possibilité de créer des "sous-programmes" augmente considérablement leurs potentialités. Il est tout à fait possible de constituer son propre système de codage et de développer des macros pour le traiter.

L'utilisation de corpus volumineux nécessite le recours à des procédures de traitement automatique, qui, si elles ne sont pas toujours complètement fiables ont au moins l'intérêt d'appliquer des principes de manière uniforme sur un grand ensemble de données, et il suffit alors pour pallier les faiblesses des traitements automatiques d'appliquer des techniques de vérification ou d'utiliser des procédures de traitement automatique contrôlées.

[108] Pour une présentation détaillée, se reporter à FUCHS *et al.* (1993) et CARRÉ *et al.* (1991)

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Afin d'être en mesure de donner accès à ce corpus à d'autres chercheurs nous avons opté, afin de mener les traitements analytiques, pour la solution d'un logiciel de traitement dont la philosophie de distribution des données correspondait à la nôtre, et dont l'utilisation soit assez répandue. Il s'agit du logiciel CHILDES, conçu à l'origine pour l'uniformisation et la mise en commun des corpus destinés à l'étude de l'acquisition du langage chez l'enfant. Les chercheurs en acquisition du langage à l'institut de psycholinguistique Max Planck de Nijmegen en 1981 se sont concertés pour la création d'un système d'archivage des transcriptions qui a abouti grâce à un financement de la fondation MacArthur, au développement au sein de l'université Carnegie Mellon d'un système d'échange de données en acquisition du langage chez l'enfant – Child Language Data Exchange System – piloté par Brian MacWhinney et Catherine Snow.

Le système CHILDES [109] a trois composants principaux : un système de transcription et de codage (CHAT), un ensemble de programmes d'analyse automatique (CLAN) et la base de données des transcriptions mises à la disposition de la communauté par les chercheurs qui utilisent le système CHILDES. En effet, s'il est libre de droit, ce système fonctionne sur le principe de l'échange. Utiliser le système revient à accepter de partager ses transcriptions avec la communauté à l'issue des analyses.

Plusieurs arguments s'opposaient à l'utilisation de ce système pour le traitement de nos données. Tout d'abord, le fait que les buts premiers de ce système sont l'étude et le partage de données en acquisition. Cet argument est réfuté parce que l'étude de l'acquisition du langage chez l'enfant doit être étalonnée par des échantillons de données sur le langage tout au long de la vie d'une part, et d'autre part les corpus en acquisition sont souvent des interactions entre adulte et enfant. Rien ne s'oppose donc à son utilisation sur des corpus de locuteurs adultes. D'autre part, le système est prévu avant tout pour l'étude de corpus oraux et nos corpus sont de type écrit cependant les chercheurs en acquisition l'utilisent également pour l'étude de l'acquisition du système de l'écrit. Nous mettons en avant aussi la volonté exprimée d'extension de ce système à d'autres types d'analyses que les études en acquisition du langage chez l'enfant. Trois chapitres de la documentation du logiciel se rapportent à ce sujet et concernent notamment la transcription pour l'analyse conversationnelle – développée sur le modèle d'un système conçu par Sacks, Schlegloff et Jefferson (1974) –, la transcription des interactions en langues signées, et le langage écrit. De plus le logiciel convient pour l'analyse dans différentes langues. Il est remarquable, encore, que le système soit assez souple pour permettre des adaptations diverses.

Pour être en mesure d'analyser nos corpus, nous avons choisi de nous en remettre à l'utilisation d'un logiciel disponible et utilisé dans la communauté internationale plutôt que de développer des solutions logicielles *ad hoc*, qui empêcheraient le partage des données. Nous avons de plus veillé à conserver une copie des corpus bruts et vérifié que les traitements effectués sur le corpus pour la mise en conformité avec le format CHILDES soient réversibles automatiquement.

Le format CHILDES est peu contraignant et permet de développer des systèmes de

[109] Les documentations du système CHILDES sont disponibles entre autres sur le site de l'université Carnegie Mellon à l'adresse suivante : http://childes.psy.cmu.edu/

en vertu de la loi du droit d'auteur.

codages propres. Chaque fois que cela était possible, nous avons opté pour la solution proposée par le système CHILDES et dans le cas contraire, nous avons implémenté nos propres systèmes de codage. Ainsi, les frimousses et la ponctuation existante par exemple ont reçu un traitement particulier puisqu'ils n'étaient pas prévus par le logiciel.

Un fichier CHILDES comporte deux parties : un bloc d'informations sur le contenu du fichier et le corpus proprement dit. Chaque session pour chaque canal fait pour nous l'objet d'un fichier CHILDES. Le bloc d'informations d'en-tête de ces fichiers contient les informations permettant de définir le corpus (session, canal, nom des participants) ainsi que des informations méta analytiques (nom de l'opérateur, date de traitement du fichier…).

La partie réservée au corpus proprement dit fait l'objet d'une structuration particulière. Ce qui a été effectivement produit par un locuteur est inscrit sur une ligne à laquelle est associée un identifiant pour le locuteur (série de trois lettres majuscules précédée d'un astérisque).

Exemple : *PCK: salut .

Sur cette ligne, appelée ligne principale, peuvent être associés directement aux unités des codes spécifiques. Une ligne principale se termine par un marqueur de fin d'énoncé : un point. Selon le manuel CHILDES, "chaque ligne principale doit coder un et un seul énoncé" [110] .

Exemple [111] : *MOT: could you please spell your name? *MAR: it's m@l a@l r@l k@l.

Pour nous, une ligne principale comporte une intervention qui porte des marques séparant les clauses. Cette attitude a été adoptée de façon à permettre la correspondance parfaite entre nos corpus brut et les corpus codés dans un souci de lisibilité, elle permet aussi de conserver les caractéristiques de la production sur les IRC, qui sont essentielles à son analyse.

À la ligne principale sont associées des lignes dépendantes qui présentent des codages pour la ligne principale à laquelle ils sont associés.

Ces lignes permettent par exemple de spécifier l'identifiant de l'allocutaire, la langue utilisée, les actes de langage, la morphologie, la syntaxe…

Exemple : *MOT: are you okay? %spa: $x:dhs $i:yq [112]

Dans cet exemple, sont codés les actes de langage du point de vue du type d'échange ($x:dhs *discussing hearer's sentiments*) et des forces illocutoires ($i:yq *ask a yes/no question*).

Pour cette étude, nous avons associé à chaque ligne principale au moins six lignes de codages qui valent pour la langue (%lan), l'allocutaire (%add), la lemmatisation (%mor), le type d'intervention (%typ), l'unité conversationnelle à laquelle l'intervention se

[110] MacWhinney (1995:8) "Each main line should code one and only one utterance."

[111] ***Issu de la documentation du logiciel http://childes.psy.cmu.edu/clan/ chat.pdf p.41***

[112] ***Issu de la documentation du logiciel http://childes.psy.cmu.edu/clan/ chat.pdf p.152***

en vertu de la loi du droit d'auteur.

rapporte (%suj), les indications de situation (%exp), les constructions syntaxiques (%syn).

Exemple : *PCK: salut . %mor: co|salut . %syn: A %lan: FR %typ: O %suj: INT %add: PCB

Les principes de codages seront présentés au cours des chapitres concernés, afin d'associer explicitement ces derniers aux questionnements qui les ont suscités.

4.5. Conclusion

Ce chapitre a présenté la façon dont nous avons traité les lignes de messages de nos corpus, selon plusieurs paramètres, nécessaires aux analyses ultérieures. Il s'est agit, en effet, d'effectuer un premier tri des interventions, en fonction de la langue utilisée par le participant. Cette section a été l'occasion de remarques préliminaires concernant le français de nos corpus. Nous avons conservé dans les corpus toutes les lignes de messages, mais elles ont été annotées pour la langue de façon à pouvoir être sélectionnées automatiquement selon ce critère.

Nous avons ici également détaillé les principes de standardisation et de lemmatisation que nous avons mis en place, sur les interventions en français. Ces traitements nous sont utiles tout au long de cette étude, mais plus spécifiquement, aux chapitres concernant le code et le lexique.

Une section importante a été consacrée à la description des mesures d'anonymisation que nous avons mises en place, et à ce qu'elles impliquent comme pertes dans le corpus. Nous avons toutefois expliqué comment limiter ces dernières, en insérant des balises dans le corpus pour les signaler et caractériser les phénomènes.

Enfin, nous avons rapidement présenté le logiciel que nous avons utilisé pour recueillir le corpus ainsi prétraité, ainsi que les critères qui ont guidé notre choix.

À l'issue des chapitres 3 et 4 on aura donc décrit les corpus tels que nous les avons collectés ainsi qu'une partie des traitements de base qui leur ont été appliqués. Il convient à présent d'entrer dans l'analyse proprement dite et nous envisagerons dans un premier temps la matérialité linguistique.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Chapitre 5 Le code écrit des IRC

La recrudescence de l'intérêt pour la chose parlée, qui est spécifique à la linguistique du XX e siècle, contribue à la réflexion sur la distinction qui paraît fondamentale entre l'oral et l'écrit. Françoise Gadet (1996:13) affirme que "différents facteurs, liés aux modifications des modalités d'échange au XX e siècle, rendent cette distinction de plus en plus fragile, au-delà de l'évidence matérielle de différences entre les chenaux".

Le dispositif de communication qui engendre nos corpus est, selon son inventeur, "basé sur le texte" [113], c'est-à-dire en fait sur l'écrit dactylographié. Nous montrerons toutefois qu'il ne répond pas à la distinction oral/écrit telle qu'elle est formulée traditionnellement. On pourrait certes arguer qu'il s'agit d'un phénomène marginal, qui ne justifie donc pas de revoir la terminologie. Mais les observations que nous faisons semblent s'appliquer à d'autres dispositifs de communication eux aussi basés sur le texte écrit tels que les courriers électroniques, les jeux en réseau, et les textos GSM qui réalisent une plus grande couverture de la population. Les campagnes publicitaires fixes (affiches, pages de magazines ou de presse) présentent des phénomènes similaires. Ce que les nouveaux moyens de communication révèlent, ce sont en partie des usages volatiles et incontrôlables de l'écrit : notes de cours, listes de courses, mémos… [114] Le primat du matériau dans la considération des distinctions est sans doute à revisiter à l'exemple du travail de Biber (1988) et peut-être même faudrait-il ainsi abandonner la

[113] Jarkko Oikarinen http://www.kumpu.org/jto/research/index.html

[114] Branca-Rosoff (1998).

en vertu de la loi du droit d'auteur.

dénomination de la distinction au profit d'une typologie des produits linguistiques, ou tout au moins revoir la définition de ces termes et affirmer plus explicitement que la distinction oral/écrit recouvre deux réalités : un matériau sémiotique et un "style" prototypique.

Au sujet de l'écrit on relève dans la littérature deux attitudes principales : celles qui adoptent une perspective autonomiste (les unités graphiques ne sont pas signes de signes mais signes de choses, la langue écrite est un système spécifique qui n'est qu'en partielle interaction avec la langue parlée) et celles qui maintiennent la relation de représentation de l'écrit à l'oral. Les systèmes d'écritures dont la relation à l'oral est distante [115] facilitent la première position, ainsi Anis, Chiss et Puech (1988:11) posent "l'indépendance, autonomie de la graphie qui exploite les trois dimensions de l'espace alors que la parole et l'écriture phonétisée se déploient dans une unique dimension spatiale chargée de figurer la linéarité du temps." Une description autonomiste du système graphique du français se trouve dans Anis *et al.* (1988:87-144). Toutefois, si des spécificités doivent être reconnues pour chacun des matériau sémiotiques, l'autonomie de l'écrit est une position qui se heurte à de graves objections. Nous verrons, à travers ce chapitre, que cette position radicale n'est pas soutenable. Incontestablement, notre corpus est de l'écrit de par la nature du canal et du matériau sémiotique. Il n'est toutefois pas un écrit prototypique.

Notre intérêt se porte donc en premier lieu sur le code. La communication basée sur l'écrit en présence de laquelle nous sommes, est synchrone [116] , mais à la différence des dispositifs synchrones basés sur l'oral [117] , outre qu'elle dépouille les messages des éléments qui accompagnent sa production (gestes, mimiques, postures, orientation des regards…), elle prive également les interlocuteurs des informations portées par la prosodie.

Dans le corpus étudié, nous avons pu relever différents phénomènes dont certains semblent relever de différents types de représentations de caractéristiques acoustiques et non-verbales de l'oral en situation.

Une première partie de ce chapitre sera consacrée à l'inventaire des divergences constatées en ce qui concerne le code graphique, en examinant les phénomènes de réduction formelle, puis les phénomènes n'impliquant pas de réduction formelle. Nous supposons en effet que la contrainte temporelle imposée par la synchronicité du dispositif influe sur la longueur des unités linguistiques, ainsi que nous l'avons vu au chapitre 3. Nous verrons toutefois que l'hypothèse de la contrainte temporelle ne suffit pas à expliquer l'ensemble des phénomènes de réduction formelle ni l'ensemble des phénomènes s'écartant de la norme de l'écrit. Dans une seconde partie, nous examinerons une spécificité reconnue des écrits Internet : les émoticones.

[115] On pourrait ainsi dessiner une échelle de distance qui placerait les systèmes mythographiques à un extrême, puis comporterait successivement les systèmes idéographiques, logosyllabiques, syllabiques puis alphabétiques.

[116] Elle requiert que les participants soient engagés dans le même temps.

[117] Téléphone, CB, interphone,…

en vertu de la loi du droit d'auteur.

5.1. Code verbal

Cette section présente le versant verbal dans le code écrit des IRC, la section suivante se consacrera aux émoticones. Nous tentons de dégager ce qui dans l'utilisation du matériau dactylographié se présente comme appartenant spécifiquement à ce type d'interaction communicative.

5.1.1. Orthographe

On a coutume de considérer l'orthographe comme une institution, née d'un besoin technique pour une lecture efficace, l'écrit présentant une moindre tolérance à l'ambiguïté que l'oral, les négociations sur le code étant potentiellement moins immédiates. Les normes de l'écrit sont pourtant quotidiennement malmenées [118] sans catastrophe.

Étant donné les caractéristiques spécifiques du code IRC, dont nous n'avons jusqu'ici donné que quelques aperçus, il est très difficile de se prononcer à propos de l'orthographe. Les graphies rencontrées dans le corpus sont plus que souvent déviantes. Dans nos corpus, 26,42% des unités linguistiques présentent au moins un écart à la norme. Nous faisons l'hypothèse ici qu'il s'agit majoritairement d'une tendance volontaire et contrôlée à tenter de représenter l'oral tel qu'il est prononcé, proféré. Le non respect manifeste de la norme est peut-être, comme le suggèrent certains puristes, une stratégie pour masquer la non maîtrise du code écrit, mais il est hasardeux de prendre position sur cette question. Il est certes une part des graphies déviantes qui ne s'expliquent pas autrement que par la notion de faute d'orthographe ou erreur de frappe – ces deux types n'ayant évidemment pas le même statut. Les premières portent soit sur des morphogrammes grammaticaux et lexicaux, soit sur des logogrammes [119] , lettres historiques et étymologiques. Nous nous contenterons sur ce point, notre intérêt n'étant pas de quantifier le degré d'acquisition de la norme, de signaler que les déviances qui répondent à la notion de faute sont bien plus importantes sur les morphogrammes, mais les stratégies d'évitement des difficultés de l'autre catégorie ne pouvant être contrôlées il ne peut s'en tirer de conclusion autre que hâtive. Pour ce qui concerne les erreurs de frappe nous avons compté les substitutions de caractères proches sur le clavier, les inversions de caractères, les espaces et apostrophes mal placées, phénomènes illustrés dans les quelques exemples suivants, et mis en relief par l'attribut gras.

Extrait du Corpus P6 *PH3: ca ba PI7 ? Extrait du Corpus P3 *PAS: plein de ptits

[118] Lucci *et al.* (1994).

[119] Au sens de Catach (1995), les logogrammes sont ensemble d'unités graphiques formant des "images de mots". La fonction essentielle des logogrammes est de distinguer formellement des homophones. Les morphogrammes grammaticaux sont les indications de flexion, tandis que les morphogrammes lexicaux sont des éléments dont le rôle est de conserver une identité formelle à des unités lexicales de même "famille".

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

ballins qui se gonflent

'b' et 'v' d'une part et 'i' et 'o' d'autre part sont proches sur le clavier français [120] , ce qui laisse penser qu'il ne s'agit dans ces cas que d'erreurs de frappe – substitution de caractères. On peut se faire la même remarque au sujet des exemples suivants, dans lesquels à l'intérieur d'une unité, ou en frontière d'unités, des caractères – ici marqués en gras – ont été inversés.

Extrait du Corpus P7 *PAK: bah jvais alelr la faire changer de la banque aujourdhui Extrait du Corpus P7 *PDS: j' aodre l'anglais mdr *PCA: ce'st vrai koi Extrait du Corpus F7 *PBP: alors PEI la demi finale d ela coupe de france ? [121]

De plus, l'utilisation des caractères comportant un accent ou une cédille – *i. e.* tous ceux qui ne se présentent pas sur les claviers prévus pour l'anglais – étant toujours problématique entre plateformes différentes, nous n'avons pas prêté attention à leur substitution par des caractères simples, cela étant depuis longtemps recommandé par la netiquette.

Parmi les déviances restantes, nous avons considéré deux catégories : d'une part une catégorie dans laquelle les déviances peuvent s'expliquer par les contraintes temporelles imposées par le dispositif, et d'autre part une catégorie rassemblant les déviances liées à une volonté de représentation de la matérialité de la chaîne parlée. Nous essayons ci-après de caractériser les autres déviances.

5.1.2. Abréviations

En premier lieu, ce qui frappe dans la matérialité de la communication IRC est l'abondance des abréviations. Ce fait pourrait s'expliquer par les contraintes temporelles générées par le dispositif. Le temps nécessaire à la production d'un message est en effet plus lent à l'écrit qu'à l'oral : la vitesse de frappe la plus élevée n'approche pas le débit d'élocution moyen (150 à 200 mots à la minute selon Levelt 1989, 1993). Rappelons que l'accessibilité d'un message à son (ou ses) destinataire(s) n'a lieu que lorsque le locuteur émetteur a validé l'envoi du message. Pour être pertinentes les interventions réactives doivent être temporellement proches de leur source donc d'une longueur limitée. Du point de vue du décodage, le traitement à l'oral est presque simultané à la production, il est donc en somme conditionné par le débit de production, tandis qu'à l'écrit les stimuli sont traités indépendamment de leur production [122] . L'écart entre la vitesse de production et la vitesse de traitement devrait donc engendrer des stratégies d'équilibrage qui en augmentant la vitesse de production ont pour conséquence de ralentir le temps de traitement. Une stratégie pour réduire le temps de production est la réduction de la forme.

120 Nous avons évidemment été obligés également de considérer les erreurs pour un clavier anglais.

121 Dans cet exemple, un caractère espace et le 'e' ont été inversés.

122 Les données chiffrées pour les vitesses de production et de traitement à l'écrit et à l'oral varient énormément notamment en fonction des stimuli, toutefois on peut dire globalement que la vitesse de traitement de stimuli oraux est inférieure à la vitesse de traitement de stimuli écrits comparables, et que la vitesse de production d'énoncé oraux est supérieure à la vitesse de production d'énoncés écrits comparables. L'écart entre les vitesses de traitement et de production est plus important pour l'écrit que pour l'oral.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

La langue contemporaine utilise des procédés d'abréviation que ce soit à l'écrit ou à l'oral. Sous le terme d'abréviation on rassemble des phénomènes de réduction de la forme d'une unité lexicale à l'exception des réductions imposées par la phonologie, "elle consiste à manifester une unité linguistique par un signifiant qui, amputé d'un ou plusieurs éléments, conserve le signifié de l'unité de départ" [123] . Ce phénomène, à l'oral, est réputé concourir à la néologie lexicale. Il économise la quantité de matériel signifiant nécessaire à la référence. On s'attend donc à ce qu'il se produise sur des unités relativement longues et fréquentes et/ou d'usage courant [124] . La fréquence des unités lexicales dépend des situations référentielles. Les unités lexicales affectées par l'abréviation sont caractéristiques, entre autres [125] , de la variation diastratique.

Nous considérerons trois types d'abréviation, conformément à l'attitude générale : l'abréviation proprement dite, en ajoutant un point sur ses spécificités à l'écrit, la siglaison et l'abréviation en acronymes.

5.1.2.1. Cas général

L'abréviation consiste le plus souvent en la réduction d'un signifiant par troncation de la fin d'un mot, on en distingue cependant deux autres configurations : troncation du début d'un mot, sélection d'une unité lexicale d'un complexe. L'abréviation intervient sur des unités lexicales longues, de ce fait les mots de composition savante sont des unités privilégiées pour l'abréviation, la troncation se réalisant sur les dernières syllabes (apocope), mais pas toujours comme une stricte décomposition : *météoro-logie* est abrégé en *météo* ; *cinémato-graphe* en *cinéma* ou *ciné* ; *kinési-thérapeute* en *kinési* puis *kiné*, etc. Les unités lexicales ordinaires longues (plus de trois syllabes) sont aussi la cible de l'abréviation selon le procédé de la troncation en finale (apocopes : *prof, fac, bac, interro, compo, manif*…) et plus rarement en initiale (aphérèses : *bus, car*). Lorsqu'un *denotatum*est désigné par une combinaison d'unités lexicales, seule l'unité lexicale la plus saillante résiste. C'est le cas pour beaucoup de référents du monde contemporain tels que *journal quotidien* [126] *, journal périodique mensuel , chemin de fer métro politain* [127] *, répétition générale* …

Dans nos corpus, ces différents types d'abréviation sont évidemment représentés (cf. liste page suivante), la question se pose de savoir si le recours à leur utilisation est un signe distinctif de ce moyen de communication.

[123] Arrivé, Gadet, Galmiche (1986)

[124] La distinction entre le caractère fréquent et courant concerne respectivement le nombre élevé d'occurrences d'un terme, et l'appartenance d'un item au vocabulaire de base, ou commun.

[125] Le parler "jeune", par exemple, en est aussi fortement empreint.

[126] L'unité soulignée est ce qui résiste à l'abréviation

[127] Pour ce cas, deux procédés séquentiels d'abréviation ont été impliqués : sélection de l'unité lexicale prégnante, puis troncation de la finale.

en vertu de la loi du droit d'auteur.

Tableau (5-1) – Principales abréviations repérées dans le corpus, classées par type.

Apocopes ado adolescent alim alimentation (électrique) ampli amplificateur appart appartement auto automatiquement autos automatiques auto automobile bermud bermuda bio biologie cam caméra cata catastrophe catho catholique ciné cinéma clito clitoris coloc colocataire config configurer croco crocodile cuni cunnilingus deg dégoûtant, dégueulasse degueu dégueulasse dispo disponible exam examen fan fanatique fana fanatique fru frustré fut futal gastro gastroentérite gynéco gynécologue hexa hexadécimal info informatique infos informations mat matin	math mathématique max maximum mayo mayonnaise muscu musculation mytho mythomane médic médicament ophtalmo ophtalmologue ordi ordinateur parano paranoïaque pass password pede pédéraste perfu perfusion perso personnellement polio poliomyélite pro professionnel prob problème prog programme pseudo pseudonyme psy psychiatre pub publicité périph périphérique quasi quasiment re rebonjour, resalut resto restaurant répu réputation scato scatologique séropo séropositif soss social stats statistiques tel téléphone télé télévision uni université (suisse)
apocopes avec modification en finale annif anniversaire apéro apéritif dico dictionnaire ex-alcolo ex-alcoolique hosto hôpital intello intellectuel	**aphérèses** culés enculés lut salut ricains américains tain putain zic musique
modification d'une partie d'un complexe lexical aprem après-midi beauf beau-frère bon app bon après midi, bon appétit comme d'hab comme d'habitude d'acc d'accord génie info génie informatique laisse be laisse béton petit dej petit déjeuner sciences po sciences politiques	Troncations de termes spécifiques à l'univers de référence IRC **apocopes** nick nickname pseudo pseudonyme chan channel op opérateur serv services **aphérèse** bot robot
sélection d'une partie d'un complexe lexical grasse grasse matinée à plus à plus tard à toute à toute à l'heure d facon de toute façon dur disque dur	**sélection d'une partie d'un complexe lexical et troncation** caps lettres capitales

Il arrive aussi que des unités lexicales soient abrégées dans le cours de la conversation, sans que cela puisse être considéré comme conventionnel, par exemple pour des unités telles que les suivantes :

gran	grand-mère
radi	radiateur
vac	vacances

dans ce cas, c'est le co-texte et l'univers de référence construit par le co-texte qui permettent de décoder l'abréviation.

On remarquera que le critère de longueur n'est pas strict puisque plusieurs unités

en vertu de la loi du droit d'auteur.

tronquées ne comportent à la base que deux syllabes. Les unités affectées sont essentiellement des noms, quelques adjectifs et adverbes, quelques rares locutions et un seul verbe.

5.1.2.2. Spécificité des procédés d'abréviation dans le code écrit

L'écrit connaît plusieurs types d'abréviation qui lui sont propres, accordant quelques arguments à l'hypothèse autonomiste. On se convaincra à la lecture des tables d'abréviations présentées comme courantes dans les dictionnaires qu'il y règne un désordre impressionnant. Les listes en effet ne se recoupent que partiellement, elles rassemblent des éléments d'utilisation générale autant que spécifique à des groupes de locuteurs et ne proposent pas de systématisation de l'ensemble [128] . Anis *et al.* (1988:139-144) proposent leur traitement au sein d'une catégorie d'unités qu'ils dénomment logogrammes et pour laquelle ils proposent 5 groupes :

[128] On pourrait objecter aux deux dernières remarques qu'il ne s'agit pas des objectifs des dictionnaires. Cependant, si la nomenclature et le choix de la présentation alphabétique les masquent, il y paraît pour les unités lexicales quelques tentatives de catégorisation.

[129] Lettres de l'alphabet.

[130] Signes de ponctuation.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Logogrammes	La série concernée est : $, £, §, &… Ces items se manifestent "par une forme matérielle indécomposable (accessible sur les claviers des machines à écrire), et sont l'équivalent d'une séquence d'alphagrammes [129] " c'est-à-dire d'un mot. Exemple : *voir* § *suivant*
Symboles mathématiques	Il s'agit là, des chiffres (qui peuvent composer des nombres), et des opérateurs. À ces derniers s'ajoutent les items suivants : =, %, <, >, mais la liste n'est pas définie, la frontière devenant floue quant à certaines unités telles que les unités de mesure. Exemple : *25% des Français ont regardé le match*.
Abréviations et sigles	Dans cette classe, selon Anis *et al.* (1988:141), "Les abréviations comme M., p., correspondent à des noms communs, alors que les sigles correspondent à des quasi noms propres". Il est aussi signalé que les unités qui composent abréviations et sigles sont "des alphagrammes détournés de leur utilisation normale" et que ce "détournement est habituellement signalé par des topogrammes [130] déviés eux aussi de leur usage normal". Exemple : *M. Janu procède à l'état des lieux.*
Logos	Anis *et al.* (1988:142) font entrer dans leur classification ce phénomène récent qui "consiste pour une société qui veut augmenter l'impact sur le public de son nom, à figer l'écriture de celui-ci dans un dispositif graphique particulier."
Topogrammes détournés	Ce cas est celui de points d'interrogation et de points d'exclamation employés seuls "pour exprimer de manière elliptique des valeurs sémantiques proches de celles qu'ils expriment dans leur usage normal". Cet usage chercherait à "simuler par ce procédé les mimiques expressives de l'échange oral" – Anis *et al.* (1988:143). Exemple : *"Marie est arrivée. – Ah ??!!".*

Pour ce qui nous concerne, nous traiterons des deux premiers types dans cette section, l'utilisation des logos n'apparaissant pas dans notre corpus et les topogrammes

[129] Lettres de l'alphabet.

[130] Signes de ponctuation.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

détournés étant traités au point concernant la ponctuation. Les sigles font l'objet de la section suivante.

Quant aux listes d'abréviations données pour la langue française [131] , dont nous mentionnions l'hétérogénéité, il est possible de repérer quelques types réguliers. L'un d'eux consiste à ne retenir qu'une lettre de l'unité lexicale. Ce procédé concerne par exemple les points cardinaux (*nord N, sud S…*). L'abréviation des unités de mesure retient une ou plusieurs lettres de l'unité lexicale de base et est invariable (*kilogramme(s) kg, minute(s) min, centimètre(s) cm…*). Les ordinaux sont convertis en chiffres arabes accompagnés de la finale de l'unité lexicale de base en exposant, fléchie pour le genre et le nombre le cas échéant (*premier 1 er , premiers 1 ers , première 1 re , deuxième 2 e …*). La sélection de la première lettre de l'unité lexicale et de la finale (à l'exposant) est utilisée pour les titres de civilité *M.* ou *M r* pour *monsieur*, *D r* pour *Docteur*, *M me* pour *madame*, *M lle* pour *mademoiselle* et quelques autres comme *C ie* pour *compagnie*… Une abréviation peut aussi, à l'écrit, ne se composer que des premières lettres du mot et dans ce cas, elle se termine par un point (*référence réf.)*. Elle peut ne comporter que la première et la dernière lettre d'une unité lexicale et dans ce cas, elle se termine par la dernière lettre du mot et ne comporte pas de point final (*boulevard bd*).

> Enfin à l'écrit toujours, une abréviation peut ne comporter que les lettres initiales d'un groupe de mots graphiques (comme un sigle) : *C.V. pour Cheval-vapeur* ou *Curriculum vitae*, *C.-à-d.* pour *C'est-à-dire*, *S.V.P.* pour *S'il vous plait*...Certains éléments affectés par ce procédé le sont aussi à l'oral et leur forme correspond alors à la prononciation des lettres de l'alphabet qui composent l'abréviation : SVP donne [esvepe], CV [seve][132]. On notera aussi qu'il se présente des cas d'homonymie : *C.V. pour Cheval-vapeur* ou *Curriculum vitae*.

Nos corpus présentent ces types d'abréviations conventionnelles normées ; on constate un écart à la norme concernant les caractères qui devraient être en exposant et la ponctuation afférente. L'abréviation des unités de mesure temporelle correspond souvent à la norme anglo-saxonne.

A. LOGOGRAMMES

Les logogrammes présents dans le corpus sont en petit nombre, ce sont *$* pour *dollar(s)* et *&* [132] pour *et*, l'utilisation de ce dernier n'étant pas limitée à la conjonction de noms propres [133] ou plus généralement de syntagmes nominaux, comme le montrent les exemples suivants :

Extrait du Corpus F6 *PBP: la je grave, je surfe & j' ecoute un mp3....vive le scsi.

131 Cf. Grevisse & Goosse (1993), Robert, Rey-Debove et Rey, (1993), Imprimerie Nationale (1990).

132 L'esperluette (de nos jours dite *et commercial*) est à l'origine une ligature de l'époque mérovingienne.

133 Selon Anis *et al.* (1988:140) & "est réservé aux titres ou noms de marques ou associations d'auteurs."

en vertu de la loi du droit d'auteur.

*Extrait du Corpus F7 *PCK: alors entre la SNCF & ma caisse....VOILA la journee....*

B. SYMBOLES MATHÉMATIQUES ET SCIENTIFIQUES

Tous les nombres dans le corpus sont convertis en chiffres arabes, à l'exception de *un* et *deux*, le symbole *%* est systématiquement utilisé pour *pourcents,* et ° pour *degré .*

D'autre part les chiffres peuvent servir à former des rébus, et il en va de même pour les symboles d'opération mathématiques :

a+ tard	à plus tard
de + en +	de plus en plus
2 rien	de rien
qlq'1	quelqu'un
7	cette
1h/cd	une heure par cd
x	fois
1/2	demi

mais nous anticipons là sur une section ultérieure.

C. ABRÉVIATIONS

Les abréviations conventionnelles de l'écrit sont peu utilisées hormis celles qui valent pour les unités de mesure. Les unités de mesure, pour lesquelles la forme abrégée est utilisée, présentes dans le corpus sont

les unités de mesure de volume informatique,

Mo	mégaoctet(s)
Go	gigaoctet(s)
mo	mégaoctet(s)
go	gigaoctet(s)
kb/s	kilobits par seconde

les unités de mesure temporelle,

j	jour(s)
sec	seconde(s)
h	heure(s)
secs	secondes
H	heure(s)
s	seconde(s)
mn	minute(s)
min	minute(s)
mins	minutes

en vertu de la loi du droit d'auteur.

les unités de mesure spatiale,

cm	centimètre(s)
m	mètre(s)
km	kilomètre(s)
m2	mètre(s) carré(s)
kms	kilomètres

les unités de mesure monétaire,

F	francs
kF	Kilofrancs [134]
Fr	francs
FF	francs français
FR	francs
dm	deutschemarks
$	dollars [135]

Les autres abréviations conventionnelles sont données dans le tableau ci-après :

Tableau (5-2) – Autres abréviations conventionnelles du corpus

[134] Cette abréviation est utilisée aussi à l'oral sur le modèle de *svp* et *cv* mentionnés plus haut.

[135] Cet item est traité au paragraphe précédent, il ne se trouve dans cette liste que pour mémoire

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

abrév.	**forme complète**
1er	premier
1 ère	première
5 eme	cinquième
2eme	deuxième
cad	c'est-à-dire
env	environ
mr	monsieur
Déc	décembre
st	Saint
ST	Saint
RDV	rendez-vous
cv	cheval vapeur
SVP	s'il vous plaît
svp	
C14	carbone 14
etc	et cætera
ex	exemple
nb	nombre
tel	téléphone

Ces dernières s'écartent de la norme au sens strict, mais pas de l'usage.

Certaines abréviations, qui ne retiennent le plus souvent que la première lettre de l'unité lexicale ne font pas partie des listes des dictionnaires :

f	fille, femme
fr	français
g	garçon
ch	cherche
m	mâle
h	homme

Elles sont utilisées dans les présentations et les interventions du type petite annonce.

Extrait du Corpus F3 *PDL: oh la la1 PDL 16 f de Belgrade Extrait du Corpus F7 *PBP: PAB tu est un f ou m

Les autres abréviations, qui valent pour le système de l'écrit sont basées sur un principe de sélection de graphèmes. Ce procédé n'est pas non plus spécifique à l'IRC, chacun d'entre nous utilise ces formes d'abréviations dans des contextes variés (listes de courses, notes de cours,…). La variation interindividuelle est très importante pour ce phénomène. À une unité lexicale peuvent correspondre plusieurs abréviations *i. e.* différentes sélections de graphèmes. Ce qui est frappant dans notre corpus, c'est que ce type d'abréviation ne semble pas complètement idiosyncrasique. Il intervient sur des unités fréquentes et/ou grammaticales. Nous donnons dans le tableau 5-3 les abréviations pour lesquelles se présentent plus de 5 occurrences mais nous signalons

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

que celles que nous ne présentons pas ici portent sur le même type d'unités.

Tableau (5-3) – Sélection de graphèmes – items ayant au moins 5 occurrences.

en vertu de la loi du droit d'auteur.

abrév.	**unité de base**
bcp	beaucoup
chq	chaque
ds	dans
jms	jamais
longtps	longtemps
mnt	maintenant
mtn	
ms	mais
m	même
mm	
mme	
ns	nous
pcq	parce que
partt	partout
pe	peut-être
plz	please
pr	pour
pk	pourquoi
prq	
pb	problème
pbm	
qd	quand
qd meme	quand même
qd m	
qq un	quelqu'un
qcq	
qlqu'un	
qq	
qq1	
qqn	
qqun	
qq chose	quelque chose
qqc	
qqch	
qqchose	
qque part	quelque part
qq	quelques
slt	salut
st	sont
tps	temps
thx	thanks
tj	toujours
tjrs	
tjs	

en vertu de la loi du droit d'auteur.

ts	tous
tt	tout
tte	toute
vs	vous

Le prototype de cette forme d'abréviation est de retenir préférentiellement les graphèmes consonnes d'un mot – ce qui est le principe de certains systèmes d'écriture [136] – et ce même si ces graphèmes ne se prononcent pas.

Il apparaît dans notre corpus, en supplément des éléments présentés dans le tableau, quelques formes spécifiques au domaine de l'informatique et des réseaux : *dl* ou *d/l* pour *downloader* et des formes spécifiques à l'IRC : *pv* pour *privé*, *msg* pour *message.*

Les phénomènes présentés dans cette section contribuent tous à réduire la forme. L'écart à la norme est certain et la variabilité non négligeable, on ne constate cependant que peu de cas d'ambiguïté, et aucune ambiguïté qui ne puisse être levée par le contexte.

5.1.2.3. Acronymes et sigles

Acronymes et sigles sont des formes d'abréviation. On a coutume de distinguer ces unités en fonction de leur rapport à l'oral. Les sigles en effet connaissent une prononciation composée des noms des lettres qui constituent le sigle, tandis que les acronymes sont conçus de sorte que leur prononciation réponde aux principes de lecture de mots. Ces formes d'abréviations sont assez présentes dans nos corpus. Certaines manifestent des similitudes d'emploi avec les frimousses, que nous présenterons à la section suivante.

La langue contemporaine utilise essentiellement les lettres initiales d'un groupe de mots formant un syntagme nominal et ayant le statut d'un nom commun ou d'un nom propre qui peut recevoir des suffixes dérivationnels (ex. CAPES – capésien).

Quelques formes se composent des lettres initiales d'une locution ainsi que pour *s'il vous plaît, SVP* et *c'est-à-dire, c-à-d*.

Les sigles présents dans le corpus sont les suivants, ils peuvent se répartir en plusieurs classes selon le domaine notionnel concerné – nous écartons en tête de liste les sigles utilisés couramment et en fin de liste ceux que nous avons rencontrés dans nos corpus seulement [137] :

[136] Cf. arabe, ou système d'écriture des Phéniciens présenté dans Cohen, M. (1953). L'inventaire des consonnes dans une langue étant toujours plus important que celui des voyelles et les patrons de consonnes formant des unités lexicales étant plus diversifiés, les consonnes sont de meilleurs candidats à représenter la prononciation.

[137] La correspondance anglaise de sigles est marquée par une couleur de caractères grise.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Courant	
hs	hors service
stp	s'il te plaît
svp	s'il vous plaît
wc	water-closet
uv	ultraviolets
Commerce	
sav	service après vente
ttc	toutes taxes comprises
Études – Diplômes	
MIAS	Mathématique Informatique et Application aux Sciences
deug	diplôme d'études universitaires général
Entreprises, institutions (organisations)	
VP	vice président
pdg	président directeur général
Informatique : matériel et logiciel	
dd	disque dur
PC	personal computer
pc	personal computer
cd	compact disc
dvd	digital versatile disc
ram	random access memory
hd	hard disc
hp	haut-parleur
avi	Audio Video Interleave (format de fichier)
mp3	Motion Picture Experts Group 1 Audi Level 3 (format de fichier)
mpg	Motion Picture Experts Group (format de fichier)
divx	Digital Video Express (format de fichier)
ide	Integrated Device Electronics
aspi	Advanced SCSI Programming Interface
scsi	Small Computer Systems Interface
Internet et réseaux	
adsl	Asymmetric Digital Subscribe Line
ip	internet protocol
lan	Local Area Network
irc	internet relay chat
icq	I Seek You(nom propre)
pppoe	Point to Point Protocol over Ethernet
smtp	simple mail transfer protocol
rip	Raster Image Processor
url	uniform resource locator
Spécifiques IRC	

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

ml	mailing list
gt	go together
brb	be right back
gtg	got to go
dcc	Direct Client Connection
kb	kick ban
lmao	laughing my ass off
lol	laughing out loud
rofl	rolling on the floor laughing
mdr	mort de rire
asv	âge sexe ville
asl	âge sexe lieu (location)
fds	fin de semaine
jdr	jeu de rôle
tlm	tout le monde
alp	à la prochaine
we	week-end

Parmi les formes spécifiques à l'IRC, trois sont très fréquentes. Il s'agit de formes composées des lettres initiales d'unités sémantiques complexes : *mdr, lol, tlm* respectivement pour *mort de rire*, *laughing out loud*, et *tout le monde.*

Tableau (5-4) – Sigles fréquents

forme	nombre de variantes	nombre total d'occurrences	nombre d'occurrences de la forme vedette
mdr	25	188	131
lol	26	733	598
tlm	2	59	59

La forme *tlm* n'est jamais employée seule tandis que *mdr* et *lol* forment souvent des interventions complètes. La forme *tlm* se démarque encore des autres par sa stabilité formelle et dans le fait que ses emplois se répartissent comme suit :

Tableau (5-5) – Répartition de la forme *tlm* selon le type d'intervention

tlm		
type d'intervention	nombre occurrences	proportion (%)
ouverture	50	84,75
clôture	7	11,86
corps	2	3,39

Elle est d'autre part en concurrence avec la forme pleine *tout le monde* beaucoup plus employée.

en vertu de la loi du droit d'auteur.

Tableau (5-6) – Répartition de la forme *tout le monde* selon le type d'intervention

tout le monde		
type d'intervention	Nb. occurrences	proportion (%)
ouverture	161	77,40
clôture	17	8,17
corps	30	14,43

En ce qui concerne *mdr* et *lol*, les formes sont beaucoup moins stables comme l'illustrent les tableaux suivants (5-7 et 5-8) qui rassemblent l'ensemble des formes repérées dans les corpus avec le nombre d'occurrences pour chaque forme :

Tableau (5-7) – Variantes de la forme *lol*

forme	Nb. occ.
lol	598
lolol	75
lololol	19
lolololol	6
lool	6
loll	3
lolololololol	3
lolll	2
lolo	2
lolololololol	2
olololol	2
llolol	1
lol))	1
lol)	1
lollll	1
lolllll	1
lololll	1
lololo	1
lololoiklololol	1
lolololll	1
lolololololololol	1
lolololçololol	1
looo	1
loool	1
looool	1
loooooooooooooool	1
;lolololol	1
	732

en vertu de la loi du droit d'auteur.

Tableau (5-8) – Variantes de la forme *mdr*

forme	Nb. occ.
mdr	134
mdrr	11
mdrrr	5
mdrrrrrrrrrrr	5
mdrrrr	4
mdrrrrrrrrrrr	4
mdrrrrrrrr	3
mdrrrrrrrrrrrr	3
mdrrrrrrrrrrrrrrr	3
mdrrrrr	2
mdrrrrrr	2
mdrrrrrrr	2
mdrrrrrrrrr	1
mdrrrrrrrrrr	1
mdrrrrrrrrrrrrr	1
mdrrrrrrrrrrrrrr	1
mdrrrrrrrrrrrrrrrr	1
mdrrrrrrrrrrrrrrrrr	1
mdrrrrrrrrrrrrrrrrrrrrrrrrrr	1
mdrrrrrrrrrrrrrrrrrrrrrrrrrrr	1
mdrrrrrrrrrrrrrrrrrrrrrrrrrrrrr	1
mdrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr	1
mdrrrrrrrrrrrrrrrrrrr

La forme complète correspondant à *lol* est utilisée dans une intervention seulement, dans la séquence d'où est issu l'extrait ci-dessus, pour expliciter *lol*.

La répartition de ces formes suivant le type d'intervention dans le corpus se présente comme suit.

Tableau (5-9) – Répartition de la forme *lol* selon le type d'intervention

type d'intervention	Nb. d'occurrences	proportion (%)
ouverture	38	5,19
corps	690	94,26
clôture	4	0,55
total	732	100,00

Un comportement similaire est observé pour *mdr,* ainsi qu'on peut l'observer dans le tableau (5-10).

Tableau (5-10) – Répartition de la forme *mdr* selon le type d'intervention

type d'intervention	Nb. d'occurrences	proportion (%)
ouverture	6	3,14
corps	184	96,34
clôture	1	0,52
	191	100,00

Ces unités sont proches du point de vue de la référence, ce qui explique leurs comportements proches. Une autre unité présente le même comportement : *ptdr* (*pété de rire*). De formation plus complexe puisque combinant le principe du rébus (ou de la sélection de graphèmes) et celui du sigle, elle manifeste des variations similaires à celles que connaît *mdr*, ainsi qu'observé dans le tableau (5-11).

Tableau (5-11) – Variantes de la forme *ptdr*

forme	Nb. occ.
ptdr	27
ptdrr	3
ptdrrrrrrr	2
pdr	1
ptdrrr	1
ptds	1

Toutefois, contrairement à ce que nous avons remarqué pour les unités de même type présentées précédemment, aucune de ces formes n'apparaît dans les séquences d'ouverture ou de clôture dans notre corpus. Ces formes fonctionnent comme des rires, régulateurs de l'activité communicative.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

L'hypothèse selon laquelle la contrainte temporelle joue sur la matérialité linguistique est donc confirmée par l'observation des procédés de réduction formelle présentés jusqu'ici, et qui sont assez couramment observés dans d'autres types de productions orales et/ou écrites. Nous nous engageons dans la section suivante à l'exposition de phénomènes plus particuliers de réduction formelle.

5.1.3. Autres phénomènes de réduction formelle

Nos corpus montrent, nous l'allons voir, que le code IRC manifeste une pseudo écriture phonétique, qui se base sur plusieurs procédés. La difficulté de l'étude du code à ce niveau réside dans le mélange des deux systèmes. Le système graphique du français entretient un rapport à l'oral dans la mesure où il est alphabétique, bien que certains tentent de lui donner une description autonome [139] . Il existe des correspondances entre les unités phonétiques et les unités graphiques qui sont une première phonologisation, ou abstraction de la réalité sonore.

Selon Catach (1995:27) "80 à 85 % des signes d'un texte quelconque sont chargés en français de transcrire les sons". Les autres signes apportent un "complément d'ordre syntagmatique (positionnel) ou paradigmatique (flexionnel ou dérivatif) ou sont des lettres étymologiques, historiques ou diacritiques". Catach (1995:27) note encore que "3 à 6 % des mots ont […] en français une graphie globale spécifique" qui utilise divers éléments du système dans un objectif spécifique de distinction formelle.

Dans un premier temps, il s'agit de se donner des critères fiables permettant de reconnaître le phénomène de représentation des sons. La difficulté à repérer les représentations de l'oral dans le code écrit des IRC est en partie liée à l'orthographe. Le code écrit normé est une représentation de l'oral qui contient en lui-même une forme d'abstraction, de systématisation et de standardisation. Le système graphémique entretient des rapports avec le système phonémique qui sont complexes [140] .

Certains phénomènes présents dans nos corpus montrent indiscutablement une tentative de représentation phonétique. Le procédé du rébus est une preuve incontestable de la représentation de sons de l'oral, sauf à dire que des lettres représentent d'autres lettres. On ne peut donc pas considérer le code comme relevant purement du système graphique et se contenter médiocrement de relever des fautes. Il reste à déterminer à quel point la représentation de l'oral intervient dans le code, quelles sont ses formes et quelles sont ses fonctions.

5.1.3.1. Le procédé du rébus

Un rébus, faut-il le rappeler, est le résultat d'une activité linguistique ludique, qui met en jeu un ensemble de formes visuelles (dessins, symboles, pictogrammes, chiffres, lettres, mots…). Un signifiant iconique évoque un signifié qui à son tour appelle un signifiant

[139] Anis *et al.* (1988)

[140] Pour un exposé on se reportera à Blanche-Benveniste (1978) et Catach (1995)

en vertu de la loi du droit d'auteur.

linguistique (forme phonétique). Le rébus utilise des identités ou ressemblances formelles (homophonies). Ce procédé a été identifié dans un certain nombre de systèmes d'écriture, et notamment celui de l'écriture hiéroglyphique dans lequel un idéogramme peut être utilisé pour représenter un homophone [141] . Ce phénomène se produit dans nos corpus sous une forme simple qui consiste à utiliser des chiffres ou des lettres pour représenter leur prononciation dans l'activité d'épellation ou de récitation de l'alphabet. Le procédé du rébus ne se soucie pas de la segmentation en unités, un caractère peut représenter la prononciation d'un mot, d'une partie d'un mot ou de plusieurs mots.

Tableau (5-12) – Rébus

bb	bébé
pd	pédé
ct	c'était
gt	j'étais
c	c'est
c	ces
C	sait
D	des
L	elle
K	cas
m	aime
R	air
v	vais
g	j'ai
J	j'y
jm	j'aime

Ainsi que le montre le dernier exemple, des combinaisons entre utilisation des lettres comme graphèmes et utilisation comme élément de rébus apparaissent. Il ne s'agit en effet pas de "j'y aime" mais bien de "j'aime" ainsi que l'illustre l'extrait ci-dessous.

Extrait du Corpus P7 *PAI: jm dire ce ke je pense c tout!

5.1.3.2. Généralisation graphémique

Cette sous-section concerne l'utilisation de caractères alphabétiques pour la représentation des sons avec ou sans transformation de la prononciation standard, et pour des unités empruntées. On serait porté à considérer ces manifestations comme des fautes d'orthographe si elles ne touchaient des unités fréquentes et ne présentaient pas une certaine cohérence à travers le corpus.

Les caractères utilisés pour représenter des sons sont les graphèmes fréquents dans leur correspondance avec un son, ou les moins ambigus. Le son [k], dans les cas où la norme orthographique n'est pas respectée, est presque systématiquement représenté par 'k'. Il arrive qu'il soit représenté par 'q' même en initiale ou en milieu de mot.

[141] cf. Berger (1891); Gardiner (1950); Sottas & Drioton (1987), Lambert (1966)

en vertu de la loi du droit d'auteur.

Extrait du Corpus F1 *PAZ: vous voulez faire koi PAY? %add: PAY Extrait du Corpus P6 *PJ8: c'est ce qon dit

Ce phénomène produit à la fois des réductions formelles et tente d'introduire de la régularité dans la conversion graphème phonème.

5.1.3.3. Chute de voyelle

La tendance à ne pas représenter les lettres qui n'ont pas de contrepartie phonétique ne se porte pas seulement sur les lettres étymologiques ou morphogrammes, elle concerne aussi des voyelles qui seraient prononcées en français standard. La chute de voyelle en position médiane est marquée dans nos corpus et ne concerne pas seulement le [œ].

Syncopes :

Extrait du Corpus F6 *PAX: b'jour Extrait du Corpus P2 *PAB: vla qui fait des pings mainan lui :/

La réduction formelle qui découle de ce procédé a aussi l'effet de représenter des "manières de parler", des phonostyles.

5.1.3.4. Assimilation

On constate aussi des représentations de phénomènes d'assimilation régressive après chute du [œ], ainsi que l'illustre l'exemple suivant.

Extrait du Corpus P2 *PAI: PAP vi bah c ce que je dis, chte donne po de cado et c tt

L'assimilation de sonorité après chute du [œ] devant la consonne sourde [s], dans nos corpus, est toujours accompagnée de la chute de cette consonne, ainsi qu'on peut l'observer dans les quelques exemples suivants

Extrait du Corpus F3 *PAJ: et chui bien placé pour parler, mes ramps sont là bas mdr Extrait du Corpus F3 *PBG: chais po Extrait du Corpus F5 *PEN: tain chui meme plus capable de boire un litre de bière cul sec :/

L'économie formelle réalisée n'est pas le seul résultat du procédé, une tentative de représentation de l'oral est ici effectivement à l'oeuvre. Quelques autres phénomènes présentent cette caractéristique, nous les présentons ci-après. L'avancement dans ce chapitre nous conduira donc à nous écarter progressivement l'hypothèse de l'effet de la contrainte temporelle comme modifiant la quantité de matériau linguistique.

5.1.3.5. Simplification de groupes consonantiques

Le phénomène est bien connu des phonéticiens. Pour l'expliquer, on recourt au schéma canonique de la syllabe CVCV. La simplification dépend de la nature des consonnes qui composent le groupe et de leur position dans la chaîne parlée.

Extrait du Corpus F4 *PAG: PAG si il est beau, c'est qu'il n'est pas de chez moi

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

%add: PCA (1) *PCA: kesstensé? %add: PAG Extrait du Corpus P6 *PD9: kesta fait de ton XXX? %add: PJ7

5.1.3.6. Jointures : Liaison et élision

Une autre manifestation à compter au rang des preuves de la représentation de sons de l'oral est celui de la jointure. La jointure regroupe deux phénomènes syntactiques, celui de la liaison et celui de l'élision [142] . La liaison ne connaît pas de contrepartie dans le système écrit [143] tandis que l'élision est marquée sur le plan graphique. Selon Marchal (1980) "il s'agit de la survivance de quelques enchaînements de consonnes de l'ancien français. Il y a liaison [144] lorsque la consonne finale d'un mot, normalement non prononcée [145] est articulée devant la voyelle ou l'*h* muet initial du mot qui suit". On notera ici le recours à la graphie – non satisfaisant – pour expliquer un phénomène oral ; la position générativiste qui consiste à donner comme forme de base d'un mot susceptible de présenter le phénomène de liaison la forme avec consonne, si elle est beaucoup plus élégante, n'explicite pas mieux la question. L'élision est considérée comme un cas d'enchaînement, les séquences d'enchaînements vocaliques étant mal tolérées en dehors des formes verbales en français.

A. LIAISON

> Les liaisons sont réparties traditionnellement en trois types : obligatoire, facultative, abusive. Ces trois types sont définis en fonction de deux règles principales concernant la liaison qui stipulent qu'il n'y a pas de liaison après un mot accentué (*i. e.* fin de groupe rythmique) et qu'il y a liaison à l'intérieur d'un groupe rythmique. La liaison est réputée obligatoire à l'intérieur du groupe nominal [lezãfã], à l'intérieur du groupe adjectival entre adverbe et adjectif [trezelegã], après certaines prépositions (*sans, dans, en*) [ãnalã], entre indice de sujet, clitique et verbe [nuzãnubliõ]... Les liaisons facultatives le seraient par l'interférence de facteurs phonétiques qui affaibliraient l'obligation de liaison pour les polysyllabes, après consonne, et lorsque la liaison formerait un groupe non euphonique[147].

[142] Ou enchaînement selon la terminologie de Delattre (1966:55), puis Encrevé (1988). Chez Delattre, le terme vaut pour les cas où la syllabification rassemble la consonne finale d'un mot et l'initiale du suivant, tandis qu'il concerne tous les cas de syllabification à travers les frontières de mots chez Encrevé.

[143] Le phénomène des consonnes euphoniques est souvent rapproché de celui de la liaison, il en est cependant tout à fait distinct.

[144] Les études récentes au sujet de la liaison (Encrevé 1988; Tranel 1995; Tranel 1996) se basent sur la structure syllabique et les règles qui lui sont associées.

[145] Léon (1992) note que les consonnes finales se perdent progressivement à partir du XI ^{e} ou XII ^{e} siècle.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Pour résumer, la liaison se fait dans des groupes de mots étroitement unis syntaxiquement et moins facilement à la mesure que le lien s'élargit. Les pauses suspendent ainsi en principe la liaison [146] .

L'apparition des phénomènes de liaison, selon Catach, N. (1995:215), "en discours est statistiquement faible". Il est assez vain de penser la liaison sans la rattacher à une étude précise de ses occurrences. La liaison facultative est en effet reconnue depuis longtemps comme un phénomène à compter parmi ceux qui présentent de la variation en fonction des registres.

> Les consonnes concernées par les phénomènes de liaison en français sont en petit nombre[148] : [z], [t], [n], [r], [p], [v], et Léon, P (1992:152) remarque que, en langue, 50% des cas possibles de liaisons se présentent avec [z], tandis que [t] et [n] se partagent à peu près également la presque totalité des 50 autres pourcents, les cas en [r] et [p] représentant moins de 1% de l'ensemble[150]. Dans nos corpus apparaissent seulement des liaisons pour les consonnes [z] et [n] la liaison est marquée sur le mot graphique de deuxième position :

Extrait Corpus P8 [147] *PZZ: PBA se vaporise PBB derrière les zoreilles Extrait Corpus F1 *PAD: 'tain tu as fais quoi aux zautre aop? lol Extrait Corpus F8 *PAV: ce serait de l'abus de pouvoir et nous sommes de bons zops

Les liaisons obligatoires sont les plus fréquentes et on ne constate pas de liaisons facultatives marquées dans la graphie. En revanche, il apparaît des liaisons remarquables [148] qu'on a coutume de porter sur le compte de l'hypercorrection [149] . Ces liaisons, dans nos corpus, peuvent être caractérisées comme ludiques et ne sont pas à considérer comme phénomène cognitivement intégré au système de la langue. On en répertorie deux types selon que la consonne de liaison est licite ou incongrue, mais il faut noter que la consonne employée fait toujours partie de l'inventaire des consonnes de liaison.

Extrait Corpus F3 *PBG: pourkoi? y a po de messants zici Extrait Corpus P8 *PAN: sBon les titi nenfants *PAN: moi j y go Extrait Corpus P6 *PA2: ca va les ti namis > =)

Un dernier cas, qui concerne la liaison, se présente dans le corpus. Seule la liaison persiste, l'unité de première position disparaissant :

Extrait Corpus F4 *PAT: il m'a demandé....z'aviez bu ou quoi?? mdr!!!!! Extrait Corpus F7 *PIW: z'avez bientot fini de crier comme ca.... ca donne mal à la tête!!!!!

[146] Sauf Liaison non enchaînée, réputée rare, excepté chez les professionnels de la parole (journaliste, hommes politiques,…)

[147] Les phénomènes auxquels cette section se rapporte sont mis en évidence dans les exemples par l'attribut gras ou souligné.

[148] Au sens de Blanche-Benveniste (1997:30).

[149] On consultera Léon (1984).

en vertu de la loi du droit d'auteur.

Extrait Corpus F7 *PBP: z'ont les yeux brides Extrait Corpus P3 *PAM: zete tarrer :P Extrait Corpus P6 *PJ7: bon allez je me casse, vraiment z'etes puerils

Ce procédé concerne exclusivement les indices de sujet.

Il est évident que la majorité des cas où devrait apparaître une liaison obligatoire reste fidèle au code écrit. Parfois au sein d'une même unité phrastique, ainsi qu'on le constate dans les deux exemples ci-dessous, on est amené à se demander pourquoi une liaison obligatoire n'est pas marquée puisqu'une autre l'est.

Extrait Corpus F1 *PAD: 'tain tu as fais quoi aux zautre a op? Lol Extrait Corpus F7 *PBP: z'ont les yeux brides

L'insuffisance du nombre de cas ne permet pas, hélas, de se prononcer sur la question.

B. ÉLISION

L'élision est très fréquente dans nos corpus. Il s'agit, nous l'avons dit plus haut d'un phénomène marqué dans la graphie.

Cas prévus par la graphie pour la frontière de mot.

Les cas où elle se produit sont bien circonscrits dans la norme orthographique [150] . Lorsqu'un mot grammatical est terminé par une voyelle et que le mot suivant commence par une voyelle, la voyelle finale du premier mot est remplacée par une apostrophe. Ce phénomène concerne exclusivement les items suivants :

pronoms	*je, me, te, se, le, la, ce*
déterminants définis	*le, la*
prépositions	*de*
conjonctifs (composés)	*lorsque, puisque, que*
	si devant *il ou ils*
adverbe négatif	*ne*

Dans nos corpus ce phénomène s'étend à d'autres cas, reflétant la prononciation effective de certains énoncés.

Cas non prévus par la graphie pour la frontière de mot.

Lorsqu'un mot grammatical de la liste précédente est suivi d'un mot qui commence par une consonne, la voyelle finale du mot grammatical peut n'être pas représentée.

Extrait du Corpus F6 *PAW: PDH? humph... s'tu veux... Extrait du Corpus F7 *PBN: jpense Extrait du Corpus F7 *PJQ: PFS checke ton brake à bras dans ton char jcrois tvas glisser hors du channel Extrait du Corpus P3 *PAM: bon jrevien plus tard :) Extrait du Corpus P8 *PBB: brb lgros qui me sert de chum mapelle :)

Lorsqu'un mot lexical est terminé par un 'e' et que le mot suivant commence par une voyelle, le 'e' peut ne pas apparaître :

[150] Riegel, Pellat et Rioul (1994:57).

en vertu de la loi du droit d'auteur.

Extrait du Corpus P6 *PCO: PI7 en a marre des gent qui l'agress avec asv en pv
Extrait du Corpus F6 *PBT: PAW par exemple.... pour la nouvelle année si tu souhaite bonnanné a qqun d'ici en tenant compte de ton heure ca gache tout le punch de la nouvelle année vu que c pas la nouvelle année encore ici

Très fréquemment dans nos corpus, les lettres finales qui n'ont pas de contrepartie phonétique disparaissent.

5.1.3.7. Combinaisons

Pour une même unité, les moyens présentés plus haut peuvent se combiner. Par exemple, la combinaison des moyens d'abréviation (raccourci de forme) présentés ci-dessus intervient dans le corpus. Ainsi, *ptdr* combine à l'initiale le procédé du rébus (ou de la sélection de graphèmes) et celui de la siglaison. L'inverse se trouve aussi, dans des unités telles que *qq1, en tt k* qui combinent l'abréviation par sélection de lettres à l'initiale et le procédé du rébus. *a+* combine la troncation et le procédé du rébus. *aop*, *sop* [151] combinent la sélection de la première lettre de l'unité et la troncation. Il apparaît aussi des combinaisons de ces procédés d'abréviation avec le phénomène de généralisation graphémique : *thx*, *pk*.

5.1.4. Phénomènes phonétiques représentés n'impliquant pas de réduction formelle

5.1.4.1. Réinterprétation

Le phénomène de réinterprétation phonétique d'un mot emprunté pour le faire correspondre au patron de la langue qui les adopte est bien connu, ici il se trouve au plan graphique : des graphèmes appartenant au système du français sont utilisés pour représenter la prononciation de mots empruntés. Les unités copiées affectées par ce procédé sont des unités fréquentes et la plupart ne sont pas d'emprunt récent. La liste ci-après présente dans la première colonne la forme rencontrée dans le corpus, tandis que la colonne de droite rassemble les formes conformes à la norme.

baill baille	bye bye
babaille	bye bye
babye	bye bye
baille	bye
baille	bye
baye	bye
bicose	because
hailleeee	hi
tchao	ciao, tchao
tcho	ciao, tchao

[151] sop et aop valent respectivement pour *super-opérateur* et *auto-opérateur*.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Inversement les utilisations de graphies anglaises pour représenter la prononciation sont rares :

chow ciao, tchao sur le modèle de *how, cow, now*.

Il est peu probable, dans ce cas, que la graphie représente la prononciation la plus courante des 11 unités anglaises terminées en 'ow' empruntées par le français et répertoriées dans le Robert [152] .

5.1.4.2. Allongements en *-euh*

On a coutume de distinguer dans la chaîne parlée entre éléments phonématiques (ou segmentaux dans la terminologie américaine *i. e.* ce qui concerne l'articulatoire) et éléments prosodiques, ou suprasegmentaux (durée, intensité, mélodie). La prosodie est réputée en français porter des indications de segmentation des énoncés, des distinctions de modalités de phrases et jouer un rôle phonostylistique.

Le phénomène ici qualifié d'allongement, marqué dans la graphie des IRC, concerne des propriétés prosodiques. Il s'agit d'allongements en '-euh' et variantes.

Extrait du Corpus P2 *(action): PAF byeuuuuhhhhh lmatchouze m'appelleuuh! Extrait du Corpus P5 *P71: ptaiiiiiiiiiiin j' arrive pas a lancer mon eggdrop ca m'énerveuuuuuh Extrait du Corpus P4 *P17: arf mé c vré le P03 y va me piquer pour que j'ai une fuiteuh Extrait du Corpus P6 *PA8: komment va ma moman ircéenneu

La présence d'un 'e' en fin de mot marque en français standard que la consonne qui le précède doit être prononcée, ce qui n'est pas toujours le cas des consonnes en fin de mot. Ce 'e' n'a donc pas de contrepartie phonétique sauf régionalismes. La littérature relève aussi la prononciation du 'e' dit caduc dans les sociolectes de Parisiens jeunes et défavorisés (Walter, H. 1977) ou encore l'apparition d'un [œ] ou d'un [ø], qu'on a alors peine à nommer caduc, en fin de mots où il n'y en avait jamais eu, et ce davantage chez les femmes que les hommes (Léon, P. 1987, Hansen, A.D. 1991). Fonagy (1989) l'identifie dans le discours "jeune et expressif".

Ce phénomène n'est pas très répandu à travers le corpus. Il intervient en général sur des unités lexicales (*vs* grammaticales) y compris sur des noms propres et dans nos corpus il se trouve toujours en fin de clause. On peut le mettre sur le compte de "manières de parler", comme de l'indication de l'insistance ou encore d'une émotivité particulière telle que l'énervement, l'agacement.

5.1.4.3. Redoublement de caractères

Ce phénomène ne concerne pas seulement les unités linguistiques, nous le verrons à la

[152] Robert, P. *et al.* (1993); *bay-window, bow-window, bungalow, cash-flow, chow-chow, know-how, marshmallow, one man show, sandow, show, slow.*

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

section suivante. Les unités linguistiques affectées par ce procédé sont à 43% des interjections et onomatopées, à 7% des noms propres de personnes et enfin dans de plus faibles proportions des noms, des verbes, des adjectifs, des adverbes et des pronoms.

Il est possible de distinguer deux cas différents : le redoublement de consonnes et le redoublement de voyelles. Il semble en effet difficile de relier dans le groupe des consonnes le redoublement de lettres correspondant à des occlusives orales ou de lettres "muettes" à une quelconque manifestation phonétique courante en français, tandis que le second pourrait être rapproché du phénomène à caractère prosodique rapidement décrit plus haut. 43% des unités affectées par ce phénomène présentent un redoublement de voyelles, 31% un redoublement de consonnes non muettes dont 90% ont le trait [+continu] [153] , 11% un redoublement de consonnes "muettes", 1% sont affectées par les deux premiers types de redoublement et 3% par le premier et le troisième type et enfin 14% présentent un redoublement de syllabes [154] .

On peut alors faire l'hypothèse que l'utilisation du redoublement de caractères est en partie un moyen pour un locuteur de démarquer visuellement son intervention de celles des autres locuteurs, cela intervenant essentiellement sur les noms propres de personne; et en partie un moyen de représenter des caractéristiques phonétiques des énoncés.

Si l'on se positionne dans cette deuxième perspective, le redoublement de caractères indiquerait la durée d'un segment. Rossi *et al.* (1981:50) rappellent que "les variations temporelles sont régies par de multiples facteurs qui correspondent à des niveaux d'analyse différents (paralinguistiques, intrinsèques et co-intrinsèques, linguistiques) et qui font de la durée un paramètre très difficile à interpréter". Si on écarte les variations temporelles intrinsèques et co-intrinsèques, pour le français, la durée est un paramètre impliqué dans le débit (émotions, attitudes, phonostyles, types de discours…), l'accentuation et le rythme (fonction démarcative, expressive, émotive) et l'intonation [155] .

Dans nos corpus, l'interprétation de ce phénomène semble dépendre de l'unité à laquelle il est attaché et véhicule essentiellement de l'expressivité et des émotions. Dans l'exemple suivant, la réponse du locuteur est affectée d'un redoublement de caractère qui véhicule une sorte d'excitation du locuteur.

Extrait du Corpus F7 *PHQ: a va ? %add: PBG *PBG: ouiiiii super merci et toi %add: PHQ

Ce type est à rapprocher de celui qui se porte sur les noms propres vocatifs en séquence d'ouverture, qui exprime l'excitation que suscite chez un locuteur l'arrivée d'un participant particulier. En revanche, le redoublement dans l'exemple suivant, renforce le dégoût exprimé par l'interjection.

Extrait du Corpus F3 *PCA: bahhhhhhhhhhhhhhh !!!!!!!!!!!!!!!!! du foot

[153] Nous incluons là les occlusives nasales bien qu'elles ne puissent appartenir strictement à cette classe n'étant continues que dans la cavité nasale.

[154] Il s'agit à quelques exceptions près d'interjections.

[155] Hirst & Di Cristo (1998:1-44) ; Di Cristo (1998:195-218).

en vertu de la loi du droit d'auteur.

L'extrait ci-dessous, lui, présente le soulignement d'un item et pourrait être rapproché de l'accent d'insistance.

Extrait du Corpus F8 *PDY: faut pas deconner, c'est pas ignooooooble de demander des explications

Cette pratique du redoublement, bien que réprouvée par la netiquette à l'article *flood*, ne fait pas l'objet de réprobation de la part des opérateurs dans nos corpus.

5.1.4.4. Casse des caractères

L'utilisation des majuscules dans les IRC a d'autres fonctions que celles déterminées par la norme graphique. Nous verrons que leur utilisation est "réglementée". Notons que la production des majuscules, sauf à ne choisir que ce mode de frappe, a un coût temporel qui peut expliquer qu'elles ne soient pas utilisées.

La majuscule de début de phrase ou celle qui signale un nom propre sont loin d'être utilisées de façon systématique. 13,05 % des interventions commencent par une majuscule. Si on retire de ce comptage les interventions qui commencent par un nom propre et les interventions qui sont complètement en majuscules le résultat se porte à 4,9 %. D'autre part, 57,12% des noms propres ne reçoivent pas de majuscule. La principale fonction de l'utilisation des majuscules est de mettre en relief tout ou une partie d'un énoncé.

Extrait du Corpus F3 *PAJ: PAG j' ai pas dit TA trompette Extrait du Corpus F6 *PAD: je t' ai dit mon bracelet QUE J AIMAIS BIEN je l' ai gardé pendant les vacances.... Extrait du Corpus F8 *PAC: ah oui DEPUIS WINDOWS oui.... Extrait du Corpus P6 *PCO: PR2005 prend son MEGA PHONE!

L'utilisation des majuscules peut être un moyen pour un locuteur de démarquer visuellement son intervention de celles des autres locuteurs.

Extrait du Corpus F7 *PED: VOUS DORMEZ? Extrait du Corpus P8 *PBH: SALUT TOUT LE MONDE

Cet emploi est réprouvé lorsqu'il n'a d'autre justification ; la netiquette réglemente en effet l'utilisation des majuscules et un certain nombre d'interventions font référence à l'utilisation des majuscules dans le canal.

Extrait du Corpus P4 *P30: BONJOURS %add: ALL *P30: J'ABITE A ROME %add: ALL *P13: baisse tes caps stp. %add: P30 Extrait du Corpus F1 *PBW: pas de majuscules c'est comme si tu criais et on n'est pas sourds...

Il est difficile de dire, pour un énoncé, si l'utilisation des majuscules mime un caractère prosodique (insistance, force de la voix) ou si elle est un moyen de démarquer visuellement tout ou partie d'énoncés, puisque les fonctions de ces procédés (attirer l'attention sur soi, sur une partie de l'énoncé, exprimer un sentiment, une émotion…) sont en intersection.

5.1.4.5. Interjections, onomatopées

Notre corpus rassemble un nombre important de formes (1621 au total) que la grammaire traditionnelle nomme des interjections ou des onomatopées. À la suite de Gadet (1989), Bruxelles et Traverso (2001:39) rappellent que *"Toute étude de corpus authentiques*

en vertu de la loi du droit d'auteur.

rencontre bien vite les "petits mots" [156] *parce qu'ils prolifèrent dans les productions orales dont ils sont une des caractéristiques les plus saillantes."* C'est une propriété de nos corpus que d'utiliser ces unités, nous postulons cependant que cela se fait dans de moindres proportions que dans les polylogues oraux spontanés, dans la mesure où le dispositif ne connaît pas les mêmes contraintes.

Le principe qui semble fonder ce groupe d'unités est qu'en dehors de leur emploi autonymique, elles n'entrent pas dans la construction de syntagmes, elles manifestent en effet une certaine autonomie syntaxique et peuvent former à elles seules un énoncé. Elles semblent donc appartenir à un niveau d'analyse spécifique, toutefois, la littérature leur reconnaît le statut d'unité linguistique.

En partant de l'inventaire des grammaires traditionnelles, dès le premier abord, on est forcé de reconnaître l'hétérogénéité [157] et tenté de proposer rapidement une première distinction [158] entre les unités utilisées pour représenter des bruits, des cris,… et habituellement qualifiées d'onomatopées, de celles utilisées pour indiquer la structuration discursive inter- ou intra-locuteur. Il faut effectivement reconnaître que toutes ces unités ne sont pas de même nature et qu'elles ne servent pas les mêmes fonctions.

Les formes que nous trouvons dans nos corpus ne respectent bien évidemment pas la norme représentée dans les dictionnaires, on s'en convaincra par le seul nombre de formes strictement différentes repérées, soit 451 pour 1632 occurrences. Bien entendu, ces unités sont affectées par les phénomènes détaillés jusqu'ici, après normalisation, on se trouve donc devant un inventaire, plus restreint, de 134 unités.

1493 lignes contiennent au moins une de ces unités (68 lignes en comportent plus d'une). On constate également qu'une part importante de ces lignes (553) ne sont constituées que d'unités de ce type et fonctionnent de façon autonome, ce qui indiquerait qu'elles participent à la régulation de l'activité communicative.

Pour chacune des unités, on ne peut le plus souvent repérer une valeur unique. Ce fait est depuis longtemps accepté dans la littérature, d'autant qu'à l'oral, des intonations différentes produisent des effets différents [159] . On peut se demander comment à l'écrit se retrouvent ces valeurs. Constate-t-on une moindre polyvalence des termes, ou non ; et dans ce cas, existe-t-il des indices sur lesquels le récepteur peut s'appuyer pour effectuer un décodage correct. Nous présentons rapidement les différents emplois de ces unités apparaissant dans nos corpus, avant de proposer une typologie.

Il est évident que les remarques que nous allons faire sont en très grande partie dictées par nos intuitions de locuteur natif, et qu'elles se basent sur une lecture qui cherche la cohérence. Les participants des canaux produisent et interprètent leurs

[156] Les auteurs empruntent ce terme à Bouchard (2000), on trouve encore dans la littérature "marqueur", "particule"… (voir notes suivantes).

[157] Barbéris (1992); Vassileva (1994).

[158] Mosegard Hansen (1998); Roulet, Auchlin, Moeschler, Schelling et Rubattel (1987) ; Gülich (1999).

[159] Morel et Danon-Boileau (1998).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

interventions *on-line*, tandis que notre point de vue est global.

Certaines unités présentent beaucoup plus d'occurrences que d'autres, c'est le cas de *bon, ah, bah, ben, haha, héhé*. On remarquera aussi que la standardisation opérée en ce domaine pose parfois problème. En effet, les phénomènes mentionnés dans les sections précédentes, gênent parfois la standardisation. Cela intervient essentiellement sur l'opposition que nous avons choisi de conserver entre *bah* et *ben*, puisqu'en effet on doit se demander si une unité qui apparaît *ba* doit être standardisée *ben* ou *bah*. Il apparaît dans nos corpus, après observation qu'il n'y a pas dans l'usage des participants une réelle opposition formelle. Dans des cas ou on attendrait *ben*, *bah* est effectivement réalisé, en revanche, le contraire ne se produit pas.

Nous examinons assez rapidement ci-après le fonctionnement des unités les plus fréquentes.

ah est produit après la prise de connaissance d'une information nouvelle ou le rappel d'une information connue mais non prise en compte, qui modifie en général le point de vue du locuteur.

Extrait du Corpus F3 *PDL: c'est koi mdr? %add: PAL (1) *PAJ: mort de rire %add: PDL (1) *PCC: mdr merde en abreje %add: PDL *PDL: ah ok %add: PAJ PCC
Extrait du Corpus P2 *PAY: puteaux c ds le 91,93,92? %add: ALL (3) *PAG: c'est a la defense %add: PAY (1) *PAG: je sais je m'y suis perdu. %add: PAY (1) *PAY: ah vi exact %add: PAG

Il peut aussi signifier l'état de surprise du locuteur, son admiration, son soulagement ou sa déception.

Extrait du Corpus F7 *PBP: PJQ c la crise :o) %add: PJQ *PJQ: PBP ah?? où ça? %add: PBP Extrait du Corpus P3 *PBX: la migraine est partie spour ca! %add: PAS, PAM, *PAM: ahh %add: PBX Extrait du Corpus P9 *PBH: Ah merde t'es sur lyon... désolé PBO, toutes mes condoléances =\ %add: PBO

Les formes *ha, rha* traduisent en général des protestations. Dans l'exemple ci-dessous, PBW vient d'accorder le statut d'opérateur à PAG, ce qui implique que ce dernier veille au bon déroulement des choses sur le canal.

Extrait du Corpus F1 *PBW: y'a pas d'raison que je sois seule à bosser ;) %add: PAG *PAG: rha %add: PBW

grr manifeste l'agacement, le mécontentement du locuteur :

Extrait du Corpus F7 *PCK: BobLeCingle ma PARLE PAS DE MA VOITURE AUJOURD'HUI STP!!!! %add: PDS (1) *PCK: grrrrrrrrrrrrrrrrr %add: PDS

oh marque tantôt la réprobation, le mécontentement du locuteur émetteur, son état de surprise ou de choc, par rapport à un énoncé ou à un événement (extrait suivant),

Extrait du Corpus F7 *PJQ: oh le vilain power abuse... %add: ALL

tantôt son admiration, son contentement :

Extrait du Corpus P3 (PBR vient d'accorder le statut d'opérateur à PCA) *PCA: ohhhhh merci mon tit PBR :o) %add: PBR

houla, houlala surviennent après des dires ou des événements inattendus ressentis comme négatifs, ils semblent manifester la surprise et éventuellement une certaine méfiance.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Extrait du Corpus P6 *PD9: oula un intello dans le chan help %add: ALL

tiens manifeste un état de surprise du locuteur sans qualifier l'événement inattendu. Il apparaît éventuellement redoublé : *tiens tiens.*

Extrait du Corpus F5 *PAL: salut PAV %add: PAV (1) *PAV: tiens lu pit %add: PAL

Il précède le plus souvent l'expression de ce qui suscite la surprise.

Extrait du Corpus P8 *PBB: tien Lambic a une amie :P %add: ALL, (PAW),

Il convient aussi pour modifier la thématique conversationnelle. L'extrait ci-dessous est tiré d'un passage qui réfère aux cinémas de New York.

Extrait du Corpus F7 *PJQ: tiens j'irais bien me faire un week en dà NYC %add: PBP

tiens sert rarement à marquer que le dire est une conséquence du comportement de l'allocutaire, ce qui le rapproche de la sémantique de sa source.

Extrait du Corpus F5 *PAY: savoir aimer c'est ne pas aimer et aimer c'est ne pas savoir!! et en plus Si tu veux le p'tit dej' au lit, dort dans la cuisine!.. Tiens!!!
Extrait du Corpus P7 *PAC: héhé doucement PBH :)) %add: PBH (1) *PBH: bin la PAC, je vais me gener tiens lol %add: PAC

Cette unité peut être combinée avec *ben.*

Les travaux sur bon sont assez peu nombreux. Caron-Pargue et Auriac (1997) regroupent les divers aspects de la signification de bon selon trois composantes : positionnement d'une frontière [160] , franchissement d'une frontière et mode de franchissement de la frontière. La première composante réfère aux fonctions d'ouverture ou de clôture d'unités discursives. La deuxième spécifie la fonction de mise en relation d'unités s'intégrant alors dans une unité plus vaste, ce qui révèlerait un aspect anaphorique de ce marqueur. La troisième composante spécifie enfin la façon dont les unités sont intégrées. Ces composantes permettent aux auteurs de dégager deux aspects distincts de la signification de bon, comme organisateur de contenu conceptuel, ou marqueur des étapes de construction du terrain commun entre les interlocuteurs. On peut alors établir deux dimensions de cette unité, fonctionnant comme connecteur pragmatique (réflexif, expressif ou phatique) ou marquant la structure conversationnelle (conclusif, explicatif, opposition argumentative).

bon, dans nos corpus est le plus souvent utilisé pour marquer une rupture de thématique conversationnelle, ainsi que l'illustre l'extrait présenté ci-dessous.

Extrait du Corpus P3 *PAG: PAS tu parles toujours aussi bien quebecois a ce que je vois %add: PAS *PAS: yep %add: PAG *PAG: ;) %add: PAG *PAG: c l'effet bianca %add: PAS *PAG: lol %add: PAS *PAS: j'edite un dico franco-quebecois ;p %add: PAG *PAS: oui %add: PAG *PAS: bianca-nana %add: PAG *PAS: ;p %add: PAG *PAS: bon je vais debrancher un truc %add: PAG *PAS: ca va peut etre me faire planter %add: PAG

Il est ainsi souvent à l'initiale des amorces de séquences de clôture.

Extrait du Corpus F3 *PDL: bon je crois que je vais partir %add: ALL

Plus rarement *bon* marque que le locuteur fait une concession. Dans ce cas *bon* valide

[160] La notion de frontière est apportée par Culioli, A. (1978, 1988)

en vertu de la loi du droit d'auteur.

une information problématique et signale qu'une réponse va y être apportée.

Extrait du Corpus F3 (action) PAM voulait aussi la surprise woiiiiiiiiin :/ %add: PAG, PCA, (7) *PCA: bon PAM je te la laisse :) %add: PAM

Cet emploi se rapproche de ceux où *bon* est associé à *ah* ou *ben,* dans ces cas, il est une sorte de validateur de l'information qui suscite l'énoncé, *ah* et *ben* contribuant respectivement les valeurs d'étonnement et de conséquence. À travers tous le corpus, *bon* n'est employé que deux fois seul, comme simple validateur d'information. D'autre part, un seul exemple présente *bon* ailleurs qu'en frontière de clause, mais jamais de façon si souple qu'on le trouve dans l'oral spontané :

…et elle euh bon faut dire qu'elle est pas maline hein mais bon elle a pas fait l'truc.

Certains emplois de *alors* sont aussi à analyser dans cette section [161] . En effet, *alors* peut marquer l'ouverture ou la clôture d'unités (changements de thématiques, unités parenthétiques, ruptures de thématique…) ; l'invitation à prendre la parole ou à poursuivre une séquence. Cette unité est quelquefois empreinte d'impatience.

Extrait du Corpus P4 *P05: ca va P03? %add: P03 *P03: tres bien merci et toi? %add: P05 *P05: ouais ca va :) %add: P03 *P05: alors koi de bo? %add: P03 *P03: bah rien de special :) %add: P05

ben et *bah* dans nos corpus sont en intersection. Après une question, ils semblent manifester l'évidence de la réponse ou l'embarras qu'elle cause.

Extrait du Corpus F3 *PCO: tu connais ussi? ;)) %add: PAE *PAE: PAE bah tu m'a donne l'url j'y suis alle %add: PCO Extrait du Corpus F4 *PAB: tain j'arrive pas a copier les chansons *.wav d'un cd sur mon DD Lolol %add: ALL *PCA: té pas doué PAB :)) %add: PAB *PCO: c ormal PAB c normal %add: PAB *PCA: tu le fais avec quoi? %add: PAB *PAB: comment je peux faire? %add: ALL *PCA: avec windac %add: PAB *PAB: ben avec rien c'est ça le probleme PCA %add: PCA

ben ou *bah* peuvent aussi introduire une conséquence [162] (conclusion, déduction), ou une concession.

Extrait du Corpus F3 *PCB: et si je l ai pas %add: PAJ (1) *PAJ: PCB bah c'est pas bien, t'es un pirate :o)))) %add: PCB Extrait du Corpus F4 (PEO a été sorti du canal par PAU, avec le message suivant : "you're repeating too much – text flood [l²]¤") *PEO: bah j'ai pas fait de repeat... se demande si PAU ets pas un peu fru ;) %add: PAU

Seul *bah* peut manifester une forme de désintérêt, résignation ou découragement.

Extrait du Corpus P4 *P05: alors koi de bo? %add: P03 *P03: bah rien de special :) %add: P05

tss et *pff* marquent en général la remise en cause de la pertinence ou du bien fondé du dire de l'allocutaire.

Extrait du Corpus F7 *PCK: PDS jaloux.... lol %add: PDS *PDS: moi? pfffffff %add: PCK

[161] Hasen-Moosegard (1997).

[162] Voir sur ce point Bruxelles *et al.* (2001)

en vertu de la loi du droit d'auteur.

Dans certains cas, ils marquent la réprobation quant à un énoncé, ou un événement. Ils peuvent intervenir seuls, précéder ou succéder un énoncé qui explicite la non pertinence. Ils ne sont combinés à aucune autre unité du type discuté ici.

bof dans nos corpus est utilisé soit en réponse à un énoncé présentant un problème, soit en réponse à une question totale appelant donc une réponse binaire (oui ou non). Dans le premier cas, illustré ci-dessous, *bof* réfute le statut de problème et est suivi dans tous les cas d'une alternative. Dans le second cas, *bof* est une réponse alternative intermédiaire.

Extrait du Corpus F8 *PAS: PDT cela dépend du nb de parisiens qui descendent %add: PDT (2) *PDT: boffffffff PAS... on les trainera pas le fond de culotte s'il faut ;) %add: PAS

Un certain nombre de formes manifestent de l'enthousiasme : *ouais, youpi, yep, yo, et yop,*.

Extrait du Corpus F4 *PDZ: ils sont tous morts? %add: PCA *PAG: youpppppiiiiiiiiii %add: PDZ

On pourrait aussi ajouter à ce petit inventaire *cool,* et *super*, qui servent les mêmes fonctions pragmatico-discurvives, mais qui pourraient être analysées comme des énoncés elliptiques, réalisant une prédication.

ouah est une unité qui connaît de nombreuses variantes formelles. L'ensemble des interventions qui la présentent semble exprimer que le locuteur est surpris agréablement.

Extrait du Corpus P6 *PJ8: mais j'en connaît pas mal des op %add: PE2, PJ7 *PJ8: :) %add: PE2, PJ7 (3) *PJ7: ouah PJ8 connait du bo monde %add: (PJ8) *PJ7: je suis impressionnée %add: PJ8

ouf exprime le soulagement. Il peut éventuellement désambiguïser l'énoncé qui pourrait être considéré comme négatif.

Extrait du Corpus F8 *PDJ: mais si on t'aime %add: PBH *PDJ: :) %add: PBH *PAH: PBH, mais si %add: PBH (3) *PBH: ouf chuis un tipeu rassuré %add: PAH, PDJ,

hmm, hum sont employés pour signifier des plaisirs sensuels réels ou imaginaires.

Extrait du Corpus F5 *PAC: hummmmmmmmmmmmmm j'ai ti hate de voir des marguerites moi %add: ALL

Les mêmes formes et *humph,* ou *humpf* sont employées pour marquer le remplissage d'un temps de réflexion, celles-ci sont souvent suivies de l'expression d'un doute.

Extrait du Corpus F6 *PDH: t'as qua ditre XXX au lieu de YYY %add: PAW (9) *PAW: XXX ? humph... s'tu veux... %add: PDH

hum et *arf* sont utilisés pour manifester une gêne qui est souvent liée au fait que l'énoncé qui suit est un FTA pour celui qui le subit [163] .

Extrait du Corpus P2 *PAY: PAP une question pas simple curiosité féminine arf, t'as quel age? %add: PAP Extrait du Corpus F8 *PBO: moi aussi je t aime :D %add: PAX *PAX: arf %add: PBO *PAX: po moi ;p %add: PBO Extrait du Corpus F6 *PBF: allez PAW arrête de bouder %add: PAW (2) *PAW: y'a pas moyen.... pas avatn d'avoir obtenu ce que je veux.... :) %add: PBF (10) *PBF: arf et tu veux quoi? %add: PAW

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Mais il arrive aussi que *arf* soit utilisé devant un énoncé impliquant que le locuteur avait un présupposé contraire à ce qui est :

Extrait du Corpus F8 *PAS: coucou PDT :))) xxxxx %add: PDT *PDT: allo PBL ;) %add: PBL *PDV: PDT aa mon aaaaaaaaaaaaaammmmmmmmmmmmmmmmmmmoooo oooooooooooooooooouuuuuuuuuuuuuuuuuuuuuuuurrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr enfin qq qui va me sortir de cet ennui %add: PDT (2) *PDW: salut PDT je ne t avais pas vue desolee %add: PDT *PDT: aaallo PDV ;)))))))))] mon chiriiiiiiiiiiiiiiiiiiiiiiiiiii %add: PDV *PDT: allo PDW ;)))))))))) %add: PDW *PBF: coucou PDT :) xxxxxxxxxxxxxxxxxxxxxx %add: PDT *PDV: PDT bon ben j'attends! kes tu me racontes? %add: PDT (4) *PDV: ... %add: PDT *PDV: bon ben je recommence à lire %add: PDT *PDT: allo beau PAS ;)))))))) %add: PAS *PAS: arf... je croyais être encore dans ton ignore list :))) %add: PDT *PDT: mais non beau PAS... jamais ;))))))))))0 %add: PAS

oups revendique une erreur, il est suivi souvent d'une excuse, d'une explication ou d'une réparation (extrait suivant),

Extrait du Corpus F3 *PAE: oops j'ai mal lu %add: PDY

mais il peut être employé par sympathie :

Extrait du Corpus P7 *PAK: merdeeeeeeeee %add: ALL (1) *PBB: quoi PAK? %add: PAK (1) *PAK: mon chum vien de briser ma carte de credit.... %add: PBB *PBB: oops %add: PAK

euh est reconnu comme unité permettant de remplir une pause, à l'oral, permettant de ne pas perdre le tour de parole. Ici le dispositif met à l'abri le locuteur de cette éventualité, puisque le message d'un locuteur n'est accessible aux autres qu'après sa composition. C'est donc la cause de l'utilisation de cette stratégie qui est ici exprimée : hésitation, timidité…

Extrait du Corpus F4 *PCO: bon kelkun connaitrais une façon originale de ce suicidé??? %add: PCO *PDR: heuuuuuuuuuuuuuuu PCO....je vais y penser..... %add: PCO Extrait du Corpus F2 *PAD: euh brb je vais faire ma lessive. à la main s'il vous plait!

Associé à *mais*, *euh* est le signe d'une protestation.

Extrait du Corpus F1 *PCJ: aaaaaaaarf Glossolalie . change de nick où je te vire mdr... %add: PAD *PCJ: vite vite %add: PAD *PAD: mais heu %add: PCJ *PAD: je peux pas lol %add: PCJ

L'extrait ci-dessous montre également que *mais*, employé seul peut assurer cette fonction de protestation.

Extrait du Corpus F3 *PBG: ben moi je pues.. merci %add: ALL *PCM: lol %add:

[163] Face Threatening Acts selon la terminologie de Brown et Levinson (1987) à propos de la politesse. La notion de face est liée à celle de territoire, elle se dédouble en face positive et face négative. Un acte menaçant pour la face d'un locuteur est un acte verbal ou non-verbal qui remet en cause un ou plusieurs aspects de la face positive ou négative, du locuteur qui l'accomplit ou de celui qui les subit. Par exemple, les offres, les promesses, sont des FTA pour la face négative de celui qui les accomplit tandis que les ordres, conseils, sont des FTA pour la face négative de celui qui les subit. Les excuses, autodépréciations sont des FTA pour la face positive de celui qui les accomplit, tandis que les critiques, injures sont des FTA pour la face positive de celui qui les subit.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

PBG (3) *(action): PBG a une tite larmichette à l' oeil *(action): PCM renifle PBG ah oui ca sent si tinsiste :)PPPPPPPP (1) *PCM: mdrrrrrrrrrrrr %add: PBG (3) *PBG: zetes messants! %add: ALL (2) *(action): PCM pas mechante du tout lol (2) *PBG: si euuuuuuuh! %add: PCM (1) *PCM: cets toi qui a dit que tu puais lol %add: PBG (2) *PBG: maiiiiiiiis %add: PCM

D'autres cas d'associations avec *mais* existent dans nos corpus : *non mais, ah mais, non mais ho, non mais bordel*... Ils marquent la clôture d'une unité dans laquelle le locuteur s'est senti défié ou attaqué.

gnak, niarf ou gnark sont produits par un locuteur qui a commis un FTA [164] – un manquement à la netiquette, dans l'exemple ci-dessous– et semblent marquer que le locuteur revendique son acte qu'il sait non conforme aux règles. Il sert en même temps d'adoucisseur, le locuteur ayant commis la faute s'auto-dépréciant.

Extrait du Corpus F7 *PCJ: pei... nantes nantes nantes nantes nantes nantes nantes nantes nantes nantes nantes nnantes nantes nantes nantes nantes nantes nantes nantes nantes nantes nantes nantes nantes nantes nantes nantes nantes antes nantes nantes nantes nantes %add: PEI *PCJ: gnark %add: PJQ

hé est utilisé pour attirer l'attention des allocutaires soit dans le but d'ouvrir une unité conversationnelle (variantes : *hého*, *pst*), ou de modifier sa thématique, soit dans le but de policer les interventions des allocutaires.

Extrait du Corpus P6 *PD1: salut les gars %add: ALL *PD1: y aurais pas des gens intelligents parmis vous? %add: ALL (PD1 se fait sortir du canal avec le message " non que des cons, la preuve") (PD1 revient) *PD1: khoo c t pour rire %add: ALL *PD1: hey vous pouvez m'expliquer un truc %add: ALL *PJ7: PJ7 il a pas trouvé ca drole je crois %add: PD1 (3) *PJ7: tu crois qu'on a envie de t'expliquer qq chose maintenant? %add: PD1

Parmi les interjections impératives, se présente *chut*. Destinée à faire taire le ou les allocutaires, elle ne se révèle pas être ressentie comme une agression de la part de ses destinataires.

Extrait du Corpus P2 *PAY: PAS chut parce que bon, le bb est pas loin %add: PAS

ouh est utilisé pour huer l'allocutaire, à la suite d'un comportement (ou d'une dénotation) évalué comme négatif.

Extrait du Corpus F6 *PBP: sur mon joli COMPAQ p100!!! %add: PDH *PDH: ouh l'antiquite %add: PBP

hein et *quoi* formulent des demandes d'approbation ou de reformulation.

Extrait du Corpus P9 *PAH: PAA avoue que taimerais ca avoir le charme de PBG hein :PPPPPPP %add: PAA Extrait du Corpus P4 *P33: g mal partt en + avec le sport d'hier %add: P13 *P33: donc c la fete koa %add: P13

zou, hop, toc manifestent la facilité ou la rapidité de l'événement dénoté par l'énoncé suivant.

Extrait du Corpus F1 *PAD: bon tiens je te confie la lourde tache de surveiller le

[164] Un acte menaçant pour la face d'un interlocuteur peut-être, comme c'est le cas ici, un énoncé qui, n'étant pas conforme aux règles imposées par l'interlocuteur, remet en cause l'autorité de ce dernier. Les autres cas sont des insultes explicites ou implicites.

en vertu de la loi du droit d'auteur.

nick XXX lol %add: PCJ (2) *PCJ: PAD ouep %add: PAD *PAD: des qu'il est dropper ZOU tu le register si je le fais pas lol %add: PCJ

beurk, piouh marquent strictement le dégoût, ils sont employé seuls à l'exception des cas où la cause du dégoût ne peut être directement inférée par l'allocutaire.

Extrait du Corpus F8 *PBL: ici les meringues elles sont racies et rances %add: PAH *PBL: beurkkk %add: PAH

na est utilisé de façon ludique, le locuteur affirme qu'aucune stratégie conative n'aura de prise sur sa position.

Extrait du Corpus F1 *PAZ: PAD... attends je fais que des stages payés la misère pour prendre 1 tite semaine de vacances... :op alors na %add: PAD

nanana, nananère sont utilisés pour narguer les allocutaires suite à un avantage exprimé ou reçu.

Extrait du Corpus F8 *PBO: moi je suis zentil tout plein et onda le sait bien ;P %add: PAS (3) *PBO: nanananereeeeeeuh %add: PAS

On trouve encore un ensemble de jurons, avec *bon sang*, *bordel*, *criss (Québec* [165] *), merde*, *putain*, *ostie de ciboire(Québec), tabernacle (Québec), zut,* etc. Le total représente 5%, la forme vedette étant *putain.*

On peut le plus souvent retrouver la cause de la production de ces unités. Ces unités sont en effet émises soit en lien avec une intervention d'un allocutaire, ou un ensemble d'interventions, soit avec une ligne ou une série de lignes systèmes, soit avec une activité immédiate du locuteur qui peut être liée ou non au dispositif.

Extrait du Corpus F8 *PAY: 'tain y'a du salut dans tous les coins tss Extrait du Corpus F4 *PAB: tain j'arrive pas a copier les chansons *.wav d'un cd sur mon DD Lolol %add: ALL Extrait du Corpus F5 *PEE: arf merde c bientot l'heure d'aller garder le mostre %add: ALL Extrait du Corpus F7 (une ligne système indique que le bot XXX a quitté le serveur) *PJQ: merde mon bot %add: ALL *PBP: heh PJQ %add: PJQ *PBP: il fait greve aussi %add: PJQ *PJQ: on se sent seuls lol %add: PBP Extrait du Corpus F7 *PFS: criss vous settes ben impoli les francais %add: ALL

Ces unités peuvent manifester de l'étonnement, ou de l'agacement.

Un certain nombre d'émissions sonores humaines, qui ne sont habituellement pas contrôlées, ni linguistiques sont représentées par des onomatopées.

Ainsi peuvent apparaître des pleurs (*snif, bouh, bouhou, ouin*), des cris de frayeur (*hi*).

Extrait du Corpus P6 *PD9: me sent morose en fait %add: PJ7 *PJ7: bon %add: PD9 (1) *PJ7: arf %add: PD9 (1) *PD9: genre coup de blues, envie de rien %add: PJ7 *PD9: snirf %add: PJ7 Extrait du Corpus P4 *(action): P03 tel un poulpe geant a ses bras et ses jambes qui repoussent %add: P17, P13, P19, *P17: hiiiiiiiiii %add: P03, P13, P19, *P17: un extraterrestre %add: P03, P13, P19,

Les formes représentant des rires ou des pleurs sont assez variées, mimant l'intensité de ces derniers.

[165] Drescher (2000), Roulet et al. (1987),

en vertu de la loi du droit d'auteur.

Extrait du Corpus F7 *PBO: bosse un peu fénéant!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! %add: PHQ *PBD: mouahahhaha PBO.... exclavagiste.... %add: PBO

Ces unités interviennent très souvent seules, en réaction à une intervention d'un allocutaire ou du locuteur lui même.

Extrait du Corpus P2 *PAY: il fait chaud tout d'un coup %add: PAC (1) *PAP: PAY pourtant jvien douvrir la fenetre.. :/ %add: PAY *PAP: hehe %add: PAY

On trouve également des représentations d'émissions sonores naturelles humaines, telles que des éternuements (*atchoum,* cf. premier extrait ci-dessous), des ronflements…

Extrait du Corpus F5 *(action): PCZ est en trainde crever %add: ALL (1) *PAB: PCZ comment ca? %add: PCZ *PCF: comment ca? %add: PCZ (1) *PCF: mdrrrrrrr %add: PCZ *PCZ: avec un p de rhumle grippe... qui le fait tousser comme un perdu %add: PAB, PCF (3) *PCF: prend du miel avec de leau chaude et du vicks %add: PCZ *PCZ: qui veut mes virus? %add: PAB, PCF *PCF: nope merci %add: PCZ (3) *PCZ: distribution gratuite %add: ALL, PAB, PCF (2) *PCZ: aaaaaaaaaattttttchoum %add: ALL, PAB, PCF***

Les ronflements (*ronfl ronfl, zzz)* surviennent dans le canal quand l'activité est très réduite. Les bruits de pets (*prout, proutos)* sont parfois employés ludiquement, à la fin d'un énoncé :

Extrait du Corpus F3 (dans ce corpus, PCO est engagé dans une unité conversationnelle dont la thématique est les flatulences, la séquence de clôture en est empreinte) (action): PCO se casse prout

Il arrive qu'ils soient l'occasion de joutes, lorsque l'un des participants commence à jurer, et seraient en ce cas à ajouter au compte des jurons :

Extrait du Corpus P6 *PD3: bah je te dis qu elle l est %add: PI7 *PA2: je veins + que une fois par mois bordel %add: PI7 *PD3: de dieu! %add: PI7 *PJ7: PA2 elle vient PI7 %add: PI7 *PI7: PD3 non je suis po aop %add: PD3 *PJ7: crotte de bique %add: PI7 *PD3: bon bah op moi de dieu! %add: PI7 *PJ7: prout %add: PI7 *PJ7: lol %add: PI7 *PI7: mdr %add: PJ7

Ce genre de joute se produit aussi avec les interjections *ah, oh* voire avec des mots lexicaux.

Extrait du Corpus F1 *PAS: rhaaaaaaaaaaaaaaaa PAG :))))))) %add: PAG (1) *PAG: rhaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa mdr %add: PAS (3) *PBN: hooooooooooooooooooooooooooo %add: PAG, PAS, *PAS: PAG RhaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaAAAAAAAAAAAAaaaaa %add: PAG (2) *PAG: PAS arrêtes sinon je vais te faire un waza %add: PAS (2) *PAG: .... %add: PAS *PAS: PAG Je te mets au défi..... héhé lol %add: PAG (…)

Enfin, *gloups* et *gruik* (bruit de gorges suscités par une émotion) semblent indiquer des contrariétés.

Des émissions sonores humaines non linguistiques contrôlées sont également représentées. Les représentations de chants et de sifflements (*toudoudou*, *fuit fuit)* ne semblent pas véritablement liés à l'activité conversationnelle ; une représentation de chant suit dans nos corpus toujours une ligne de message action produite par le même locuteur. Il arrive en effet que des locuteurs spécifient ce qu'ils sont en train de faire par le biais des lignes d'action – par exemple, écouter de la musique – sans que cela ne puisse

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

se rattacher à aucune unité conversationnelle en cours. L'intervention onomatopéique est alors un prolongement informatif de la ligne d'action.

Des bruits de mandibules (*miam-miam*, *slurp*) apparaissent lorsque des thématiques culinaires ou sensuelles sont en cours :

Extrait du Corpus F5 *PAC: hummmmmmmmmmmmmmmmmmmmmm qui parle de sirop d'érable là :P j'y vais dans quelques jours miam miam menoummmmmmmmmmm :P %add: ALL

Des représentations de bruits de bisou (*smack)* et d'applaudissements (*clap clap*)remplacent le faire, impossible avec ce moyen de communication.

Extrait du Corpus F8 *PAH: PDE, tu peux, j'adore quand tu fais ca ;o) %add: PDE (16) *PDE: quand je fais quoi PAH ??? %add: PAH (5) *PDE: quand je t'embrasse ALBATOR? alors re-belote smack :) %add: PAH

Enfin, quelques émissions sonores non humaines sont représentées. Ainsi, des bruits de chocs (*blam, boum, paf, pan, pof, pouf, poum)* viennent agrémenter la conversation :

Extrait du Corpus P3 *(action): Trois fois, adjugé vendu %add: PAS, PAK, PAM *PAB: mdr %add: PAS, PAK, PAM *PAS: ;p`%add: PAB *PAB: pan %add: PAS, PAK, PAM

Dans l'extrait ci-dessus, l'onomatopée vient renforcer la mise en scène construite par le locuteur PAB, en représentant un événement non linguistique.

bzz-bzz, *bzz* imitent le vol d'insecte, par exemple, le premier est utilisé dans une session pour dire bonjour à un participant dont le surnom est celui d'un insecte, et le second est employé comme verbe

Extrait du Corpus F7 *PFN: ca cause pas les mouches ca bzzzzzzzzz %add: PLJ

Des bruits d'artefacts (*bip-bip, pouet-pouet, pouet*) et des cris d'animaux (*coin-coin*, *meuh*, *miaou*) surviennent également. Ces derniers, ainsi que le pré-babillage (*areuh*) sont souvent employés en remplacement des termes d'ouvertures. Quelquefois, ils ont des emplois plus spécifiques, ainsi que le présente l'exemple ci-dessous :

Extrait du Corpus P3 *PAG: PAS ben oui je ronronne %add: PAS (8) *PAG: miaouuuuuuu %add: PAS

Les onomatopées ont la particularité de pouvoir être employées comme des unités lexicales. C'est le cas également de certaines unités que nous avons classées dans le groupe des jurons. Une autre singularité est qu'elles peuvent prendre des affixes *-os; re-*, par exemple *proutos, repouet*.

Il reste bien sûr un certain nombre d'unités qui ne sont pas mentionnées ici, c'est que leur petit nombre ou leur fonctionnement particulier les met en question. Ainsi, *Bingo,* ou *rebelote,* sont des termes de jeux qui sont utilisés seuls sans référence au jeu. Le premier signalerait que l'allocutaire a deviné une information que le locuteur n'a pas explicitement donnée (mais qu'il ne cachait pas non plus) et *rebelote* signale que la situation s'est déjà produite. Leur utilisation est très rare.

Les interjections formulent des réactions face à de nouvelles situations et/ou elles permettent de fluidifier les transitions conversationnelles.

La spécificité que ces dernières présentent dans nos corpus est en premier lieu leur

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

faible incidence au cours d'un tour de parole ou d'une clause. Elles sont en règle générale des indices de frontières d'unités.

Nous avons observé, ainsi que le remarque Mosegaard-Hansen (1997) que certaines peuvent se combiner avec d'autres, et qu'il existe toujours un ordre pour cette combinaison.

On peut distinguer différentes natures parmi les unités retenues. Certaines sont par ailleurs des unités lexicales, et il ne semble pas qu'on puisse faire de réelle restriction quant à la catégorie morphologique, puisqu'on trouve des verbes fléchis (*tiens, voyons, dites, dis-donc*), des noms (*merde*), des adjectifs ou adverbes (*bon, bien, cool, super*). Dans cette classe, on notera le cas particulier des jurons, qui se distinguent des injures dans la mesure où ils se produisent en réaction à une situation tandis que l'injure est adressée et répond à une syntaxe [166] . Si l'interjection est un signe linguistique, le signifiant est une forme lexicalisée et expressive ou une forme lexicale privée de son sens original. On argue que l'interjection serait un mot sans valeur conceptuelle (pas de sens dénotatif, référentiel) mais signalant un ensemble de situations émotionnelles.

Dans nos corpus, les unités étudiées ont potentiellement trois fonctions : expressive, impérative, représentative, structurante. Certaines peuvent porter plusieurs de ces fonctions dans une même occurrence.

L'information véhiculée est assez diverse. Elle peut concerner l'attitude émotionnelle initiative ou réactive du locuteur, son attitude épistémique initiative ou réactive, la structuration discursive, l'établissement ou le maintien du contact (on constate en effet des cas où un locuteur réagit immédiatement après une intervention par une interjection seule, puis il formule une intervention plus longue, ce qui semblerait être une stratégie de maintien du contact, étant donné la contrainte temporelle et la contrainte de surcharge).

Dans la communication, ce qui concerne les émotions peut être porté par le lexique, l'intonation, le débit, les mimiques, la gestuelle, les postures... Une interjection peut avoir, nous l'avons vu, des valeurs paradoxales [167] , à l'écrit c'est le système d'inférence des intentions et réactions d'autrui qui se met en jeu, aucun indice linguistique [168] ne permet de départager les cas.

On peut distinguer différents types d'émotions, celles qui transparaissent malgré le locuteur, et celles qu'il souhaite faire paraître. Elles peuvent être plus ou moins contrôlées, plus ou moins volontairement dévoilées. Le dispositif des IRC, comme l'écrit traditionnel, permet au locuteur de contrôler tout ce qu'il communique, les émotions qui paraissent donc sont volontairement représentées, "mimées". Tout ce qui se joue dans les IRC se fait par l'échange, verbal ou pseudo verbal, il semble important pour les participants de se construire une identité affective et émotionnelle.

166 Ruwet (1982) ; Milner (1978).

167 Par exemple, *ah* peut exprimer la surprise agréable ou désagréable.

168 Le paralinguistique étant de surcroît exclu.

en vertu de la loi du droit d'auteur.

5.1.4.6. Particularités

Divers types de variation au plan phonétique du français : diatopique (géographique), diastratique (social) et diaphasique (situationnel) sont représentés dans la graphie. Pour ce faire une partie des procédés présentés plus haut sont utilisés, en étant parfois détournés.

- Français québecois

Extrait du Corpus F7 *PJQ: PFS checke ton brake à bras dans ton char jcrois tvas glisser hors du channel %add: PFS *PFS: ah ouain... tu la meme po mon accent man %add: PJQ

(en français québécois, *checke*=vérifier, regarder, *brake à bras* = frein à main)

- Français des étrangers (imitation)

Extrait du Corpus P3 *PAS: maah! les filles elles diiisent rocco il a une grosse biiitte, rocco y fait mal. ma qué dans les films ils preferent les grosses Extrait du Corpus P3 *PAJ: contrerement à mon pseudo je ne souis pas italian Extrait du Corpus P3 *PAS: bcp de tetes connues pour moi ici dans le haut peuple... *PAG: n'est il pas? %add: PAS

- Français infantile

Extrait du Corpus P4 *P33: ze zure ke c vréeuh

- Français soigné

Extrait du Corpus F8 *PAS: autant dire que l'on n'aime plus personne comme ça :) %add: PAH

Dans ces cas, les procédés de représentation phonétiques sont seulement une partie des moyens employés : on constate en effet des choix lexicaux particuliers (*brake à bras, char*…), et des choix de moyens syntaxiques (cf. *n'est il pas* dans l'extrait du corpus P3 ci-dessus : l'inversion de l'indice sujet impersonnel dans la structure interrogative [169]).

5.1.5. Signes de ponctuation

Traditionnellement, sont distingués comme appartenant au système de la ponctuation une dizaine de signes : le point, le point d'interrogation, le point d'exclamation, les points de suspension, la virgule, le point virgule, les deux points, les guillemets, les parenthèses, les tirets. Il est toutefois à noter que dans cette catégorie pourraient être ajoutés les espaces (segmentation en mots), les majuscules de début de phrase, les retours et sauts de lignes, les alinéas et nombre d'autres attributs de mise en forme du texte qui contribuent à la structuration du texte [170] . Notons que certaines de ces formes peuvent être employées

[169] Cette structure, grammaticalement correcte, calquée sur les *question-tags* de l'anglais, est souvent ressentie comme "snob".

[170] Catach (1980)

en vertu de la loi du droit d'auteur.

à d'autres fonctions. Anis *et al.* (1988) nomment l'ensemble de ces formes des *topogrammes* lorsqu'elles sont utilisées pour organiser un texte dans sa séquentialité et pour donner des indications énonciatives (modalités de phrases,…) tandis que les éléments – le point d'interrogation, le point d'exclamation, les points de suspension – qui peuvent être employés sans support, seuls ou combinés "pour exprimer de manière elliptique des valeurs sémantiques proches de celles qu'ils expriment dans leur usage normal" [171] sont dénommés *topogrammes détournés*. Un autre sous-ensemble de ces formes est utilisé pour d'autres fonctions : signalement des abréviations, écriture des nombres décimaux…

Les signes de ponctuation appartiennent strictement au code écrit et n'entretiennent pas un rapport simple avec la structure intonative de l'oral. Ils forment un système de signes graphiques qui contribuent à l'organisation de la séquentialité d'un texte, à sa segmentation en unités.

Ainsi, la virgule permet de former des groupes syntagmatiques, en marquant une frontière entre des termes de même fonction dans la coordination, ou en isolant les groupes fonctionnels en incise ou en détachement. Le point virgule, les deux points et la virgule peuvent séparer des sous phrases liées par une relation sémantique. Le point est utilisé pour segmenter le texte en phrases simples ou complexes essentiellement, cependant des auteurs signalent des usages littéraires contemporains dans lesquels l'encadrement d'unités linguistiques non autonomes par des points et donc détachement d'un élément de phrase a fonction de mise en relief. Les points de suspension sont donnés par la norme comme segmentant le texte en phrases avec prolongement sémantique. Il s'apparente à une sorte de troncation du matériel signifiant ouvrant sur de l'implicite ou sur du non pertinent (par ex. citation tronquée). Les points de suspension sont beaucoup utilisés à l'écrit dans le discours rapporté direct pour marquer par une sorte de métaphore iconique – l'espace physique représente une dimension temporelle – le débit, le rythme dans la chaîne parlée comme signe des émotions qui en sont la cause.

Le point d'interrogation marque les phrases interrogatives directes, il est des moins ambivalents parmi les signes de ponctuation. Avec le point d'interrogation, le point d'exclamation est rattaché au plan énonciatif. Ses valeurs sont si diverses qu'il est difficile d'en donner la liste : surprise, étonnement, perplexité, injonction…

Ces signes de ponctuation sont utilisés à l'écrit d'une façon singulière dans le discours rapporté direct, pour mimer en particulier des pauses de l'oral causées par des réalités diverses. La ponctuation n'étant pas clairement fixée par la norme, l'usage en dispose assez variablement.

Dans nos corpus, l'utilisation de signes de ponctuation est déviante au plan de la forme et de la fonction. Au plan de la forme, on note que les signes peuvent être employés seuls, ou bien en combinaisons, excepté pour la virgule et le deux points. Nous avons relevé 112 types différents de combinaisons de signes de ponctuation traditionnels, ces unités étant elles aussi la cible du redoublement de caractères, qui peuvent être regroupés en quelques classes du point de vue de la forme.

[171] Anis *et al.* (1988:143)

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

- points et suites de points
- suites de points d'interrogation
- suites de points d'exclamation
- combinaisons de points d'exclamation et de points d'interrogation
- combinaisons de points d'exclamation et de points

Seuls 26,5 % des interventions sont marquées par au moins un signe de ponctuation ayant des fonctions syntagmatiques ou discursives. 5,06 % des interventions comportent une virgule effectivement utilisée comme signe de ponctuation. Pour expliquer cette faible apparition, on pourrait objecter que la ponctuation devrait intervenir dans une grande partie des cas en fin d'intervention, et que par conséquent elle est implicite. Cependant, à l'écrit comme à l'oral, la syntaxe d'une phrase interrogative peut être la même que celle d'une phrase assertive et il n'est dans ce cas à l'écrit pas d'autre moyen de désambiguïser que d'utiliser le point d'interrogation. On observera d'autre part, qu'une intervention n'est pas toujours un énoncé fini.

Du point de vue de la fonction de ces signes dans les interventions, on ne retiendra que celles qui sont remarquables.

La majorité des interventions marquées par un point d'interrogation ont une structure ambiguë du point de vue du type de phrase. Et un grand nombre de ces structures, ambiguës quant au type de phrase, ne sont pas marquées par un point d'interrogation.

Des interventions constituées uniquement de signes de ponctuation servent à marquer des réactions non verbales telles que l'étonnement, la perplexité…

Extrait du Corpus P6 *PA1: jleur demande si ils font tourner l' adsl sur macosx il me reponde aplé france telecom %add: PI7 *PI7: qmdr %add: PA1 *PI7: mdr %add: PA1 *PI7: ba qu'es t'a a leur poser des question con aussi :) %add: PA1 *PA1: ... %add: PI7

En ce qui concerne les points, nous parlons de suite de points plutôt que de points de suspension dans la mesure où ces unités comportent de deux à *n* points. Les emplois de suites de points se partagent les valeurs de points et celles de points de suspension décrits par la norme. On peut supposer que des stratégies de facilitation du décodage amènent les locuteurs à utiliser deux points de suite plutôt qu'un seul, ou encore qu'il s'agit d'erreurs de frappe. Les suites de points ont aussi une fonction suspensive ainsi que le montrent les exemples suivants :

Extrait du Corpus F4 *PAM: JE BOSSE MOI nan mais ho lol %add: PAG *PAM: fau pas croire %add: PAG (2) *PAG: PAM ben c'est ce qu'on fait.... on croit pas %add: PAM Extrait du Corpus F3 *PZZ: PCK doit aller chez l'ophtalmo depuis .. 2 ans

Dans les IRC, pour se signaler comme présent dans un canal, un participant peut utiliser la commande *away* [172] . Certains utilisent des ponctuations pour signaler qu'ils sont de retour devant leur clavier et disponibles. L'envoi d'un message constitué d'un point seulement, par exemple, a l'avantage de faire apparaître le surnom de son émetteur dans l'espace d'activité du canal *i. e.* la zone sur laquelle se concentre le plus l'attention de tous

[172] Cette commande fonctionne sur le mode *on/off.*

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

les locuteurs participants et lui permet donc de susciter probablement plus de réactions à son égard. Mais les caractères de ponctuation employés comme phatiques ne relèvent que d'un phénomène marginal.

Dans nos corpus, l'usage du deux points et du point virgule sont presque exclusivement réservés à la composition d'unités particulières : les frimousses, dont nous traitons à la section suivante. La restriction s'applique au deux points qui est quelquefois utilisé pour signaler que ce qui suit découle de ce qui précède, il fonctionne alors comme un présentatif, un introducteur d'explication.

Extrait du Corpus P6 *PCO: PE7 se casse: il a cours de math! Extrait du Corpus P3 *PZZ: PAS fait une vois voix effemine efféminée: ouiiiiiiii

Enfin, on remarque à travers les corpus un faible marquage du discours rapporté direct. Les locuteurs se citent assez souvent les uns les autres et cela revêt une forme particulière, dans des cas précis. Si les paroles à rapporter ont été proférées dans la même session, le locuteur citant utilise une petite série de commandes qui permet de recopier les paroles à rapporter (le texte est rigoureusement le même) ainsi que le surnom de leur émetteur.

Extrait du Corpus F3 *PCA marchez bien sur els clous (7) *PED <PCA> marchez bien sur els clous

Dans l'exemple ci-dessus, le locuteur répète seulement les paroles d'un autre, mais le plus souvent, un énoncé original est lié à la citation, par exemple le locuteur citant estime que le locuteur à qui il s'adresse ne saura pas à laquelle des interventions il réfère, ainsi qu'il en est le cas dans l'extrait présenté ci-dessous.

Extrait du Corpus F5 *PBE: c'est quoi l'extension des domaines marocains...tunisiens? *PBE: .ma ? (10) *PAX: <PBE> c'est quoi l'extension des domaines marocains...tunisiens?----------que veux-tu dire? %add: PBE

Il s'insère alors le plus souvent un groupe de caractères séparateurs, ainsi qu'on peut l'observer dans l'exemple précédent, entre la citation et la partie originale de l'énoncé. Ce groupe de caractères peut être une série de tirets, une série d'espaces, ou des caractères combinés pour former une flèche ayant alors de surcroît une fonction de monstration.

Extrait du Corpus F8 *PAV: <PEK> donne moi un voice <------ encore un qui croit que je suis à son service

Dans le cas où la quantité de parole à rapporter est minime, ou lorsque les paroles à rapporter sont temporellement éloignées, cette fonctionnalité n'est pas utilisée (le texte est modifié).

5.2. Groupes de caractères

5.2.1. Définitions

Une des caractéristiques les plus connues de l'activité communicative non formelle en réseau est qu'elle met en œuvre des groupes de caractères alphanumériques. Ces

en vertu de la loi du droit d'auteur.

unités, d'après les informations dont nous disposons, sont apparues avec l'écrit électronique. Ces groupes de caractères sont nommés, par les utilisateurs eux-mêmes, des *smileys* et c'est là le seul terme que nous avons rencontré dans notre corpus pour les désigner. D'autres noms sont employés pourtant en français, par des locuteurs qui se disent soucieux de défendre [173] la langue française : *souriards, émoticones, binettes, trombines, frimousses* [174] …

Les noms donnés à ces unités laissent deviner leur forme et/ou leur fonction, ainsi du nom anglais *smiley* dont l'adaptation pour le français est *souriard*.

Nous commencerons cette section par un rappel du principe de construction de ces unités et une mise en évidence de leur nature iconique. Puis nous exposons une petite enquête que nous avons menée à leur sujet en dehors de nos corpus. Nous comparons ensuite les résultats obtenus, avec ceux que nous obtenons sur notre corpus, en essayant de dégager les caractéristiques majeures de l'emploi de ces unités dans nos corpus.

5.2.1.1. Principe de construction des unités

Le principe de construction des unités est simple : une combinaison de caractères qui s'interprète en effectuant mentalement une rotation de l'image de 90° vers la droite.

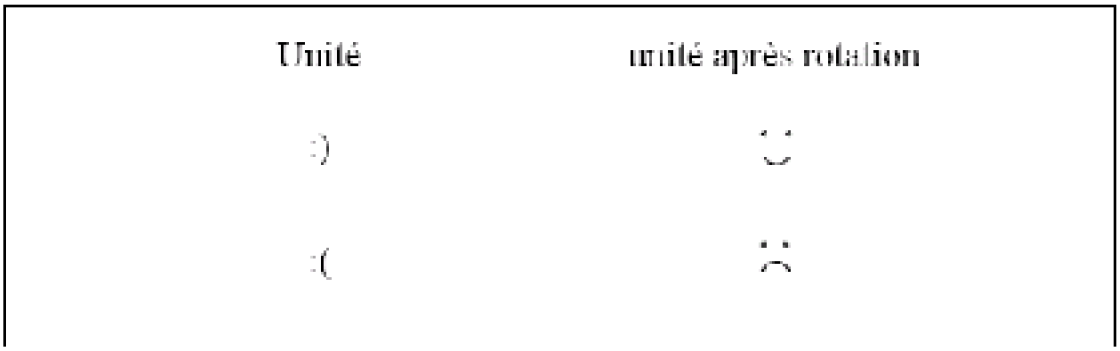

Illustration (5-1) – Exemples d'unités

Il se trouve quelques unités qui ne respectent pas cette consigne (*i. e.* qui ne nécessitent pas d'effectuer de rotation mentale), nous en présentons quelques-unes Illustration (5-2) :

[173] Entre autres leur position est qu'il vaut mieux utiliser des termes existants dans la langue (traduction) plutôt que copier le mot étranger (en particulier anglais).

[174] Le terme retenu par la DGLF est *frimousse*, tandis qu'Anis, J (1998) lui préfère *trombine*.

en vertu de la loi du droit d'auteur.

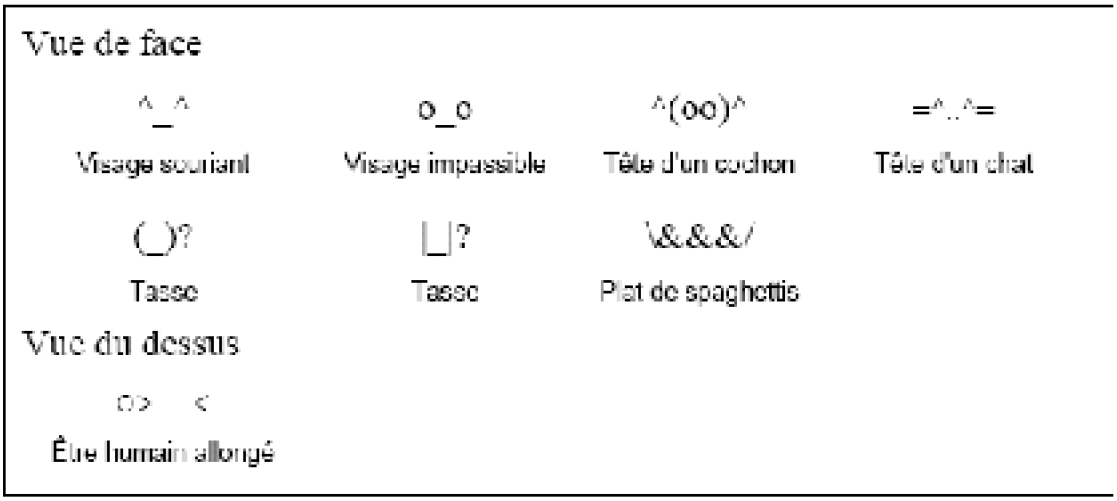

Illustration (5-2) – Exemples d'unités ne répondant pas au principe de base

Ce procédé est connu au-delà de la construction de *smileys* sous le nom d'art ASCII [175] et est utilisé pour pallier le manque de possibilités graphiques de la transmission numérique de texte.

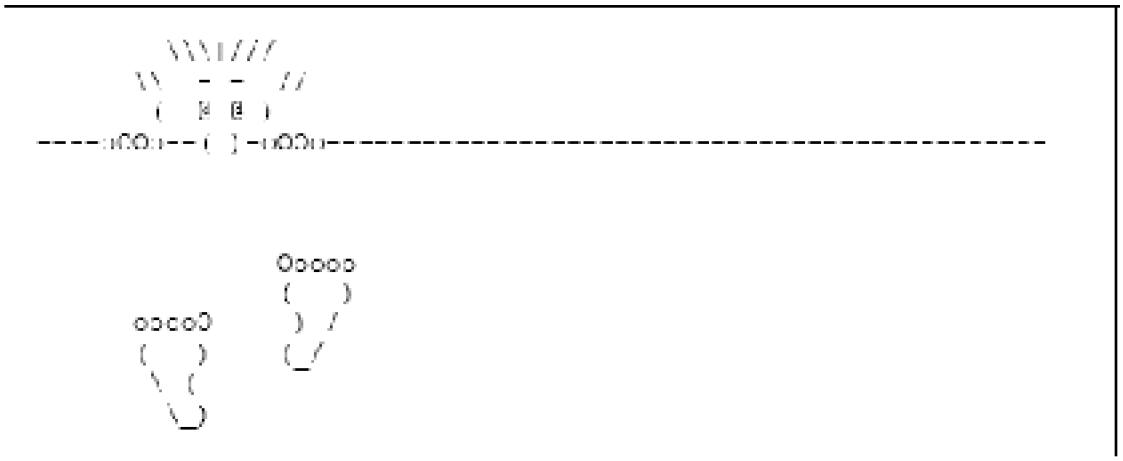

Illustration (5-3) – Exemple d'Art ASCII

Quelques unités semblent manifester une représentation acoustique : *xox, xxx*. Elles apparaissent presque exclusivement dans les séquences d'ouverture et de clôture.

On peut donc dégager trois types d'unités en fonction de leur forme

- Représentation de mimiques
- Représentation d'objets ou êtres
- Représentation de sons non verbaux

[175] Utilisation de caractères ASCII (American Standard Code for Information Interchange – "code américain standard pour l'échange d'informations"), avec une police de caractères à empattement fixe, pour dépasser les impossibilités graphiques. Un espace écran ou page est alors une trame sur laquelle on dispose des caractères vides (espaces) et des caractères pleins comme dans la technique du point de croix.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Ces types d'unités manifestent une évidente iconisation du non-verbal.

La netiquette n'est pas normative sur la question : chacun est libre d'inventer de nouvelles formes, leur succès étant ensuite sans doute lié à leur iconicité (facilité d'interprétation), leur utilité (fréquence) et la rapidité de leur production.

Dans notre corpus on repère quelques cas explicites de non compréhension de ces unités (nous disposons les unités en question en gras dans les exemples ci-après), qu'elles fassent partie ou non des unités fréquentes :

Extrait du Corpus F3 *PCC: si je continu je vais me faire dessus . %add: PCO ... *PAJ: PCC ça se soigne ;(%add: PCC *PCC: qqc te derange PAJ ..... %add: PAJ *PAJ: PCC non pkoi ? %add: PCC *PCC: je sais pas j ai du mal comprendre ton smiley %add: PAJ

et des cas de création explicite d'unités :

Extrait du Corpus F7 *PFZ: eh les gars j' ais invente 1 nouveau smiley ! et le voila : {:A) %add: ALL *PFZ: heheh %add: ALL *PJQ: essaye celui là, ':4o) %add: PFZ *PJQ: c'est pour les clowns %add: PFZ *PFZ: hheheh PJQ non le tien c pour les petits mais le mien es specialement cree pour s'amuser ! %add: PJQ

Au principe de construction de base, il faut ajouter, d'après nos observations l'amplification par redoublement du caractère terminal constitutif d'une forme et la réduction formelle par suppression d'un caractère central.

5.2.1.2. Dictionnaires de *smileys* sur le Web

On trouve sur le Web quantité de pages à propos des *smileys*. Ces dernières présentent le principe général de création des frimousses et des listes d'items accompagnés de leurs significations. Il se trouve également des recommandations fonctionnelles quant à leur utilisation. Les trois extraits de pages présentés ci-dessous sont des exemples types de présentation de listes de *smileys*.

Dits aussi en français souriards, binettes, trombines ou émoticons, les smileys permettent de mettre de l'intonation dans un mail et d'ajouter nuances ou précisions aux propos. http://www.arobase.org/ecole/abrev.htm Ces symboles sont utilisés un peu partout sur Internet pour exprimer un sentiment, une émotion à un moment donné.

http://gla.ecoledoc.lip6.fr/~moinet/Net/Smileys/body_smileys.html Les smileys ou EmotIcons sont des petits bonhommes qui représentent l'émotion de l'utilisateur. Il suffit de pencher sa tête à gauche pour voir une petite tête. Ils sont utilisés pour mettre des nuances dans un message. http://www.hiersay.net/smileys.asp

On notera la forte tendance présentée par ces extraits à soutenir que les *smileys* ont fonction de "représentation des émotions". Nous observerons cependant dans les "dictionnaires" de *smileys* que les gloses s'éloignent de la stricte référence à l'émotion.

Pour ces observations, nous avons recherché les pages Web se présentant comme des dictionnaires de *smileys*. Nous avons sélectionné les pages pertinentes figurant parmi les dix premiers éléments des résultats de la recherche sur trois moteurs de recherche couramment utilisés par les internautes français : Altavista®, Voilà® et Google™ [176] .

[176] Rappelons que le moteur de recherche Yahoo, très utilisé par les internautes utilise les mêmes index que Google.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Cette sélection présente donc les pages potentiellement les plus consultées à un moment donné [177] par les internautes francophones, même si elles ne sont pas forcément les meilleures [178] .

Tableau (5-13) – Références des dictionnaires de frimousses observés

Titre de la page	URL	Nb. d'items
Les sMiLeYS	http://perso.wanadoo.fr/stephanie.crottet/smileys.htm	126
smileys	http://gla.ecoledoc.lip6.fr/~moinet/Net/Smileys/body_smileys.html	136
Les smileys	http://pages.infinit.net/duotang/smileys.htm	52
Les smileys	http://smileys2000.online.fr/	149
Les smileys	http://www.arobase.org/ecole/abrev.htm	8
Quelques Smileys	http://www.cybercable.tm.fr/~bretzner/smileys.html	120
Les smileys ou EmotIcons	http://www.hiersay.net/smileys.asp	117
Dico des smileys	http://www.linux-france.org/prj/jargonf/general/smileys.html#liste	618
Les smileys :-)	http://www.multimania.com/madmac1/smileys.html	134
Le dictionnaire des smileys	http://www.chez.com/talou/smileys.htm	125
Smileys ou emoticons	http://www.leucemie-unle.org/listes_diff/smileys/smileys_frame.html	321
	Total	1900

Ces onze pages présentent 964 *smileys* différents. La première remarque au sujet de ces listes est leur non concordance : les listes ne sont pas similaires ni du point de vue des entrées, ni du point de vue des gloses associées. D'une part, au sein d'une même page peuvent se présenter des *smileys* identiques pour des descriptions sémantiquement différentes, il se manifesterait donc des phénomènes de polysémie/homonymie reconnus. Sur une même page toujours, des descriptions identiques peuvent être accordées à des items différents. Et d'autre part, sur l'ensemble des pages observées, une même unité peut recevoir des acceptions différentes, ou encore des items différents peuvent être associés à des descriptions identiques, ce qui révélerait donc des incohérences. Enfin les unités polysémiques/homonymiques ou synonymes ne sont pas reconnues comme telles sur toutes les pages observées. Un très grand nombre d'unités n'appartiennent qu'à une seule page et un très grand nombre d'entre elles portent des significations si particulières qu'on peut se demander si elles sont véritablement utiles et utilisées.

La comparaison des listes de *smileys* permet de montrer qu'il naît une sorte de mythologie autour de ce phénomène. La classification des items présents dans le corpus montre que la profusion de *smileys* présentés sur les pages Web ne reflète pas une réalité. Nous concentrerons dans un premier temps notre regard sur les items communs à

[177] Les résultats de requêtes identiques sur le Web peuvent varier grandement dans le temps en raison de l'évolution et de la dynamique constante de cette partie d'Internet.

[178] Une des problématiques liée au réseau Internet est la question de la validation de l'information.

en vertu de la loi du droit d'auteur.

cinq pages au moins.

Tableau (5-14) – Liste des *smileys* communs à cinq pages au moins – la colonne de gauche contient les frimousses, celle de droite comporte le nombre de pages sur lesquelles l'unité se trouve

en vertu de la loi du droit d'auteur.

;-)	10
:-D	10
:-/	10
:-)	10
:-X	9
:-p	9
:-o	9
:->	9
:*)	9
:'-)	9
:-(	9
:-S	8
:-{}	8
:-{)	8
:^)	8
:-@	8
:-?	8
:-*	8
:'-(	8
:-&	8
8:-)	7
{:-)	7
[:-)	7
@:-)	7
:-E	7
:-9	7
:-7	7
:-[	7
:)	7
:(	7
:-#	7
*:o)	7
(:-)	7
(-:	7
X-(	6
C=:-)	6
B:-)	6
B-)	6
8:]	6
8-)	6
3:-o	6
>:->	6
=:-)	6
}:-(	6
|-O	6

|-I	6
@=	6
:-Q	6
:-F	6
:D	6
:-C	6
:-~)	6
:-|	6
:-)~	6
:-	6
%-)	6
-0,06	6
8)	5
0:-)	5
>;->	5
<:-I	5
[:]	5
:v)	5
:O	5
:-I	5
:8)	5
:>	5
:<	5
:]	5
:@	5
::-)	5
:/7)	5
:*	5
-:-)	5
':-)	5
-:-(	5
:-'	5
.-)	5
,-)	5
*<:-)	5
(:I	5

Les stratégies de définition des items sont hétérogènes tant à travers les dix pages utilisées qu'au sein d'une même page.

Il est à noter dans un premier temps qu'il n'y a pas d'illustration des emplois. La paraphrase est quelquefois utilisée "*Euh... Tu baves, là, j'crois !"*et semble indiquer que les *smileys* ne représentent pas seulement des émotions. D'autre part, dans ces listes, les descriptions s'entendent en émission ou en réception : "l'utilisateur se moque de vous", "on se moque" ; "Je tire la langue" ; "smiley (vous) tirant la langue". Certains *smileys* sont

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

décrits comme autocentrés (dire quelque chose de soi), d'autres sont présentés comme allocentrés (dire quelque chose de l'allocutaire). Des items sont décrits en eux-mêmes : "smiley orateur", "vue de 3/4 d'un elfe" pour ce qu'ils représentent icôniquement. Pour certains items on ne fait mention que de leur utilisation, par exemple : "à prendre au second degré", "employé avec des phrases ironiques". On se rendra compte de ces phénomènes à la lecture du tableau suivant qui est une compilation des pages mentionnées plus haut. Ce tableau reporte les mentions exactes des pages observées, nous avons simplement conservé une seule des descriptions lorsque plusieurs descriptions similaires se présentaient du type "l'utilisateur se moque", "on se moque" et uniformisé les descriptions (formes à l'infinitif). Les parenthèses signalent des descriptions doubles : lorsqu' a été trouvé sur une page "bavard" et sur une autre "être bavard" nous avons rassemblé ces descriptions en "(être) bavard".

Tableau (5-15) – Compilation des descriptions de frimousses

en vertu de la loi du droit d'auteur.

Frimousses	descriptions relevées	Nb.
:-D	rire	
	moquerie	
	parler trop	
	gaucher qui vous rase	15
:-o	crier	
	être étonné	
	naissance	
	chanter	
	bailler	15
:-p	(être) bavard	
	(confier un) secret	
	tirer la langue	
	langue a fourché	
	camarade	14
:->	être content	
	être diabolique	
	rire beaucoup	
	rire sarcastique	
	être moqueur	13
:-X	ne rien pouvoir dire / motus	
	bisous	
	nœud papillon	13
:-(	ne pas être content	
	être triste	
	faire la gueule	
	être déçu	
	expression de désapprobation	12
:-)	rire	
	comédie	
	être heureux	
	« employé avec des phrases ironiques »	
	sourire	
	bonne humeur	12
:-E	avoir les dents pourries	
	avoir de belles dents	
	être déçu	
	avoir des dents proéminentes	
	être un vampire avec dents proéminentes	
	être un vampire aux jolies dents	12
;-)	clin d'œil	
	« à prendre au second degré »	
	s'être fait cogner à force de raconter des blagues	
	s'être fait cogner	12
:-*	après avoir mangé quelque chose d'aigre ou amer	

en vertu de la loi du droit d'auteur.

	faire un bisou	
	(avoir des) nausées	
	se tromper	11
:-/	pas ma faute	
	être sceptique	
	sourire forcé	
	être un gaucher indécis	11
8:-)	(être) une petite fille	
	avoir ses lunettes sur le front	
	(être) sorcier	
	(être) mickey mouse	11
:-#	(porter un) appareil dentaire	
	avoir pris un coup dans la bouche	
	motus	
	avoir dit une parole de trop	10
:-&	être muet	
	être en colère	
	être très fâché	
	manger des spaghettis	
	avoir la langue nouée	
	avoir la langue bien pendue	10
:)	être heureux	
	sourire	
	(être) lépreux	10
:*)	être ivre	
	porter la moustache	
	être euphorique	10
:-?	fumer la pipe	
	être fumeur	
	être dubitatif	10
:^)	(avoir) le nez cassé	
	être ivre	
	(avoir) un grand nez	
	vue de 3/4 d'un elfe	10
:-S	"Après un commentaire BIZARRE"	
	incohérence	
	raconter n'importe quoi	10
:'-)	pleurer de rire	9
:-[	crétin triste	
	être un vampire	
	être un imbécile / crétin	
	faire la moue	9
:-{)	porter la moustache	9
:-I	faire un demi sourire.	
	hmm	

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

	indifférence	
	être pensif	9
:-Q	être fumeur	
	essayer de toucher son nez avec sa langue	9
{:-)	porter une perruque	
	avoir une coiffure à la mode	
	avoir une raie au milieu	
	avoir une nouvelle coupe de cheveux	9
(-:	australien	
	être gaucher	8
(:-)	(porter un) casque de moto.	
	être chauve	
	(avoir une) grosse face	
	être agréablement surpris	
	être un crâneur	8
:(	ne pas être content	
	être triste	
	être déçu	8
:'-(	pleurer	
	être triste	8
:-@	crier	
	jurer de dire la vérité	
	porter une barbe permanentée	8
:-{}	avoir du rouge à lèvres	8
:-|	avoir une journée ordinaire	
	être indifférent	
	être sévère	
	se lécher les babines	8
:-7	fumer la pipe	
	dire un secret	
	"je me suis forcé pour l'écrire"	
	"juste après avoir fait une déclaration désabusée"	
	Rire jaune	8
:-C	être incrédule	
	être un clochard	
	être vraiment fâché	
	être fainéant	
	être très très malheureux	8
|-O	bailler	
	ronfler	
	être fatigué	
	naissance	8
8-)	porter des lunettes de soleil	
	porter des lunettes	
	smiley nageur	

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

	être tout excité	8
*:o)	Bozo le Clown	7
:-	avoir la bouche cousue	
	être un homme	
	neutre	
	neutre, "remarque ambiguë"	
	Programmer en PROLOG	7
:-~)	avoir un rhume	
	avoir froid	7
:-9	se lécher les babines	
	siffler	7
:-F	s'être cassé une dent	
	tirer la langue	
	être un vampire auquel il manque une dent	7
@:-)	Indien (de l'Est)	
	porter un turban	
	avoir une nouvelle coupe de cheveux	
	avoir les cheveux bouclés	7
[:-)	écouter un baladeur	7
}:-(	avoir son postiche à l'envers !	
	Bête à cornes	
	se prendre la tête	
	vient de s'envoler!	7
=:-)	être punk	
	être coiffé en brosse	
	avoir la tête en forme de tuyau	7
>:->	être démoniaque	7
B-)	porter de grosses lunettes	
	porter des lunettes	
	Allons à la plage, les amis !	
	Batman	7
C=:-)	être un chef	7
X-(	venir de mourir	7
-0,06	être mort	
	avoir le cerveau mort	
	avoir pété un plomb	
	être débile !!	6
%-)	avoir pété un plomb	
	loucher	
	avoir passé des heures et des heures devant l'écran.	6
(:I	être intello	
	être un crâne d'oeuf	6
.-)	être borgne	6
:-)~	baver	
	cafouiller	

en vertu de la loi du droit d'auteur.

	radoter	6	
:<	être déçu		
	être très triste	6	
:D	rire	6	
:O	bailler		
	crier	6	
:v)	avoir le nez cassé à gauche		
	avoir le nez qui pointe à gauche	6	
@=	bien que quelque peu déformé, être partisan de la guerre nucléaire		
	être pro-nucléaire		
	un champignon magique	6	
|-I	dormir		
	sommeil, ennui	6	
>;->	clin d'oeil démoniaque	6	
3:-o	être une vache	6	
8)	(être) grenouille	6	
8:]	(être) gorille	6	
B:-)	porter ses lunettes sur le front		
	porter des lunettes sur la tête	6	
*<:-)	porter le bonnet du Père Noël		
	être le Père Noël		
	porter un chapeau	5	
,-)	bien qu'invisible, je fais un clin d'oeil		
	être un borgne qui fait un clin d'oeil		
	ne dormir que d'un oeil !	5	
:-'	cracher		
	fumer	5	
-:-(	punk méchant		
	punk	5	
':-)	avoir accidentellement rasé un de ses sourcil ce matin	5	
-:-)	être un punk		
	(être) un indien	5	
:*	bisous	5	
:/7)	(être) Cyrano de Bergerac	5	
::-)	porter des lunettes	5	
:@	hurler		
	être interloqué	5	
:]	sourire timide		
	vouloir être un ami	5	
:-<	être très triste		
	faire la gueule		
	porter une moustache	5	
:=)	avoir deux nez	5	
:>	sourire jusqu'aux oreilles		

en vertu de la loi du droit d'auteur.

	sourire espiègle		
	sarcastique		
	Hmm, laissez-moi y réfléchir	5	
:-0	être orateur		
	"remarque choquante"		
	pas de hurlements! (Labo en paix)	5	
:8)	(être) un cochon	5	
:-J	blaguer		
	sourire à gauche	5	
:Q	crier avec un herpès		
	Utilisateur de vi [179] disant : Comment on sort de ce foutu éditeur EMACS ?		
	Euh... Tu baves, là, j'crois !		
	Quoi ?	5	
:-t	garder un visage impassible		
	rire jaune		
	être en colère	5	
:-W	être langue de vipère	5	
[:]	être un robot	5	
<:-I	être un fou		
	être un cancre	5	
0:-)	être innocent		
	être un ange	5	
O-)	(être) cyclope		
	être un plongeur	5	

Si certains items présentent des phénomènes de polysémie parallèles avec ceux qu'on rencontre dans le lexique (les unités seraient utilisées pour dénoter des réalités ayant un noyau sémantique commun, ou étant liées sur le plan pragmatique), une très grande partie reçoit des descriptions significations très éloignées les unes des autres voire sans lien (ex. "être ivre" et "porter la moustache"). Les émoticones seraient donc ambiguës étant donné, sans doute, les moyens limités de représentation iconique. Dans notre corpus, cependant, nous ne retrouvons pas toutes ses significations. Une étude même superficielle de ces unités montre la large part d'irréalité que présentent les pages qui revendiquent le statut de dictionnaires de *smileys*. Cette dernière s'explique par une surestimation et une valorisation du degré de cryptonymie rencontré dans les messages électroniques.

5.2.2. Emplois des frimousses

5.2.2.1. Des pages Web au corpus

[179] Lire [viaj] nom d'un éditeur de texte (Visual Interface).

Seule une vingtaine d'éléments des listes étudiées apparaissent dans notre corpus. Les significations présentées pour ces éléments ne sont qu'en intersection partielle avec leurs emplois dans notre corpus. Le tableau ci-dessous présente une croix pour chaque emploi recensé.

Tableau (5-16) – Comparaison des emplois du corpus et des descriptions des dictionnaires de frimousses

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Frimousse	Descriptions	emplois recensés
:-D	rire	X
	moquerie	de soi-même
	parler trop	-
	gaucher qui vous rase	-
:-p	bavard	-
	secret	-
	tire langue	X
	langue a fourché	-
	camarade	-
:-X	ne rien pouvoir dire / motus	-
	bisous	X
	nœud papillon	-
:-(	ne pas être content	X
	être triste	X
	faire la gueule	X
	être déçu	X
	expression de désapprobation	X
:-)	rire	X
	comédie	-
	être heureux	X
	« employé avec des phrases ironiques »	X
	sourire	X
	bonne humeur	X
;-)	clin d'œil	X
	« à prendre au second degré »	X
	s'être fait cogner à force de raconter des blagues	-
	s'être fait cogner	-
:)	être heureux	X
	sourire	X
	(être) lépreux	-
:*)	être ivre	-
	porter la moustache	-
	être euphorique	X
:(	ne pas être content	X compatissant
	être triste	X
	être déçu	X
:D	rire	X
:O	bailler	X
	crier	-
:=)	avoir deux nez	-
:>	sourire jusqu'aux oreilles	-
	sourire espiègle	X
	sarcastique	-

en vertu de la loi du droit d'auteur.

	Hmm, laissez-moi y réfléchir	-
:')	pleurer	X (pour rire)
	être triste	X
:-))	Double menton	-
	Je suis très content.	X
	L'interlocuteur a un double menton	-
	très heureux	X
:-)))	Je rigole énormément	X
	Triple menton (très gras)	-
:/)	*Ce n'est pas drôle*	-
	C'est pas marrant!	-
:o)	Nez rouge	-
;-(	Avoir envie de pleurer	-
	Oeil au beurre noir	-
=)	yeux exorbités	-

On notera que les significations particulières et qui s'écartent de la vocation première des frimousses ne sont pas représentées dans notre corpus à une exception près :

Extrait du Corpus F7 *PZZ: PBO arrete pas de bailler :O

Mais on remarquera immédiatement la redondance entre le verbal et la frimousse.

5.2.2.2. Inventaire des unités du corpus

Notre corpus présente 248 types de frimousses différentes au sens strict du terme (232 types sans distinction de casse) pour 2059 occurrences ce qui donne donc une moyenne de 8,86 occurrences pour chaque type. Mais cela va sans compter un examen plus détaillé des formes qui permet sans mal de repérer des classes.

Une dizaine de formes vedettes (fréquentes) peut être dégagée de l'ensemble, formes pour lesquelles il existe des variantes :

Tableau (5-17) – Types de frimousses

en vertu de la loi du droit d'auteur.

forme	Occurrences exactes	Occurrences exactes +variantes
:)	644	920
:(	64	81
:p	202	241
;)	182	255
:/	46	50
:o)	45	58
=)	38	40
;o)	14	19
;p	30	38
xox	10	60
xxx	20	117

Si on classe les unités en fonction de leur forme exacte, la figure (5-1) présente une forte proportion de formes rangées dans une catégorie "autres", non satisfaisante pour la raison qu'il est possible de repérer des séries de formes qui se rattachent à un fonctionnement commun.

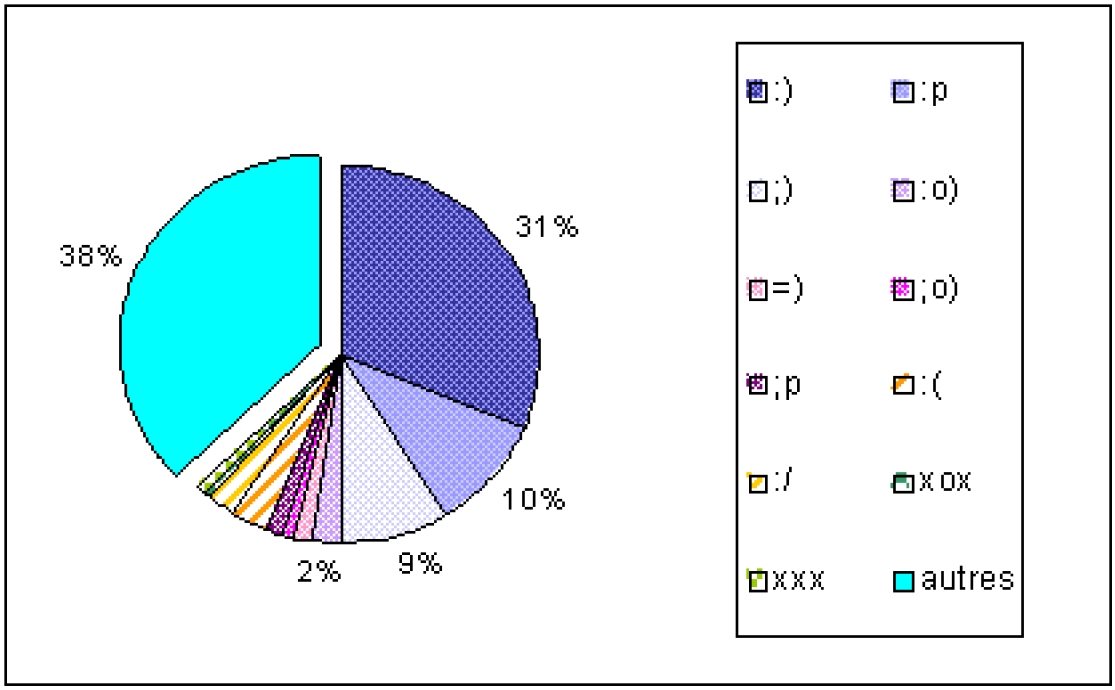

Figure (5-1) – Répartition des frimousses par type – occurrences exactes

La répartition des frimousses par type selon les occurrences exactes laisse une large part d'items non classifiés. Nous avons considéré comme variantes d'une même frimousse le cas où un redoublement de caractères apparaît et/ou où seule la casse des caractères change. En effet, dans les emplois présents dans le corpus, le redoublement de caractères et les différences de casse semblent exprimer un degré pour l'émotion représentée (ou sont utilisées pour mettre en évidence un message). Nous donnons

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

ci-après le tableau des variantes pour la frimousse vedette :

Tableau (5-18) – Variantes de la forme vedette

forme	occurrences
:)	644
:))	121
:)))	58
:))))	21
:))))	1
:::::	18
:))))))	10
:::::::	9
:))))))))	8
:::::::::	3
:))))))))))	1
:))))))))))	4
:))))))))))))	4
:))))))))))))))	2
:))))))))))))))	2
:))))))))))))))))	2
:))))))))))))))))	2
:))))))))))))))))))	1
:))))))))))))))))))))	1
:))))))))))))))))))))))	2
:))))))))))))))))))))))))	1
:))))))))))))))))))))))))))	1
:)))))))))))))))))))))))))

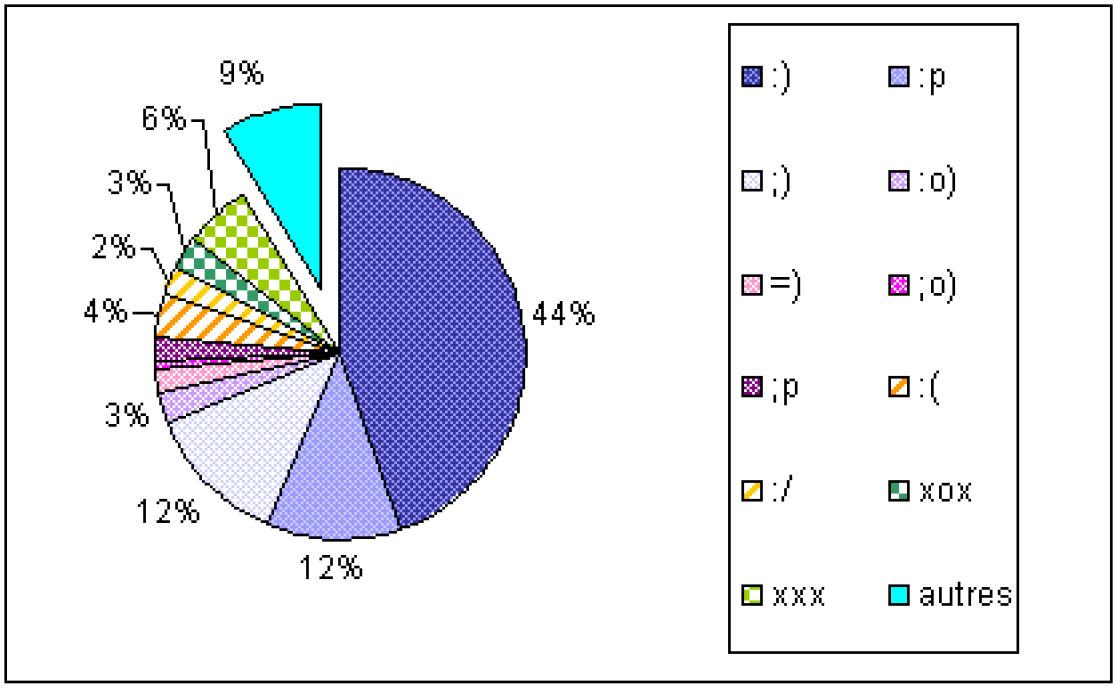

Figure (5-2) – Répartition des frimousses par type en regroupant les variantes

Après regroupement des variantes, la part de frimousses non prises en compte dans le classement est fortement diminuée. Un petit nombre d'items est fréquemment utilisé sans recours à l'explicitation. Il semble donc raisonnable de poser que l'utilisation de ces items est conventionnalisée ou tout au moins présente une certaine stabilité.

5.2.2.3. Caractéristiques générales

1908 interventions comportent un *smiley*. On notera que la moyenne pour le nombre d'occurrences par intervention est 0,17 ; l'utilisation des émoticones n'est donc pas systématique.

Ces unités apparaissent majoritairement en fin d'intervention (1857 interventions se terminent par une frimousse) – ce qui suggère que leur production intervient en second plan, après composition du message linguistique. Elles peuvent constituer une intervention à elles seules (320 interventions), en réaction à une intervention précédente d'un autre locuteur. Les quelques unités qui apparaissent en d'autres endroits (141 interventions sont concernées) ne permettent pas d'établir une confiance absolue dans leur positionnement. Il se trouve néanmoins dans notre corpus que la majorité des exemples de ce type se présente en début d'intervention ou en frontière de clause, ainsi que l'illustre l'extrait suivant.

Extrait du Corpus P9 *PBG: NP , ne te meprends po :) je ne fais que rester poli moua .

Lorsqu'on trouve un *smiley* en début d'intervention, il est en général une réaction à une intervention précédente d'un autre locuteur et est suivi d'un énoncé interventif initiatif. En ce qui concerne les autres positions possibles, les pages de frimousses issues du Web étudiées laissent penser que certains *smileys* pourraient être utilisés comme des mots, on

en vertu de la loi du droit d'auteur.

ne constate cependant dans notre corpus que de rares cas où une frimousse prend la place d'un mot, ainsi que l'illustrent les deux extraits ci-dessous.

Extrait du Corpus F6 *PCN: CD PAN et PCN on a pas de * mais ø)) *PAN: PCN pas la peine de me citer LOL *PCN: mais si Extrait du Corpus P7 *(action): NP fait des xxxxxxxx a sa deesse

Après avoir dessiné à grands traits le positionnement de ces unités dans les énoncés, il convient de chercher leur fonction.

5.2.2.4. Statut des frimousses dans le corpus

On peut distinguer dans notre corpus deux catégories de frimousses : positives et négatives. Ce qui est à relever, c'est que beaucoup de frimousses apparaissent dans des séquences d'ouverture ou de clôture. Ce sont la plupart du temps des formes positives. Les formes négatives apparaissent en séquence de clôture pour manifester la déception liée au départ de quelqu'un. Les frimousses qui se présentent dans le corps des conversations sont aussi essentiellement de forme positive. Les frimousses de forme négative apparaissent pour une très grande part en combinaison avec des formes positives et sont souvent elles-mêmes à ne pas prendre au premier degré. Les interactions conflictuelles ne présentent pas de frimousses négatives, le mécontentement apparaît alors dans les mots.

La présentation de l'extrait ci-dessous permet d'illustrer ce phénomène et permettra, nous l'espérons, au lecteur de se faire une idée de la façon dont se jouent les échanges sur les IRC. Les lignes de messages sont présentées sous leur forme anonymée. Un astérisque précède le code du locuteur émetteur du message et ce code est suivi de deux points. Vient ensuite le message à proprement parler, qui peut se présenter sur plusieurs lignes. La ligne précédée de "%add:" qui suit chaque message présente le code du locuteur auquel s'adresse le message, lorsqu'il a été identifié. Il s'agit là d'une information que nous avons apportée dans nos corpus de façon à délimiter des unités conversationnelles [180], elle n'apparaît pas dans le corpus brut. Nous avons ensuite, le cas échéant indiqué entre parenthèses le nombre de messages intercalés n'appartenant pas à l'unité thématique présentée.

Extrait du Corpus F8 [181] *PDY: PDV t'as rien d'autre a faire que des copier/coller des privates que tu as ? %add: PDV (4) *PDV: PDY il s'agissait pas d'un private, et merde !!! et merde et merde %add: PDY (2) *PDV: quand on sait pas on la ferme %add: PDY (7) *PDY: ben j'en sais rien PDV, vu que tu fais ca tout le temps %add: PDV (10) *PDV: PDY tu me connais d'où pour dire que je fais ça tout le temps ! tu dois confondre %add: PDY (12) *PDY: PDV ben je te connais ici, et je sais que c'est un truc que tu fais souvent %add: PDV (2) *PDY: et c'est pas genial , sachant qu'un private ne porte pas son nom pour rien %add: PDV (2) *PDV: PDY PDV ben je te connais ici, et je sais que c'est un truc que tu fais souvent <-- je ne sais pas si tu cherches une dispute ou une conversation, mais je ne fais pas ça ! et je

[180] Ces unités sont délimitées par topics conversationnels – Cf. Chapitre 6.

[181] ***Les nombres qui apparaissent entre parenthèses sur une ligne correspondent au nombre d'interventions appartenant à d'autres unités conversationnelles.***

en vertu de la loi du droit d'auteur.

t'avoues un truc, tu m'enerves %add: PDY *PAV: mca veut dire quoi private..... %add: PDY, PDV, (1) *PDY: en tout cas, t'aimes bien les copier/coller:) %add: PDV (3) *PDY: et puis non, franchement je m'en fous que tu fasses ca, je cherche pas du tout une dispute %add: PDV (4) *PDV: PDY peut etre mais pas des privés! tu t'amenes et tu fais que dire des trucs que tu sors de je ne sais où %add: PDY *PDY: je dis juste que c'est pas genial pour ceux qui t'ecrivent en private, c tout %add: PDV (6) *PAK: finalement, je crois que le "(6) polemique, insulte, jugement sans foncdement" remporte tois les suffrages.... %add: PDY, PDV, ALL

Nous ne présentons ici que le début de cette séquence qui compte 223 interventions et se trouve parmi les plus conflictuelles de notre corpus. Il n'y figure qu'un *smiley*, de forme positive.

5.2.2.5. Statuts particuliers d'items

Dans notre Tableau (5-17) *supra*, il faut noter le statut particulier de deux items qualifiés comme frimousses. Il s'agit des éléments *xxx* et *xox*. Ces derniers se trouvent dans notre corpus essentiellement dans les séquences d'ouverture et de clôture. Ils copient les rituels de salutation des conversations en face à face et représentent la bise –d'une façon acoustique.

Tableau (5-19) – Liste des *smileys* communs à cinq pages au moins

item	occurrences exactes	occurrences variantes	occurrences dans les séquences d'ouverture et de clôture
xox	10	60	58
xxx	20	117	101

Les exemples qui suivent illustrent les emplois de ces unités.

Extrait du Corpus F1 *PBQ: kikoo PBW!!! :))) xxxxx %add: PBW Extrait du Corpus F3 *PCM: bye bye PBG xxxxxxxxxxx %add: PBG

Ces dernières sont parfois utilisées en remplacement de mots.

Extrait du Corpus F1 *PBD: xxxxx de P99 il est la

5.2.3. De l'usage des frimousses

L'inventaire très important de formes rencontrées, tant sur les pages Web, que dans nos corpus, laissait entrevoir l'existence possible d'un système de signes complexe. Toutefois, nous avons été amenées à considérer que les formes les plus fréquemment utilisées, représentant 91% des occurrences, peuvent être réparties en deux classes : formes positives et formes négatives – si on écarte le statut particulier des items *xxx* et *xox*.

On note que les formes positives sont en plus grand nombre que les formes négatives (7% du total). Les unités positives dans nos corpus ont essentiellement pour fonction de manifester de la bonne humeur, de la gaieté, et ce surtout dans les séquences d'ouverture. Dans le cours des conversations, elles signalent aussi parfois que le

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

message auquel elles sont associées n'est pas sérieux – elles indiquent alors l'inadéquation du contenu propositionnel à la réalité comme fait revendiqué par l'émetteur du message.

Les formes négatives, moins fréquentes, manifestent de la déception ou tristesse dans les séquences de clôture. Les unités négatives, en dehors des séquences de clôture, quand bien même elles participent de la régulation de l'activité, sont très souvent antiphrastiques.

Les frimousses sont beaucoup utilisées dans les séquences d'ouverture et de clôture, ou, tels les rires, comme régulateurs de l'activité communicative. Moins fréquemment et par imitation du paraverbal, l'utilisation des frimousses dans le corps des conversations est à rapprocher de la notion de modalité. Elle permet de communiquer l'attitude du sujet parlant à l'égard de ce qu'il dit.

5.3. Conclusion

Au terme de ce chapitre, nous avons passé en revue un ensemble important de phénomènes qui manifestent un écart à la norme de l'écrit. La représentation de la chaîne parlée libérée du carcan de la norme allie des procédés qui permettent une réduction de la forme, et des procédés qui tentent d'épaissir les dimensions du matériau.

Les Français sont réputés très attachés au respect de la norme prescriptive, et ce, tout particulièrement pour l'écrit. L'examen de nos corpus indique que les internautes francophones dérogent largement à cette règle. Nos corpus présentent un écrit non figé et non stable en proie à des variations individuelles importantes. Ce que montre cette analyse du code, c'est entre autres que les utilisateurs semblent ne pouvoir ou ne vouloir se contenter de représenter le verbal dépouillé de ses corrélats paraverbaux, au risque parfois de ralentir l'encodage et le décodage et d'appeler à la mobilisation de capacités cognitives habituellement non sollicitées par l'écriture ou la lecture. On ne peut assurément pas invoquer la contrainte temporelle pour expliquer l'ensemble des phénomènes observés. Un certain nombre de procédés génèrent incontestablement une réduction formelle, mais ce n'est pas le cas de tous. De plus, il est raisonnable de se demander dans quelles mesures la réduction formelle permet un gain de temps en production. En effet, soit on suppose que les formes réduites font partie de la représentation des unités dans le lexique mental, et dans ce cas, leur génération est automatique, soit on suppose qu'un algorithme de réduction est à l'œuvre et augmente alors le temps de traitement. Dans un cas comme dans l'autre, il peut y avoir un bénéfice si l'augmentation du temps de traitement de l'unité préalable à la production est compensée par un temps de production de la forme réduite inférieur au temps de production de l'unité complète.

Quoi qu'il en soit, le fait que l'écart à la norme ne favorise pas toujours une réduction formelle implique que d'autres motivations sont en jeu. À l'oral, les phénomènes connus de réduction formelle ne sont pas seulement dus à la contrainte temporelle et au principe

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

d'économie de la langue, ils servent aussi de signes d'appartenance sociale. Pour l'écrit normé, les réductions formelles ont une fonction essentielle d'économie. Dans l'écrit des IRC les réductions formelles semblent pour la plupart assortie d'une fonction sociale.

L'analyse nous amène ainsi à caractériser l'écrit des IRC par les quatre points suivants. Le code des IRC n'est pas seulement une simplification du système de l'écrit, il en est un travestissement qui permet la représentation de caractéristiques de l'oral que les dimensions de l'écrit ne représentent pas. Nous avons observé en effet que des phénomènes qui n'impliquent pas de réduction formelle, les petits mots et les frimousses pallient l'unidimensionnalité du canal/medium [182] accentuée par le dispositif très contraint des IRC, puisqu'ils contribuent en partie à faire savoir aux allocutaires la façon dont ils appréhendent les informations, ou la façon dont ils la réalisent. Il faut aussi prendre en compte le fait que le canal/medium est presque le seul moyen de se manifester à l'autre [183] , il est ainsi raisonnable de penser que le code est utilisé et travesti par les participants pour se créer une identité (pallier la monocanalité). De plus, l'activité centrale est l'échange de messages, il est donc probable que l'intérêt de la chose soit volontairement augmenté par les participants en exploitant toutes les possibilités offertes par le dispositif, quand bien même cela présente un écart à la norme. Enfin, outre sa fonction de représentation, le code présente une dimension sociale, il est utilisé comme signe distinctif d'appartenance à une communauté.

[182] Cette dénomination n'est pas la manifestation d'une distinction théorique, elle nous permet seulement de distinguer deux homonymes de notre texte : le *canal*, élément d'un dispositif IRC ne se confondra pas avec le *canal* en tant que support physique nécessaire à la manifestation du code. Nous opposons ainsi les IRC à la situation en face à face qui, elle, permet une communication multi-canale (visuel, auditif, proxémique…) et peut utiliser les dimensions de l'oral.

[183] Les commandes présentées aux chapitres 2 et 3 sont un autre moyen de se manifester à l'autre. Il est aussi possible d'utiliser d'autres dispositifs de communication humaine en même temps que les IRC.

en vertu de la loi du droit d'auteur.

en vertu de la loi du droit d'auteur.

Chapitre 6 Unités conversationnelles

Nous avons présenté, au chapitre 2, les caractéristiques du dispositif d'où sont issus nos corpus. Ces corpus présentent des discours produits dans des situations naturelles. Certes le medium est nouveau et utilisé par une faible proportion de locuteurs, il ne s'agit pourtant pas de situations artificielles telles qu'on les rencontrerait dans la mise en situation de locuteurs pour l'observation de leurs usages linguistiques. L'objet technique qui permet la communication linguistique conditionne la structure spatio-temporelle de l'activité communicative.

Nous avons vu, au chapitre 2, que les communications IRC se réalisent dans des canaux. Ces derniers ont des propriétés particulières (nom, sujet, statut public ou privé, modéré ou non modéré, nombre de participants) et il leur est associé un code de conduite particulier (netiquette). Les participants ont également des statuts dans le canal (opérateurs, *voices*, simples utilisateurs) qui leur donnent des autorisations particulières quant à l'envoi de messages et à l'utilisation de commandes. Enfin, l'activité d'un canal est constituée des messages (simples et actions) et des commandes des participants (entrées et sorties du canal, changements de surnom des participants, commandes des opérateurs sur les participants ou sur les propriétés du canal) [184] .L'espace d'activité d'un canal est un centre où s'enregistrent séquentiellement les interventions publiques (messages et commandes) des utilisateurs connectés. Rappelons que tout utilisateur connecté a accès à l'ensemble de l'activité publique du canal depuis son arrivée dans le canal [185] .

[184] Cf. chapitre 3.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Nous essaierons ci-après de caractériser les spécificités de l'interaction induites par le dispositif.

En effet, une lecture même superficielle – si toutefois elle est possible – des différentes sessions d'enregistrement montre qu'il convient de porter une attention particulière à la façon dont les messages s'échangent entre les participants d'un canal, et implique un traitement particulier des corpus.

Depuis les années soixante-dix les interactions communicatives sont l'objet d'un champ de recherche transdisciplinaire, il existe donc différents courants qui sont en partie dessinés par les objectifs d'analyse [186] . Dans la diversité des interactions communicatives la linguistique, qui a tardivement manifesté l'intégration véritable de l'interaction à son champ d'étude, s'intéresse essentiellement aux interactions verbales. Kerbrat-Orecchioni (1996:7-8) dégage les paramètres nécessaires à l'établissement d'une typologie de ces dernières. Ce sont : le cadre spatio-temporel dans lequel s'inscrit l'interaction, la nature du canal, le nombre et la nature de ses participants, son but, son degré de formalité et son style. Il nous faut ici ajouter, que les propriétés du dispositif technique, sont également importantes à prendre en compte.

Dans un premier temps, nous essaierons de proposer une caractérisation du type d'interaction en présence duquel nous sommes, à la lumière des paramètres mentionnés ci-dessus. Puis nous entrerons plus profondément dans l'analyse en observant ce qui se joue pour le réglage de l'alternance des tours de parole. Notre intérêt se portera ensuite sur les différentes unités qu'il est possible de dégager dans nos corpus en référence au modèle d'analyse proposé par l'école de Genève. Enfin, nous observerons quelques caractéristiques particulières qu'entretiennent les différentes interactions au sein d'une même session d'enregistrement.

6.1. Caractérisation du type de comportement verbal, ou type d'interaction.

Sont donc identifiés différents types d'interactions verbales [187] en fonction des paramètres mentionnés en introduction. Nos corpus semblent pouvoir être caractérisés comme un ensemble de conversations qui ont la particularité d'être écrites.

Selon André-Larochebouvy (1984) "la conversation est une interaction verbale réciproque ; elle comporte des phrases et des formules fortement ritualisées dont l'analyse est essentielle à la connaissance de la structure de la conversation ; elle requiert

[185] L'activité du canal s'affiche séquentiellement dans une fenêtre, que nous avons appelé fenêtre d'activité, dont on peut faire défiler le contenu pendant la session (intervalle durant lequel l'utilisateur est connecté).

[186] Pour une présentation de ces différentes approches, on se reportera à Kerbrat-Orecchioni (1996:10-15).

[187] Pour une typologie, on se reportera à Kerbrat-Orecchioni (1991, 1996).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

la présence d'au moins deux participants pourvus de droits égaux au tour de parole ; c'est une activité spontanée et gratuite ; c'est une unité dans le temps et dans l'espace."

Cette définition de la conversation met en avant plusieurs des propriétés d'un acte communicatif verbal qui coïncident avec les propriétés de nos corpus. Nous sommes effectivement en présence d'interactions verbales réciproques synchrones. L'activité que nous observons est spontanée, elle n'est nullement imposée au sujet parlant.

La notion de "gratuité" qui apparaît dans la citation peut s'entendre de plusieurs façons. La "gratuité" appelle là, évidemment, la question des buts communicatifs et requiert qu'on s'interroge sur sa pertinence pour caractériser le phénomène. Cette question peut s'envisager de façon négative. Dans une activité qui ne serait pas gratuite, on suppose que les participants tirent ou perdent bénéfice dans cette activité. Cette notion concerne évidemment les intérêts généraux des participants. Ces derniers peuvent avoir des conséquences en particulier sur la négociation de la thématique, sur la longueur des messages linguistiques et sur leur structure. Il reste alors à qualifier le bénéfice reçu ou perdu par l'effet de la communication. Se peut-il que des individus communiquent s'ils n'espèrent y trouver aucun bénéfice ? On s'aperçoit que la notion de "gratuité" ne peut être clairement et précisément abordée qu'en excluant les bénéfices internes à la communication : les participants n'ont d'autre but que celui de communiquer et d'en tirer les bénéfices inhérents. On peut relever que les corpus recueillis présentent une grande liberté de la part des locuteurs, à la fois dans les thématiques abordées et le type de discours produit. Les échanges ne semblent pas contraints et les locuteurs ne semblent avoir, en règle générale, d'autre but que celui de communiquer. Certains cherchent toutefois à obtenir des réponses à des questions précises, parfois techniques ; d'autres encore cherchent à établir une relation privilégiée avec une ou plusieurs autres personnes (ils produisent des messages de type petites-annonces) [188] ; mais ce n'est pas le cas général.

Des phénomènes ont été le plus évidemment repérés pour l'autre facette de la notion de "gratuité" [189] , c'est-à-dire lorsque les intérêts financiers des participants sont en jeu. On peut concevoir que les participants soient rémunérés pour une activité impliquant l'interaction verbale, ou qu'ils paient et dans ce cas, on peut difficilement imaginer qu'ils n'en tirent pas un bénéfice quelconque. Ainsi, on a pu remarquer des différences notables dans les écrits transmis avec port payé (qui a remplacé progressivement le port-dû), le prix étant conditionné non seulement par la distance, le temps de transport, mais aussi par le poids voire le nombre de mots [190] . Les télégrammes ainsi sont connus pour cet aspect, également les petites annonces présentent des phénomènes liés à des aspects financiers [191] . La gratuité financière de l'activité de communication IRC n'est évidemment

188 Nous serons amenée par la suite à revenir sur cette question.

189 Facette qui n'est pas explicitement contenue dans la citation d'André-Larochebouvy.

190 En 1909, le tarif du port d'une carte postale est moindre si cette dernière ne comporte pas plus de cinq mots. En revanche, le pneumatique ne connaît pas de tarif conditionné par le nombre de mots ; seule la dimension des dépêches impose une limite.

191 On remarquera toutefois, que les petites annonces gratuites et sans limitation de taille présentent les mêmes phénomènes.

en vertu de la loi du droit d'auteur.

pas véritable pour tous les locuteurs. Cette activité nécessite en effet un équipement informatique et un abonnement auprès d'un fournisseur d'accès Internet [192] . Toutefois il semble que le coût supporté par les participants n'est pas réellement mesuré [193] et n'a pas plus d'effet que dans les conversations téléphoniques. Il n'en est fait que très rarement mention au cours de l'interaction, toujours avec humour et jamais en termes monétaires. Cela ne semble pas représenter une préoccupation majeure des participants ni une entrave à leur activité communicative.

Extrait du Corpus P3 *PAJ: Mon provider me fait un crédit d'octets PAB ;/ %add: PAB

Les raisons explicitées du départ d'un participant sont bien autres que des questions pécuniaires, ainsi que le montrent les quelques exemples ci-après, qui pourraient à quelques détails près, être issus de conversations téléphoniques [194] .

Extrait du Corpus F2 *PAD: euh brb je vais faire ma lessive. à la main s'il vous plait! Extrait du Corpus F3 *PAY: 'vais prendre une douche en écoutant du metal a fond, ca devrait me detendre... Extrait du Corpus F4 *PCD: l'ecole a ttend pas Extrait du Corpus F4 *PAB: bon... temps d'aller faire les courses moi..... xxxxxxx tous Extrait du Corpus F4 *PAY: bon break faut faire les courses a+ ! Extrait du Corpus F8 *PBO: bon encore 15 min et retour au boulot moi...

Le plus souvent, les motifs évoqués pour justifier un départ, sont des activités incompatibles avec l'activité de CMO.

Le contenu de la citation d'André-Larochebouvy doit être nuancé, si on souhaite qu'il convienne pour les IRC. La question du droit égal au tour de parole doit en effet être repensée en fonction du dispositif même de communication, puisque certains locuteurs ont le pouvoir d'exclure d'autres locuteurs de la communication et que tout locuteur peut ignorer [195] tout autre locuteur. Néanmoins ces restrictions n'affectent que de façon minime les échanges observés et l'exercice de la permutation des rôles d'émetteur et de récepteur est la règle à quelques rares exceptions près ainsi que le montre l'extrait ci-dessous, dans lequel un utilisateur privilégié suspend le droit de parole d'un autre locuteur.

Extrait du Corpus P5 [196] P55: !b P58 *** XXXX has joined #PCanal P55: zut rate :) P50: loll P55: !b P58 *** RobotLogiciel sets mode: +b *!*?------@*.XXXX.XXX.XX P55: laaa *** XXXX has quit IRC (Quit: ¤14-(-12VirusScript7 ME14-)(-15BeTa14-)-

[192] Les coûts ne sont pas forcément supportés directement par les participants, les situations étant très diverses (à domicile chez les parents, au lycée ou à l'université dans une salle de travail, dans un lieu public…).

[193] L'ordinateur est considéré comme un investissement global.

[194] Sauf utilisation de téléphones mobiles.

[195] Notons que le canal/medium facilite l'action d'ignorer un autre locuteur. Il est assez difficile de s'empêcher d'entendre un locuteur particulier alors qu'il est aisé de ne pas lire les interventions d'un locuteur.

[196] ***L'opérateur est, dans cet extrait, assisté d'un robot logiciel. Ce dernier repère le message de l'opérateur (!b P58) et le transforme en une commande.***

en vertu de la loi du droit d'auteur.

10Get it at6 http://www.XXXX.XXXX.XXX/ 10and6 http://www.XXXX.XXX/¤) *P55: now l'emmerdeur il se la ferme *P55: :) *P55: mais je ne t'empeche pas d'etre ici :)

Le locuteur ne peut ici plus produire de message. Des situations équivalentes peuvent se produire en face à face, mais ici, le dispositif fait que l'emploi de la "force" n'atteint pas la personne physiquement.

Les locuteurs exclus sont très souvent des participants qui n'ont pas respecté la netiquette et qui n'étaient pas engagés dans une interaction. Une seule fois sur tout le corpus un locuteur ayant le statut d'opérateur a restreint le droit au tour de parole aux seuls locuteurs ayant le statut d'opérateur, la situation n'ayant duré que quelques secondes et n'ayant entraîné aucune rupture d'interaction.

Extrait du Corpus F7 *** PJQ sets mode: +m *PJQ: mode sieste lol %add: ALL *PBP: hehehe %add: PJQ *PJQ: sont fatiguants avec leurs inepties et leurs discours vains %add: PBP *PBP: c nul PJQ on voit pas les conmneries des autres %add: PJQ *PBP: lol %add: PJQ *** PJQ sets mode: -m *PJQ: tiens pour toi :) %add: PBP *PJQ: lis et prend ton pied :) %add: PBP

La censure des opérateurs intervient surtout pour faire respecter trois points de la netiquette du canal : utiliser les lettres capitales à bon escient, parler français et ne pas faire de publicité. Dans le cas de la langue utilisée et des lettres majuscules, un avertissement est formulé au participant fautif avant toute mesure répressive.

Extrait du Corpus F1 *PBW: PCM.... en francais ici !!!!!!!! %add: PCM Extrait du Corpus F3 *PCM: PBV pas de majuscules merci et parle francais %add: PBV

En revanche les utilisateurs qui n'interviennent que pour faire de la publicité pour d'autres canaux sont éjectés du canal sans avertissement, que leur activité intervienne sur le canal lui-même ou auprès de l'ensemble des participants d'un canal par le biais de messages privés.

Le recours à la force qui entame la réciprocité n'est ainsi qu'une entrave mineure à l'activité communicative des canaux.

La réciprocité cède aussi à la possibilité qu'ont les participants d'ignorer techniquement les interventions d'un autre participant. Bien sûr il existe la possibilité ordinaire, connue en conversation en face à face qui consiste à ne pas prêter attention à un locuteur dont on est l'allocutaire. Dans le cas des IRC, cette non réponse n'est un affront que dans la mesure où l'allocutaire est identifiable. Un grand nombre d'interventions ne reçoivent pas de réponse. Cela concerne 9,16 % des interventions sur l'ensemble du corpus. Les interventions ignorées sont des interventions qui n'ont pas d'allocutaire précis. Cette catégorie peut être divisée en 2 classes : les interventions dans d'autres langues [197] et les interventions pour lesquelles aucun des participants ne s'est reconnu comme allocutaire. Les premières représentent 3,25 % des interventions de l'ensemble du corpus et 40,52 % des interventions ignorées. Les secondes sont pour la majorité des interventions adressées à l'assemblée ou à un locuteur absent. Cette seconde catégorie représente 5,91 % de l'ensemble. Ce tableau est à nuancer puisque rien ne dit que les locuteurs qui n'ont pas obtenu de réponse dans la fenêtre d'activité du

[197] Et plus particulièrement celles qui n'ont pas d'allocutaire précis (ce qui est le cas de la majorité des interventions dans d'autres langues que le français).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

canal n'en ont pas reçu en privé. En effet, 17,35 % des interventions ignorées sont des interventions du type petite annonce qui appellent souvent des réponses en privé. La responsabilité de l'acte d'ignorance ne peut donc dans ce cas incomber aux participants d'un canal. On compte encore 6% des interventions ignorées qui n'appelaient pas de réponse particulières, étant des descriptions d'activité ne s'intégrant à aucune des interactions en cours. Enfin, 1,96 % sont des FTA [198] pour la face positive de celui qui les subit, ce dernier n'étant pas clairement et univoquement identifié.

Extrait du Corpus P6 *PE8: les francai c'ai que des pd

La liberté d'ignorer un locuteur n'est spécifique à l'IRC dans sa version technique. Un utilisateur peut exécuter une requête d'ignorance sur un utilisateur particulier. Le résultat de cette commande est que les interventions du locuteur ignoré ne sont absolument pas présentées au locuteur ignorant. On peut voir ci-dessous une illustration de ce phénomène.

Fenêtre du participant X	Fenêtre du participant Y	Fenêtre du participant Z
C as joined Xcanal Z: t'aV k pa parler X: Z tu ménerv /ignore Z Y: eh oh :(Y: Z aret X: ca boum Y??? Y: X ouep	C as joined Xcanal Z: t'aV k pa parler X: Z tu ménerv /ignore Z Y: eh oh :(Z: X te pff Y: Z aret X: ca boum Y??? Y: X ouep	C as joined Xcanal Z: t'aV k pa parler X: Z tu ménerv /ignore Z Y: eh oh :(Z: X te pff Y: Z aret X: ca boum Y??? Y: X ouep

Ce participant ignorant ne possède même pas l'information de l'événement d'intervention du participant ignoré sauf à le déduire des informations présentes dans les interventions des participants non ignorés. L'extrait de corpus ci-dessous présente une intervention faisant référence à ce type de situations. PDY est ignoré d'un participant, mais son attitude montre qu'il ne sait pas de quoi il s'agit – ce locuteur continue de s'adresser en effet au locuteur qui l'ignore. Un autre locuteur explique donc à PDY de quoi il retourne :

Extrait du Corpus F8 *PDJ: PDY qd tu ignores, c'est mm sur le chan, elle ne voit plus tes messages %add: PDY

La citation d'André-Larochebouvy omet ou confond certains des critères typologiques en matière d'interaction verbale. Le but de l'interaction, son degré de formalité et son style sont peu pris en compte [199] . Kerbrat-Orecchioni (1996:8) formule la question de la gratuité d'une façon plus explicite. Elle indique en effet que la conversation est un type d'interaction dont les participants "n'ont pas d'autre but avoué que le seul plaisir de converser, [qu']elle a enfin un caractère familier et improvisé : thèmes abordés, durée de l'échange, ordre des prises de tour". Ces caractéristiques correspondent pleinement à ce que nous trouvons dans nos corpus. Le but global des interactions ne semble pas autre que celui d'entretenir un lien et de participer à une communauté.

[198] Face Threatening Acts selon la terminologie de Brown et Levinson (1987) reprise et aménagée par Catherine Kerbrat-Orecchioni (1992 et 1995) entre autres.

[199] Le but de l'interaction est n'est évoqué que dans son caractère "gratuit", et le style dans son caractère "spontané".

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

6.1.1. Contexte

L'analyse des conversations pose comme caractères du contexte, le site, les participants et le but de l'interaction. Kerbrat-Orecchioni (1996) distingue deux rôles principaux du contexte. Du côté de la production, le contexte serait décisif pour les thèmes abordés, les formes d'adresse, le registre et les actes de langage. Et du côté de l'interprétation pour le décodage de l'implicite. Nous donnons ci-après les propriétés générales de nos corpus concernant le contexte.

6.1.1.1. Le site

Pour nos corpus, le cadre spatio-temporel est particulier puisque le cadre spatial partagé est virtuel au sens que donne à ce terme la cybernétique, et que le cadre temporel est repéré par rapport au temps GMT, puisque les participants à l'interaction peuvent être engagés dans des fuseaux horaires différents. Les exemples qui suivent illustrent cette propriété pour les dimensions spatiales puis temporelles.

Extrait du Corpus F6 *PCR: TU ES OU? Extrait du Corpus P1 *PAD: kelle heure il est chez toi?

Les locuteurs-scripteurs ne partagent donc qu'une partie du site, celui-ci peut-être considéré comme dédoublé, ainsi que nous le verrons au chapitre 9.

Cette particularité n'est pas exclusive pour la communication IRC, des remarques similaires peuvent être faites au sujet de la conversation téléphonique. Si on pouvait dire à son sujet qu'un des locuteurs (le locuteur appelant) au moins connaissait la position géographique globale de l'allocutaire, les techniques actuelles complexifient les situations. La présentation du numéro peut permettre que chacun des locuteurs identifie la position de l'autre, mais le transfert d'appel, et les téléphones mobiles font que souvent aucun des locuteurs ne connaît a priori la position de l'autre. Dans la communication IRC, à moins d'une mention explicite, les interlocuteurs ne connaissent pas les propriétés spatio-temporelles de leurs situations de communication respectives.

Du côté du temporel, ainsi, la littérature montre que les rituels d'ouverture et de clôture doivent être adaptés. Ainsi dans nos corpus se manifestent quelques spécificités à ce sujet :

Extrait du Corpus F7 [200] *PAK: ca va ce matin? euhhhh et après-midi? lol %add: ALL

Le locuteur PAK n'est ainsi certainement pas engagé dans le fuseau horaire de France métropolitaine [201] . Son message est produit en une seule fois. Dans ce cas, au moment de produire *euh* – dernière limite possible à laquelle le locuteur s'est rendu compte de son erreur, deux stratégies sont possibles : soit il corrige son intervention avant de l'envoyer, soit il poursuit en ayant pleine conscience que son erreur sera interprétée. La correction

[200] ***Cet enregistrement a eu lieu un après-midi en France.***

[201] Ou il feint de ne pas l'être.

en vertu de la loi du droit d'auteur.

du message étant moins coûteuse (supprimer le point d'interrogation et poursuivre, ou supprimer sept caractères et reformuler), on peut supposer que le temps d'hésitation marqué ici par l'interjection *euhhhh* est délibérément représenté, laissant apparaître aux allocutaires la situation particulière du locuteur ; la conjonction de coordination utilisée pour réparer en complétant le message permet au locuteur de s'adresser à la fois à ceux qui se trouvent partager la même condition temporelle et à ceux pour qui l'après-midi est commencée.

Cette situation se retrouve sporadiquement au long de nos corpus. Certains locuteurs ne laissent pas passer les rituels d'ouverture inadaptés. Il y a même de temps en temps débat à ce sujet.

Extrait du Corpus F6 *PAW: 'soir PBT %add: PBT (4) *PBT: salut PAW (c l'apres midi ici) %add: PAW (2) *PAW: (je sais, mais si je devais tenir compte du décalage horaire chq fois que je dis bonjour....) %add: PBT (2) *PBT: (c'est important quand meme) %add: PAW (6) *PAW: (important pour qui?) %add: PBT (6) *PBT: (PAW par exemple.... pour la nouvelle année si tu souhaite bonnanné a qqun d'ici en tenant compte de ton heure ca gache tout le punch de la nouvelle année vu que c pas la nouvelle année encore ici) %add: PAW (3) *PAW: PBT c'est pas totalement faux... mais ça n'arrive qu'une fois l'an %add: PBT

L'adaptation au moment se manifeste essentiellement dans les séquences d'ouverture et de clôture ainsi que dans les thématiques du corps des conversations – elle concerne alors les propriétés temporelles culturelles qui font le rythme de la vie en société occidentale (moment des repas, moment de travail, jours de repos, saisons, périodes de vacances, fêtes religieuses, fêtes locales, anniversaires…).

Le discours doit être approprié au moment et au lieu. Ce dernier est ici virtuel mais il a une manifestation physique concrète, la fenêtre d'activité d'un canal en étant la représentation. Les participants ont conscience à la fois d'être, virtuellement, dans un lieu public et actuellement dans un lieu privé ou public distinct. Les messages échangés se caractérisent par une assez large liberté d'expression. Cependant, un certain nombre de règles explicitées dans la netiquette [202] doivent être respectées. Il s'agit d'une sorte de contrat communicatif associé à un canal et au dispositif et par là au site.

Ainsi, des recommandations sont faites à différents niveaux. On remarque que la plupart sont négatives : ce qui est spécifié est davantage ce qu'il ne faut pas faire, la sanction étant l'interdiction pour un participant, à plus ou moins long terme, de rejoindre le canal sur lequel a été commise la faute.

Au niveau matériel, l'utilisation des majuscules et de la ponctuation ne doivent pas différer des usages habituels, en particulier l'utilisation des capitales doit être réservée à des cas précis et exceptionnels. Cette contrainte permet de conserver la possibilité d'associer une fonction à l'utilisation de ces éléments (insistance…) et de préserver la

[202] Rappelons que la netiquette (network etiquette) rassemble les consignes de bonne conduite sur Internet. Des adaptations locales existent, mais elles reprennent la plupart des points de la RFC1855. Les RFC (Request for Comments) sont des articles qui rassemblent les informations techniques (protocoles et normes) relatives à Internet. Ces articles sont identifiés par un numéro à quatre chiffres et sont disponibles sur de nombreux serveurs. On trouvera en annexe une reproduction de ce document ainsi que quelques exemples de netiquettes spécifiques aux IRC

en vertu de la loi du droit d'auteur.

lisibilité.

Il est recommandé également de ne pas produire des messages trop longs, d'une part pour ne pas monopoliser l'espace de conversation, et d'autre part parce que la visualisation de l'espace d'activité n'est pas le même pour tous (taille et définition des écrans) et peut ainsi amener l'écran d'un utilisateur à défiler d'un coup avant qu'il ait pu prendre connaissance de toutes les interventions. La répétition ainsi que le découpage intempestif d'un message (*flood*) sont considérés comme de mauvaises attitudes. Ces contraintes permettent elles aussi de préserver la lisibilité, elles sont également en rapport avec la gestion des tours de parole.

En ce qui concerne le code, nous l'avons dit les canaux étudiés souhaitent imposer le français. Les interventions dans d'autres langues s'exposent donc au risque d'expulsion. Produire des messages dans un code adapté est donc ainsi un paramètre d'adaptation au site.

Il est conseillé encore de toujours utiliser les formules de salutation et d'au revoir. Ainsi, il ne serait pas nécessaire de saluer chacun personnellement dans un canal. Et notamment utiliser une procédure automatique saluant chaque participant à son entrée dans le canal, est réprouvée. En ce qui concerne les adieux, la recommandation est la même et est assortie cependant de la nécessité d'attendre qu'une réponse y soit apportée par les participants avec lesquelles l'interaction se termine, il arrive en effet que la communication soit interrompue par des problèmes techniques, il convient alors de s'assurer que le message a bien été reçu.

Il est également précisé qu'il ne faut pas imaginer que des participants inconnus vont accepter de converser, et de surcroît en privé, avec n'importe qui. Les internautes ne devraient donc pas s'en offusquer.

Les informations personnelles ne doivent pas être demandées avant d'avoir tissé un lien avec un participant, et il est demandé de respecter l'anonymat des participants dans tous les cas. On recommande également d'être poli, amical et serviable.

Les participants ne devraient pas quémander un statut particulier. Ils ne devraient pas non plus utiliser de scripts sans les comprendre, ni utiliser des robots logiciels sans les déclarer.

L'ouverture de canaux privés ne devrait se faire qu'avec des participants connus pour ne pas déranger n'importe qui.

Le cas où une intervention reste sans réponse ne devrait pas provoquer d'agressivité puisqu'il est difficile de savoir s'il s'agit véritablement d'une non réponse ou d'un problème technique.

Les participants doivent également avertir les destinataires et recueillir leur consentement avant d'envoyer de trop grandes quantités d'information.

La publicité est réprouvée, les invitations en masse (à rejoindre un autre canal), les messages privés envoyés en masse, les annonces pour d'autres serveurs, ainsi que les messages commerciaux sont réprouvés.

Il s'agit là du recensement théorique des règles explicites associées au site (et qui

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

sont valables pour les deux canaux étudiés). Évidemment toutes ces règles sont enfreintes, volontairement ou par méconnaissance et quelquefois par jeu, nous aurons l'occasion d'en discuter plus loin, et nous verrons que la pratique est plus souple que ce que les règles édictées permettent d'attendre.

6.1.1.2. Participants

Au sujet des participants, les paramètres à prendre en considération pour caractériser les interactions sont leur nombre, leurs caractéristiques individuelles, et les relations qu'ils entretiennent.

Le nombre de participants engagés dans une conversation joue sur la structure de la conversation, des interactions et des échanges. Selon André-Larochebouvy (1984:47) "la conversation triadique est fondamentalement différente de la conversation dyadique, tandis que les conversations à quatre et plus se réduisent à des structures dyadiques et triadiques". Dans les canaux IRC, le nombre de participants varie en permanence. Comme dans un bar, des gens entrent et sortent à tout instant. Les *logs* sont l'enregistrement des interactions des participants dans le canal principal. La différence avec ce que serait l'enregistrement des interactions dans un bar réside dans le fait que les interventions de chacun dans le canal principal sont perçues de tous tandis que les interactions privées entre les connectés à un canal sont insoupçonnables sauf mention explicite dans le canal principal. On note également qu'un participant peut alors mener plusieurs interactions en parallèle. C'est le cas dans l'extrait suivant, dans lequel le locuteur PBW est engagé dans trois unités différentes, que nous ne présentons pas entièrement pour ne pas alourdir l'exposé.

***Extrait du Corpus F1 *PAS: T'es pas encore couché PBW??????? %suj: DORM %add: PBW *PBW: y'a pas d'raison que je sois seule à bosser ;) %suj: TPEH %add: PAG *PAG: rha %suj: TPEH %add: PBW *PBW: mais si PAS ;) ... je suis levée ;) %suj: DORM %add: PAS *PAK: s %suj: INDET %add: INDET *PAS: rhaaaaaaaaaaaaaaaa PAG :)))))) %suj: WAZZZ %add: PAG *PBU: et toi PBW t'as fini tes impôts??? %suj: TFTIP %add: PBW *PAG: rhaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa mdr %suj: WAZZZ %add: PAS *PAS: PBW Tu dors 4 heures par nuit??????????? %suj: DORM %add: PBW *PBC: bonjour P97 %add: P97 *PBW: j'en ai encore PBU... mais j'arrive à la fin ;)))))))) ouffffffffffff %suj: TFTIP %add: PBU *PBN: hooooooooooooooooooooooooo %add: ALL *PAS: PAG Rhaaaaaaaaaaaaaaaaaaa

conversationnel qui fournit des indices de la distance qui existe entre les interlocuteurs [203] . Rappelons qu'en théorie, les participants d'un canal peuvent se connaître réellement, se connaître seulement par les interactions qu'ils ont menées dans le canal ou dans d'autres lieux sur le réseau, qu'ils peuvent éventuellement s'être rencontrés réellement *a posteriori*, ou ne pas se connaître du tout. On considérera deux types de locuteurs : les locuteurs privilégiés, qui sont intéressés à participer au canal principal et semblent pour la plupart y participer régulièrement et les locuteurs occasionnels qui utilisent le canal principal comme un passage vers des canaux privés.

6.2. Gestion des tours de parole

L'analyse conversationnelle pose comme principe de la conversation l'alternance qui dans une situation idéale produit un équilibrage relatif de la longueur des tours et de la focalisation du discours. Cela signifie que dans une interaction le locuteur courant a le droit de conserver la parole pendant un intervalle de temps (au-delà duquel un allocutaire pourra réclamer justement la parole) et donc le devoir de la céder, l'allocutaire étant alors en devoir de la prendre. Le tour est donc borné par deux changements de locuteur. La situation idéale dégagée par les chercheurs en analyse des conversations veut ainsi qu'une seule personne parle à la fois, et ce, que ce soit sans négociation, ou avec négociation implicite ou explicite. Dans les interactions verbales sur IRC, cette situation n'est pas négociée, elle est imposée par le dispositif. En effet, les lignes de message apparaissent séquentiellement dans la fenêtre d'activité du canal, dans leur ordre d'arrivée sur le serveur [204] , il ne peut donc y avoir de chevauchement [205] , et d'autre part, les messages s'affichant d'un seul bloc, il n'y a pas d'attente de la fin de la production d'un locuteur.

De même, la situation idéale ne présenterait pas de silences. Dans les IRC il est difficile de considérer cette question dans la mesure où des difficultés techniques peuvent entraîner des silences qui ne reflètent pas le silence des participants (silence, il y a, en réception seulement pour tout ou partie des utilisateurs). D'autre part, les lignes systèmes sont produites assez régulièrement par les différentes opérations que réalisent les utilisateurs, la fenêtre d'un canal est très rarement figée. Lorsque cela se produit, il arrive

[203] Les chercheurs en analyse conversationnelle qualifient la relation interpersonnelle à l'aide des concepts de relation horizontale et verticale. *relation horizontale* désigne le lien qui unit les interlocuteurs au plan de la distance (des facteurs externes – s'ils se connaissent, appartiennent à un groupe social particulier – ou internes – négociés lors d'interactions – peuvent déterminer ou modifier ce lien…). *relation verticale* désigne les rapports hiérarchiques entre les interlocuteurs (des facteurs externes – biologiques, socioprofessionnels – ou des facteurs internes – quantité de parole, nombre de prises de parole, nombre d'intervention initiatives, indices proxémiques et kinésiques…) – cf. Kerbrat-Orecchioni (1995).

[204] Deux paquets d'informations ne peuvent arriver strictement en même temps.

[205] Deux interventions peuvent avoir été produites dans le même temps, mais cette situation, si elle peut être inférée, n'est pas physiquement représentée et cela ne perturbe pas ni la production, ni la perception des messages.

en vertu de la loi du droit d'auteur.

que les locuteurs y fassent référence ultérieurement, ainsi que l'illustre l'extrait du Corpus F8 ci-après.

Extrait du Corpus F8 *PAC: houla le gros blanc d'un coup lol %add: ALL

On peut également signaler que les locuteurs sont plus tolérants que dans la conversation en face à face en ce qui concerne le délai entre l'apparition de leur message dans la fenêtre d'activité du canal et l'apparition du message réactif d'un allocutaire, les manifestations d'impatience sont rares.

Extrait du Corpus P3 *PBZ: tu ma pas repondu PAK ..... tes qui? %add: PAK

Le réglage de l'alternance est un point crucial de la structure conversationnelle. En règle générale, ce sont les participants qui le négocient au long de l'interaction. Kerbrat-Orecchioni (1996:30) rappelle que "le changement de tour doit normalement s'effectuer à un point de transition possible". Ce point de transition [206] doit être inscrit dans l'énoncé. Ces signaux de fin de tour, ou de transition possible peuvent dans la conversation prototypique être de nature verbale ou non verbale.

Les signaux de nature non verbale relevés dans la littérature sont des signaux prosodiques (courbe intonative, ralentissement du débit, chute de l'intensité articulatoire) ou mimogestuels (regard soutenu, arrêt des gestes accompagnant la production verbale), il va sans dire qu'ils ne concernent pas notre cas. Les signaux de nature verbale sont essentiellement la complétude syntactico-sémantique, et la complétude pragmatique (l'énoncé doit être pertinent). Certains petits mots sont également analysés comme des morphèmes signalant [207] la clôture (*bon, voilà*) ou appelant l'allocutaire à prendre la parole (*hein*). Dans la conversation, il peut arriver que l'alternance ne se produise pas à une place transitionnelle. Ces cas sont dénommés chevauchements, interruptions et intrusions.

Dans nos corpus, nous l'avons dit, le dispositif met en défaut l'analyse. Ce n'est que sur des signaux de nature verbale que peut s'appuyer l'alternance des tours de parole.

Cependant la situation est complexifiée par un phénomène que nous avons attribué à la contrainte temporelle. Un utilisateur, sauf problème technique, erreur de frappe ou intervention musclée d'un opérateur, a la pleine maîtrise de la longueur de son message. Toutefois, pour être pertinentes, les interventions réactives doivent être temporellement proches de leur source, donc d'une longueur limitée. Or, tant que le locuteur émetteur n'a pas signifié l'envoi du message (en appuyant sur une touche spéciale), ce dernier n'est pas visible des autres utilisateurs. Nous l'avons dit, la stratégie employée pour pallier les délais imposés est alors parfois de découper les interventions. Notons que la netiquette formule précisément des recommandations risquant de susciter ces stratégies de découpage : les messages trop longs doivent être évités pour ne pas monopoliser l'espace de conversation, la visualisation de l'espace d'activité étant variable et pouvant ainsi amener l'écran d'un utilisateur à défiler d'un coup avant qu'il ait pu prendre connaissance de toutes les interventions. Or, les locuteurs n'emploient aucun signe systématique spécifique pour marquer que la parole n'est pas à prendre.

[206] "place transitionnelle" selon les termes de Kerbrat-Orecchioni (1996).

[207] Exemples de Kerbrat-Orecchioni (1996).

en vertu de la loi du droit d'auteur.

D'autre part, le principe de l'alternance suppose que la fin du tour d'un locuteur génère le début de celui d'un autre. Dans le cas d'interactions impliquant plus de deux participants, ainsi qu'il en est souvent le cas dans nos corpus, la sélection du successeur est problématique. Les deux cas de figure relevés par la littérature [208] se trouvent : il s'agit de la sélection du successeur par le locuteur, et de l'autosélection du successeur.

6.2.1. Statut du non locuteur, identification du/des destinataire(s)

Le statut du non locuteur (non scripteur) dans les communications IRC est assez difficile à évaluer et requiert qu'on apporte quelques précisions sur la façon de considérer l'objet. Une conversation n'est pas un objet statique, c'est un objet qui se construit dans le temps. Lorsqu'on parle de *non locuteur* soit cela suppose qu'on considère la conversation à un point de son déroulement, et donc que les locuteurs (en puissance) qui ne sont pas en train de produire un message se voient attribuer ce statut ; soit cela suppose qu'on ait délimité l'objet et qu'on soit capable de déterminer parmi les locuteurs (en puissance) lesquels ont produit des messages constitutifs de la conversation. En vérité, la situation est très complexe, et il est certain que, pour un locuteur (en puissance) donné, le statut par rapport au groupe conversationnel varie moins que son rôle à l'intérieur du groupe conversationnel.

Ainsi, Goffman (1973) distingue les participants ratifiés, des simples spectateurs. Les critères qui permettent la distinction ne sont pour la plupart pas applicables au dispositif IRC. En effet, les destinataires directs et indirects, considérés comme faisant partie du groupe conversationnel, le sont par l'arrangement physique du groupe et par le comportement non verbal de ses membres. Il s'agit là d'indices proxémiques, de questions d'orientation du corps et de direction des regards qui ne sont absolument pas des indices visibles dans notre cas ni pour les locuteurs eux-mêmes ni pour l'observation. Dans les IRC, on ne pourra donc identifier avec certitude et précision quels sont les participants spectateurs. Quant au statut de participant ratifié, il est également difficile à établir, dans la mesure où il est difficile de connaître l'histoire conversationnelle des participants d'un canal [209] .

À l'intérieur du groupe conversationnel, lorsqu'une intervention a été produite, l'identification du (ou des) destinataire(s) direct(s) de cette intervention se base à la fois sur des indices non verbaux (orientation du corps, direction du regard) et sur des indices verbaux tels que des termes d'adresse dans des segments métacommunicatifs. Ces derniers permettent de repérer l'identité du destinataire de façon non ambiguë. Dans les situations traditionnelles, on relève que ces segments métacommunicatifs ne sont pas systématiques [210] .

Dans nos corpus, l'émetteur d'une intervention est identifié par le fait que toute ligne

[208] Sacks *et al.* (1974), Kerbrat-Orecchioni (1996), Traverso (1995), Witko-Commeau (1995).

[209] On peut en effet considérer que l'inventaire des membres d'un groupe conversationnel soit plus important que celui des participants ayant produit des messages. Un locuteur connecté, connu des membres produisant des messages, pouvant être considéré par ces derniers comme destinataire de leurs interventions.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

de message produite comporte dans la fenêtre d'activité en prélude le surnom du locuteur, borné à gauche et à droite par des caractères spéciaux. Il est très souvent possible d'identifier le destinataire direct d'un message, puisque la désignation du destinataire par l'émetteur est fréquente et elle se fait par des moyens linguistiques : les messages comportent très souvent un segment métacommunicatif vocatif permettant d'identifier l'allocutaire.

Extrait du Corpus F3 *PCB: ou puis je trouver TOUTES les commandes du dos please? %suj: CdDOS %add: ALL (1) *PAJ: PCB sur le manuel:) %suj: CdDos %add: PCB (1) *PCB: et si je l ai pas %suj: CdDos %add: PAJ (1) *PAJ: PCB bah c'est pas bien, t'es un pirate :o)))) %suj: CdDos %add: PCB *PCB: nooooon %suj: CdDos %add: PAJ *PDL: mais vous etes toutes la journee sur net? voua n'avez pas mal au yeux? %suj: zetnet %add: ALL *PBD: PAJ: ce serait-y pas l'hopital que si fout de la charite ? :p %suj: CdDos %add: PAJ *PCO: PDL on commence a avoir les yeux rouges %suj: zetnet %add: PDL

Une autre possibilité d'identification, mise en oeuvre par l'émetteur d'un message est le procédé de la reprise. Ce dernier a plusieurs formes :

Il peut être une répétition du message interventif auquel le locuteur répond, ainsi qu'on en trouve un exemple dans l'extrait ci-dessous :

Extrait du Corpus F5 *PBE: c'est quoi l'extension des domaines marocains...tunisiens? %add: ALL *PBE: .ma ? %add: ALL (10) *PAX: <PBE> c'est quoi l'extension des domaines marocains...tunisiens?----------que veux-tu dire? %add: PBE

La reprise peut ne concerner qu'une partie de l'énoncé auquel il est fait écho.

Extrait du Corpus F8 *PAC: Dell ils ont une politique tres agressive question prix... mais est-ce fiable? %add: PAS [211] *PAS: fiable oui. mais très très mauvais pour le traitement des commandes %add: PAC

Dans cet extrait, un item lexical repris, *fiable* permet d'identifier l'énoncé auquel le message répond et permet donc de connaître avec certitude l'allocutaire. Mais la reprise peut également concerner une unité phrastique comme c'est le cas ci-dessous.

Extrait du Corpus F5 *PEN: 14 tain c'est cherrrrrrrrrrrr %add: PDV (1) *PDV: oui c cher %add: PEN

D'autres messages présentent, en réponse à une question totale, qui appellerait donc un *oui* ou un *non,* une formulation complète ainsi qu'on les trouve dans les ouvrages de didactique de français :

Extrait du Corpus F1 *PBW: PBM ;)))))))))))))) tu vas bien ;) %add: PBM (3) *PBM: salut PBW :)) xxxxx %add: PBW (1) *PBM: oui je vais pas trop mal et toi? :) %add: PBW Extrait du Corpus P3 *PAJ: PBE tu es le diable toi %add: PBE (7) *PBE: non je suis pas le diable %add: PAJ

[210] Ils le sont en revanche dans la plupart des situations de communication dans lesquelles les interlocuteurs ne disposent pas de suffisamment d'indices situationnels, ou disposent d'indices peu fiables, pour identifier le locuteur et son allocutaire, ou encore lorsque les conséquences potentielles d'une méprise sur le destinataire et/ou l'émetteur sont coûteuses.

[211] L'allocutaire est ici identifié par un principe de cohérence.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

On remarque que lorsque le message initiatif est éloigné du message réactif, les procédés permettant de réduire les possibilités d'interprétation sont additionnés. Ainsi, dans les extraits ci-dessous, le vocatif et la reprise sont combinés.

Extrait du Corpus P3 *PAS: PAG tu ronronnes? %add: PAG (8) *PAG: PAS ben oui je ronronne %add: PAS Extrait du Corpus F8 *PBF: alors ton depart approche ? :) %add: PEF (12) *PEF: PBF oui ca approche :)) enfin j'ai encore du temps %add: PBF

En effet, ce double procédé est presque indispensable au bon déroulement de l'interaction dans les cas où le locuteur auquel s'adresse le message réactif a produit plusieurs messages auquel l'auteur du message réactif est susceptible de répondre.

Quelquefois, les locuteurs jouent sur ce principe. Dans l'extrait présenté ci-après, deux unités conversationnelles avec des inventaires de participants différents sont en cours. L'une est relative à des problèmes de grève des trains, l'autre à une effraction commise sur le véhicule d'un participant (PCK). L'inventaire des participants pour les deux unités conversationnelles est assez stable à ce stade. Pour l'unité conversationnelle notée OPEL, les participants sont PCK et PBS, tandis que pour l'unité codée SNCF il s'agit de PFN, PCJ, et PIB.

Extrait du Corpus F7 *PCK: oui ils m'ont niqué la serrure avant passager %suj: OPEL %add: PDS *PFN: a ta place j'irais a la gare de bonne heure..... %suj: SNCF %add: PCJ *PCK: bon on change de sujet stp %suj: OPEL %add: PDS *PDS: a carrement %suj: OPEL %add: PCK *PDS: pas du boulot pro ca %suj: OPEL %add: PCK (1) *PCJ: j'peux meme pas, j'ai cours, deja que je vais partir en plein milieu... %suj: SNCF-CRS %add: PFN *PDS: fallaiçs mettre un system linux embarqué pour l'alarme %suj: OPEL %add: PCK *PCK: heureusement que j'avais une alarme %suj: OPEL %add: PDS *PCK: qui les a dérangé... ils ont rien pris %suj: OPEL %add: PDS *PDS: ah oki %suj: OPEL %add: PCK *PDS: tu les as butté? %suj: OPEL %add: PCK *PIB: chanceux va %suj: SNCF-CRS %add: PCJ *PIB: PCJ %suj: SNCF-CRS %add: PCJ *PDS: pas changed18nceux... intelligent %suj: OPEL %add: PIB

Tous les messages ne sont pas marqués de la sorte, aussi, le plus souvent les locuteurs doivent se baser sur des indices de cohérence. Ces derniers appellent non seulement un système de décodage d'indices linguistiques, mais également un système d'inférence sur les intentions d'autrui, *i. e.* un modèle de l'allocutaire et des interventions initiatives et réactives qu'il est susceptible de produire.

La distinction destinataire privilégié/destinataire secondaire est également difficile à établir dans la mesure où aucun autre indice que ceux apparaissant dans les messages n'existe. Il est toutefois des cas qui manifestent qu'un participant se considère comme un destinataire secondaire d'une intervention adressée à un autre locuteur.

Extrait du Corpus F1 *PBW: PAY.... tu sais que le québec peu aussi être très ensoleillé ? %add: PAY *PBW: peut sera mieux %add: PAY *PAZ: lol PBW :) laisse moi rire %add: PBW

Ce sont là des productions de rires, ou de frimousses, adressées au locuteur émetteur de l'intervention pour laquelle un participant se considère destinataire secondaire.

en vertu de la loi du droit d'auteur.

6.2.2. Changement de tour et successeur

Nous avons donc vu que l'alternance des tours de parole repose à la fois sur des questions de place transitionnelle et de sélection du locuteur "suivant". Évidemment, les principes sont très souvent transgressés, volontairement ou non, les indices permettant l'identification des places transitionnelles ou du successeurs étant potentiellement flous.

Dans l'extrait ci-dessous, le message action "interrompt" le message de PEP qui n'est manifestement pas complet syntaxiquement. On ne peut toutefois pas incriminer le locuteur PAW d'avoir commis volontairement une interruption. Ce premier type de contrevenue au principe de l'alternance est en effet induit par le dispositif qui ne garantit en rien qu'un message envoyé n'arrive juste après un message non fini alors que l'émetteur du second message n'en avait pas pris connaissance au moment de l'envoi. D'autre part, aucun signe n'étant disponible pour connaître les intentions de prise de parole d'autrui, il est raisonnable de penser que la règle de l'alternance ne s'applique qu'au niveau des interventions initiatives et réactives, c'est-à-dire qu'un locuteur n'attend en principe pas son tour pour envoyer un message, il n'attend que matière à agir et réagir.

Extrait du Corpus F6 *PEP: ce qu'il y a de bien avec le thé chaud %suj: THE %add: ALL *(action): PAW n'a juridiquement aucune envie de parler... %suj: JASR %add: PDZ *PEP: c'est que ça brûle %suj: THE %add: ALL *PCN: lol)) %suj: THE %add: PEP *PCM: ssss %suj: INDET %add: INDET *PCN: mais de l eau froide %suj: THE %add: PEP *PEP: et qu'une brûlure, ça occasionne des arrêts maladie %suj: THE %add: ALL, PCN

Dans ce même extrait, une intervention apparaît à une place non transitionnelle. La seconde intervention de PCN survient après l'intervention en deux messages de PEP, qui semble syntaxiquement, sémantiquement et pragmatiquement complète, et qui pourtant ne l'était pas, ainsi que le montre le troisième message de PEP.

Il survient effectivement, dans nos corpus, très souvent des cas où l'alternance n'apparaît pas en place transitionnelle, à cause des contraintes imposées par le dispositif qui font que souvent on observe des croisements tels que ceux apparaissant dans l'extrait ci-dessous.

Extrait du Corpus P3 *PCC: yo qlq'1 peut m; envoyer un ping [212] ? %suj: EPIN %add: ALL *PAU: PBR melle de toi ton cul hein! %suj: OUBV-OPP %add: PBR *PCA: lol %suj: OUBV-OPP %add: PAU *PAC: tien attrape PCC %suj: EPIN %add: PCC *PAS: bah desole le sex fait parti de la vie je ne voulais pas choquer ;p~ %suj: OUBV %add: PBX *PBX: tiens un bouquet d icmp [213] %suj: EPIN %add: PCC (PAC vient de kicker PCC avec le message " ping :)))))) 4*12*4* 12S4k12A4 S12c4R12i4P12t 4*12*4*¤") *PBX: comme c est sympa %suj: EPIN %add: PCC *(action): PCA éclate un néon sur la tête de PAU! %suj: OUBV-OPP %add: ALL (PAU) *PAS: c jolie :) %suj: EPIN?OUBV-OPP %add: PAC?PCA *PAS: ca fait

[212] ***"Ping" est un des outils de déboguage réseau les plus utiles. Son nom lui vient d'un sonar de recherche de sous-marin qui envoie un son et attend un ping en retour.***

[213] ***Internet Control Message Protocol, documenté dans la RFC 792***

en vertu de la loi du droit d'auteur.

toujours plaisir %suj: EPIN?OUBV-OPP %add: PAC?PCA

Dans cet extrait, plusieurs conversations sont menées en parallèle, à tel point qu'il est parfois difficile de rattacher certaines interventions à des unités thématiques. En cas de doute, nous avons marqué dans les corpus les différentes possibilités, et noté "INDET" les cas où nous n'avons trouvé aucune possibilité cohérente. En règle générale, cela ne semble pas perturber la communication entre les internautes, en tout cas pas de façon explicite.

Cependant, cela se produit dans l'extrait ci-dessous. Nous avons ajouté dans la présentation de cet extrait dans une colonne à gauche des numéros de ligne permettant de faire ressortir lesdits croisements. Les numéros en gras sont les numéros des messages, les numéros entre parenthèses ceux des messages auxquels ils répondent.

Extrait du Corpus F5 1 *PEF: t'essayais d'ouvrir le fichier depuis le CD %add: PEE 2 (1) *PEE: oue PEF mdr %add: PEF 3 (2) *PEF: c ptet un prob d'acces en lecture seule alors... %add: PEE 4 *PEF: copie le sur ton HD et enleve l'attribut Read-Only du fichier %add: PEE 5 (3-4) *PEE: .....tu me parle la france s'il te plait lol %add: PEF 6 *PEF: puis réessaie %add: PEE 7 (4) *PEE: je peux pas le copier sur mon hd lol %add: PEF 8 *PEE: j'ai pas la plce lol %add: PEF 9 (5) *PEF: commence par copier le fichier sur ton disque... %add: PEE 10 *PEF: arf lol %add: PEE 11 (6) *PEE: ma soeur le fera %add: PEF

Le "raté", ici, se situe au message 9, qui ne tient pas compte des messages 7 et 8, situation reconnue par le message 10, dans lequel le locuteur s'en amuse.

En ce qui concerne les cas relevés et catégorisés par la littérature, le chevauchement est dans sa conception matérielle rendu impossible par le dispositif. On pourrait toutefois considérer que le "raté" que nous venons de décrire s'analyse en ces termes. En effet, PEF a un problème de bureautique qu'il a soumis à l'assemblée. PEF questionne PEE pour tenter d'émettre un diagnostic, ce qu'il réalise en 3. Au message 4 PEF formule des instructions de contournement du problème. En 5, l'intervention de PEE concerne le code et survient alors que PEF n'avait manifestement l'intention de "rendre la parole". Les messages 7 et 8 formulent une opposition à une proposition contenue dans le message 4. En (9) enfin, conformément à la demande de reformulation PEF compose un message qui n'est pas cohérent avec les deux messages précédents. 7, 8 et 9 ont été produits dans un intervalle de temps très court qui s'apparente à un chevauchement.

L'interruption, quant à elle, n'est possible qu'à condition que le locuteur ait découpé son tour de parole en plusieurs messages. Un locuteur a donc le loisir de s'exposer à l'interruption ou de l'interdire. Dans l'extrait ci-dessous, P13 formule une intervention longue, qui ne s'expose pas à l'interruption, tandis que P26 découpe son intervention en plusieurs messages et s'expose donc à l'interruption de P21.

Extrait du Corpus P6 *P13: ensuite avant de dire que je suis une gauloise, ou que nous sommes des gaulois, tu devrais relire tes livres d'histoire... mais il est possible que tu ne saches pas lire %suj: MQGL %add: P26 *P26: hahahaha %suj: MQGL %add: P13 *P26: nan pigeonne %suj: MQGL %add: P13 *P26: si je savais apas lire %suj: MQGL %add: P13 *P21: hahahahahaha %suj: MQGL %add: P13, P26 *P26: je vois pas comment %suj: MQGL %add: P13 *P26: \je pourrais chatter %suj: MQGL %add: P13

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Notons qu'il est difficile de déterminer si une unité syntaxique a été volontairement scindée en plusieurs messages par un locuteur, ou si le locuteur attache après coup un autre segment à son message initial, profitant de la structure syntaxique amorcée.

Les interruptions relevées peuvent être de nature coopérative ou de véritables FTAs.

Le phénomène de l'intrusion est assez difficile à évaluer. Il concerne les cas où un locuteur illégitime parasite le circuit interlocutif. Cette situation peut se manifester explicitement, un locuteur en faisant la remarque, ainsi que l'illustre l'extrait ci-dessous.

Extrait du Corpus P2 *PAB: PAG on t a pas demandé bordel :/

Dans l'extrait ci-dessous se présente une subtilisation du tour de parole :

Extrait du Corpus P6 *PAS: et toi PCA qui es tu [214] ? %suj: CQJX %add: PCA *PAU: c ma meuf :/ %suj: CQJX %add: PAS

Ces subtilisations ne se font pas toujours au profit de celui qui prend la parole. L'extrait ci-dessous montre plusieurs cas de subtilisation de la parole.

Extrait du Corpus F8 *PDW: c est deja pas mal d etre beau de l interieur PAS %add: PAS (5) *PBL: on doit faire quoi kan on est moche de l interieur et de l exterieur? %add: PDV (1) *PDT: on fait son possible PBL ;) %add: PBL *PDW: tu crois qu il y a des gens moches de l interieur et de l exterieur aussi? %add: PBL *PAK: oui... moi :) %add: PDW

Le modèle hiérarchique définit des unités dialogales – l'interaction, l'échange, la séquence – et des unités monologales – l'intervention, l'acte de langage. Toutefois, Jeanneret (1999) explore un certain nombre de cas qui n'entrent pas dans ce moule. Elle distingue la production d'une unité discursive par au moins deux locuteurs de l'articulation en échange de deux interventions. Pour Jeanneret (1999) "il y a coénonciation [215] de deux segments quand ils ne peuvent se résoudre ni en échange, ni en paire adjacente."

Jeanneret (1999:13-14) relève dans ses corpus 7 types de coénonciations :

un connecteur est attaché au premier segment	1.
un connecteur est attaché au 2e segment	2.
coénonciation avec du discours rapporté	3.
relative produite par un locuteur différent de celui de l'antécédent	4.
le deuxième segment est précédé d'une reprise d'un matériel du premier segment	5.
reprises plus larges	6.
coénonciations exposées (le deuxième segment donne lieu à une séquence lattérale).	7.

[214] Notons qu'une question sur l'identité d'un allocutaire ne signifie pas forcément que les locuteurs ne se "connaissent" pas, mais aussi qu'ils ne se "reconnaissent" pas, les changements de surnoms étant fréquents.

[215] La coénonciation apparaît dans la littérature sous différents termes. Jefferson (1973), et Sacks (1992) parlent de *collaborative sentences*. André-Larochebouvy (1984) parle de *complétion du tour de parole*, Loufrani (1981 et 1985) de *locuteur collectif*. Enfin, Gülich (1986), Lerner (1991) et Coates (1994), cités par Jeanneret (1999) utilisent respectivement les termes d'*achèvement interactif du discours*, d'*unité syntaxique produite conjointement* et de *construction conjointe d'énoncés*.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Trois fonctions sont identifiées. Jeanneret (1999:112) indique que la fonction interactive entre les segments coénoncés peut être rituelle, argumentative, ou de reformulation non paraphrastique. Elle est rituelle dans le cas où la complétion intervient parce que le locuteur du segment coénoncé pense que le segment auquel il l'attache n'est pas complet et est menaçant pour la face de l'interlocuteur ou conduirait à un malentendu. Ainsi Jeanneret (1999:122) rappelle que "Toute intervention se doit d'être complète pour ménager la face de son interlocuteur et permettre aux interlocuteurs de bien se comprendre". La fonction est argumentative dans les coénonciations composées *i. e*. commençant par un connecteur. Enfin, elle est de reformulation non paraphrastique selon Roulet (1987:115) lorsqu' "un énonciateur tente de mieux satisfaire à la complétude interactive en présentant l'intervention principale comme une nouvelle formulation liée à un changement de perspective énonciative indiqué par le connecteur d'un premier mouvement discursif […]". Ce type se marque obligatoirement par un connecteur reformulatif et selon Jeanneret (1999:116) "implique une réinterprétation de l'énoncé source par l'énoncé doublon et donc une subordination rétroactive du premier au second". Tandis que la reformulation paraphrastique "se marque par une proximité sémantique de l'énoncé doublon par rapport à l'énoncé source".

Évidemment cela pose le problème de la complétude d'une unité et de la façon dont les locuteurs l'envisagent. Les phénomènes de coénonciation sont assez rares dans nos corpus. Ils ne sont jamais invités par le locuteur de la première partie de l'unité. On notera que le connecteur n'est pas présent dans tous les cas de complétion :

Extrait du Corpus F4 *PCA: bah on se fera un chapon a la place :) %add: PDR (1) *PDR: un chapon en méchoui..... c'est peu ;) %add: PCA (4) *PCA: a la croute de sel :) %suj: jpqs %add: PDR *PDR: tu fais ca comment PCA? %add: PCA (3) *PCA: tu entoure ta bestiole avec une pate faite de gros sel et d'eau.... et tu enfournes %add: PDR *PDW: 30 Minutes pa rlivre %add: PDR (1) *PDR: ca doit être bon PCA :) ... moi je l'enduis de moutarde ;) %add: PCA *PCA: ca cuit dans son propre jus... excellent %add: PDR

L'extrait suivant présente plusieurs cas de coénonciation. Les locuteurs s'interrogent à propos du rôle des opérateurs et formulent des propositions :

Extrait du Corpus F7 *PJQ: ceci n'est pas notre rôle %add: PFL *PFL: lol PJQ %add: PJQ *PFL: j'y crois vraiment =) %add: PJQ (2) *PJQ: moi aussi :) %add: PFL, PJQ, (1) *PBN: 'servir et protéger' %add: PFL, PBP, PJQ (6) *PBP: servir prtoger et surtout etre courtois %add: PBN (1) *PJQ: servir et faire des bavures %add: PBP, PBN, *PJQ: lolol %add: PBN, PBP, *PBP: oui %add: PJQ *PFL: et oper tous les heureux possesseurs du ratascript même si ils connaissent que dalle à irc %add: PBN, PJQ, PBP,

Le premier cas de coénonciation utilise un procédé de rattachement par reprise du matériel énoncé. Le deuxième cas de coénonciation ne reprend pas la totalité des termes et détourne l'orientation argumentative de l'intervention. Ce détournement est amplifié par un autre locuteur, PFL, qui avec un attachement classique par un connecteur poursuit le texte de PJQ. Les phénomènes de coénonciation observés dans le corpus ont très souvent un aspect ludique. Ainsi, dans l'extrait suivant, quelques locuteurs tentent de convaincre un aop d'effectuer une opération qui ne fait pas partie de ses prérogatives.

Extrait du Corpus F7 *PBP: oui on diras rien %add: PBS *PBS: trop tard, le big

en vertu de la loi du droit d'auteur.

boss a tout loggé ce soir en première instance je perd mon aop et du coup mes chances de founderage [216] de dans deux ans %add: PBP *PBP: tain je suis plus op et j'ai toutes les pubs ;o)) %add: ALL *PHQ: qui auraient lieux ds un an %add: PBS *PBS: oui exact %add: PHQ

La relative est produite par un locuteur différent de celui de l'antécédent générant une incohérence, toutefois validée par l'auteur du premier segment.

Jeanneret (1999) identifie des phénomènes de coénonciation en réparation. Dans nos corpus la réparation est le plus souvent autoeffectuée et pallie des problèmes de frappe essentiellement.

6.3. Unités conversationnelles

Nous avons questionné le statut des locuteurs dans le cours de l'interaction. La question se pose ensuite d'identifier des unités conversationnelles afin de pouvoir dégager les caractéristiques de la structuration de la conversation dans les IRC. Les concepts descriptifs proposés par Roulet (1981, 1986, 1989) et Kerbrat-Orecchioni (1990) pour la conversation sont ceux qui serviront de base à cette étude rapide. Le modèle hiérarchique initié par l'école genevoise distingue cinq rangs d'analyse principaux. L'incursion ou interaction est une unité définie par une certaine stabilité des participants et du cadre spatio-temporel, elle s'analyse en séquences, une séquence étant une suite d'échanges ayant une cohérence sémantique et pragmatique. L'échange se compose d'interventions [217] , unités fonctionnelles indépendantes de la notion de tour de parole. Il existe plusieurs types d'interventions : initiative, réactive, évaluative. L'intervention n'est pas la plus petite unité. L'unité de base est l'acte de langage. L'intervention, ainsi, peut être composite et comporter un acte directeur et des actes subordonnés.

Évidemment le découpage de nos corpus en "séquences" pose problème. En effet, le nombre de participants est en constante variation, et la prise de parole n'est pas contrainte par le dispositif. Ceci fait que, une suite quelconque de lignes de message peut paraître, à un lecteur non averti, tout à fait incohérente. Les participants semblent pourtant s'y retrouver. L'hypothèse est que les locuteurs catégorisent les lignes de messages qui apparaissent séquentiellement en fonction d'une représentation de la structure de la conversation en face à face. Ainsi, nous avons découpé le corpus en nous basant sur la structure prototypique de la conversation.

6.3.1. Séquences d'ouverture et de clôture

[216] Création d'un canal public.

[217] Notons que jusqu'ici, le terme d'intervention n'a pas été employé en ce sens. En règle générale, une ligne de message contient une intervention, mais il se peut qu'une intervention soit répartie sur plusieurs lignes de message, ou qu'une ligne de message porte plusieurs interventions.

en vertu de la loi du droit d'auteur.

Le modèle hiérarchique propose un découpage des interactions en séquences ou épisodes.

L'interaction est définie par une stabilité du cadre spatio-temporel et des participants engagés. La séquence est définie comme un ensemble d'échanges ayant une cohérence sémantique et pragmatique.

Il semble difficile d'après nos corpus de donner une représentation des unités interactionnelles qui s'y tiennent, correspondant au cas prototypique c'est-à-dire commençant par une séquence d'ouverture et se terminant par une séquence de clôture.

Tableau (6-1) – Structure prototypique de la conversation

Séquence d'ouverture (salutations, salutations complémentaires)
Corps
Séquence de clôture (salutations, vœux…)

Selon Goffman (1974), les rituels d'accès que sont les salutations et les adieux permettent respectivement l'établissement du contact, et l'accompagnement de la séparation. Ces rituels consistent donc en l'établissement d'un accord entre les partenaires conversationnels sur les limites de l'interaction. Ils relèvent de la politesse.

Dans nos corpus, les séquences d'ouverture et de clôture interviennent en général indépendamment des thématiques conversationnelles. Ce fait n'est pas propre à l'IRC. On signale en effet dans la littérature que la modification du nombre des participants est liée à la réalisation des rituels d'ouverture et de clôture lorsque les interactions ne sont pas formelles.

Dans l'extrait ci-dessous, l'arrivée d'un participant génère de nombreuses salutations, elle ne met cependant pas fin aux interactions en cours.

Extrait du Corpus F1 *PBW: PBU.... tu as fini de pelleter chez toi????.... tu veux te tenir en forme????... alors... viens pour déblayer chez moi :) %suj: TPDC %add: PBU (10) *PBU: voudrais bien PBW mais je ne peut pas je recommence à travailler à matin à 8 heures %suj: TPDC %add: PBW (2) *PBW: bon courage PBU.... tu dois avoir un skidoo alors ;) %suj: TPDC %add: PBU (1) (PAS entre dans le channel) *PAS: hééééééé %add: ALL *PBW: PAS ;)))))))))))))) tu vas bien ;)))))))))))? %add: PAS *PAS: Lu PAY :)))))))) %add: PAY *PAS: Lu PAZ :)))))))) %add: PAZ *PAG: PAS :)) %add: PAS *PAS: PBW :))))))))))))))) %add: PBW *PAB: slut PAS %add: PAS *PAZ: coucou PAS :) %add: PAS (1) *PBU: ouin mais c pas vraiement un temps pour faire du skidoo la neige est pas bonne %suj: TPDC %add: PBW *PAS: Oui mais j'ai pas dormi des masses lololol %add: PBW (3) *PAS: Salut PAB :)))))))))))) %add: PAB *PBD: PAS kikou %add: PAS (1) *PAS: Salut PCJ :)))))))) %add: PCJ *PAS: PBD :))))))))))))))))) %add: PBD *PCJ: hello PAS %add: PAS *PBW: c'est vrai PBU... elle est collante et pesante ;(%suj: TPDC %add: PBU

Le départ d'un participant engagé dans une unité conversationnelle, lorsqu'il est accompagné d'une séquence de clôture, ne clôt pas forcément l'unité conversationnelle. En effet, si cette unité intéresse plus de deux participants, elle peut se poursuivre. L'extrait

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

présenté ci-dessous illustre ce cas. Nous n'avons évidemment pas reproduit l'unité entière. Celle-ci débute par une question d'un utilisateur concernant des problèmes rencontrés par les opérateurs du canal puis se centre progressivement sur le comportement discutable d'un opérateur particulier.

Extrait du Corpus F2 *PBT: PAD j'etais la qd le probleme s'est passé %add: PAD (1) *PAD: PBT bah il arrete pas de faire chier VVV à chaque fois qu'elle vient %add: PBT *PAL: PAG il a abuse de son pouvoir et s'est entete dans ses conneries %add: PAG (2) *PBT: les pb relationnelles entre un aop et un user le doiventy pas se terminer pas une shitlist %add: PAG (1) *PAG: a ok... %add: PAL, PBT (2) *PAG: tu sais je ne suis pas tous les ragots de ### moi lolol %add: PAL *PBT: bon aller je me casse %add: ALL (1) *PBT: bye bye les enfants %typ: clôture %add: ALL *PAG: a+ PBT %typ: clôture %add: PBT *PAL: A+ PBT %typ: clôture %add: PBT *PAD: ah ok VVV qui se fait foutre dehors à chaque fois qu'elle vient c un ragot %add: PAG *PBL: a+ aop %add: PBT *PAL: PAG c pas un ragot c une realité %add: PAG

L'utilisateur PBT est engagé dans cette unité depuis son commencement. Après une série d'échanges, il annonce son départ puis produit une formule de clôture à laquelle trois des participants répondent avant de poursuivre l'unité conversationnelle en cours. Dans cet extrait, les interventions mentionnées ont été portées en gras. VVV représente le surnom d'un participant absent de la session, et ### le nom du canal.

L'arrivée d'un participant accompagné d'une séquence d'ouverture n'ouvre pas non plus forcément de nouvelles unités.

Le dispositif joue en partie sur la structure des conversations, ainsi, le fait qu'il soit un lieu ouvert autorise l'arrivée et le départ de participants à tout moment, ce qui a pour conséquence les phénomènes présentés ici. Il ne s'agit pourtant pas d'une spécificité des IRC, c'est une caractéristique des lieux ouverts [218] . Toutefois, la probabilité qu'une unité conversationnelle ne soit pas interrompue de la sorte est bien moindre avec ce dispositif. Le grand nombre de participants possibles est en partie responsable de ce fait.

On peut constater des différences dans la gestion des séquences d'ouverture et de clôture à différents niveaux.

On note une différence importante entre le nombre d'interventions pouvant être comptabilisées comme appartenant à des séquences d'ouverture (salutations et salutations complémentaires) et le nombre d'intervention appartenant à des séquences de clôture (pré-clôtures et au revoirs). Deux hypothèses peuvent être émises à ce sujet. Entrer dans un canal implique probablement qu'on cherche à entrer en contact avec des participants. Trois cas de figure se présentent alors. Première possibilité : L'utilisateur entrant est déjà venu et est en relation avec d'autres allocutaires qu'il suppose pouvoir trouver dans la liste des participants du canal pour engager une session privée. Dans ce cas, il n'a pas besoin de se manifester. Deuxième possibilité : l'utilisateur entrant n'est

[218] On notera que, dans des situations plus traditionnelles, si une interaction dans laquelle la réciprocité est faible est engagée, les séquences d'ouverture ou de clôture faisant suite ou précédant des arrivées ou des départs isolés se feront en privé dans l'audience ou pas du tout (sauf cas exceptionnels alors ressentis comme formes d'impolitesse). En revanche, moins la situation est formelle et plus le degré de réciprocité est haut, plus la place pour les séquences d'ouverture et de clôture est aisée.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

jamais venu ou est rarement venu et/ou n'est pas en relation particulière avec d'autres utilisateurs. Pour engager des conversations, qu'elles soient publiques ou privées, il est préférable qu'il se signale sur le canal principal. Troisième possibilité, l'utilisateur entrant est en relation privilégiée avec d'autres utilisateurs du canal principal ou fait partie de l'équipe des opérateurs.

Les participants qui formulent des interventions appartenant à des séquences d'ouverture sont donc à la fois des utilisateurs qui vont poursuivre une activité sur le canal principal et des utilisateurs qui vont poursuivre une activité dans des canaux privés, ces derniers ne formuleront donc pas de séquences de clôture.

On constate par ailleurs que les salutations d'ouverture sont plus fréquemment adressées que les salutations de clôture, elles sont donc en plus grand nombre.

On note une différence dans la formulation des salutations selon le destinataire et selon l'émetteur.

Les participants qui ne semblent pas entretenir de relation stable avec d'autres participants du canal ne reçoivent pas toujours et n'attendent pas forcément de salutations en retour des leurs.

Extrait du Corpus P7 *PCD: salut les nanas je suis là rien que pour vous %add: ALL Extrait du Corpus F3 *PAQ: salut qqn peut me convertir qqchose qui est en hexa? %add: ALL

Cela a lieu lorsque les locuteurs ont des buts précis, ainsi qu'on l'observe dans les exemples ci-dessus, qui illustrent deux des cas présents dans nos corpus, c'est-à-dire établir le contact avec une catégorie spécifiée d'interlocuteurs ou pour obtenir une information ou une aide précise.

Les salutations émises par ce type de locuteurs sont adressées explicitement ou implicitement à l'assemblée et ne reçoivent que sporadiquement des salutations publiques en retour, en règle générale, un opérateur s'en charge, parfois collectivement.

Extrait du Corpus F4 *PCY: Salut tout le monde %add: ALL (4) *PCY: chalu tou le monde %add: ALL *PDG: hi %add: ALL *PBN: salut tlm %add: ALL (2) *PCA: salut les arrivants %add: PCY, PDG, PDN

Dans de rares cas, un locuteur qui n'a pas été salué en retour proteste.

Extrait du Corpus F5 *PDP: bonsoir %add: ALL (2) *PBU: lu PDP %add: PDP (20) *PDP: J'ai dit BONSOIR a tous!!! %add: ALL *PAY: bonsoir %add: PDP *PBU: PDP et moi j'ai dit salut PDP %add: PDP

Il arrive également que l'activité d'un locuteur dans le canal se réduise à en appeler un autre et qu'aucune réponse n'y soit apportée.

Extrait du Corpus F6 *PCE: P94 est ce que tu es la? %add: P94

On peut alors supposer soit que l'interlocuteur n'est pas présent (le départ du participant ne tarde pas, ce qui est le cas de cet extrait), soit qu'une conversation privée est engagée.

Certains participants, à leur arrivée, déclenchent des séries de salutations. Des indices, tels que des allusions à histoires conversationnelles, permettent de comprendre que ces locuteurs se connaissent.

en vertu de la loi du droit d'auteur.

Extrait du Corpus P3 (PIW entre dans le canal) *PAB: rew %add: ALL *PZZ: PAB est back from .. heee .... is back %add: ALL (1) *PAG: rrrrrrrrrrre PAB %add: PAB *PAD: PAB re mon amour :) %add: PAB (1) *PAM: re PAB :) %add: PAB Extrait du Corpus F7 (PIW entre dans le canal) *PIW: allo tout le monde ; %add: ALL *PIW: allo PEI ;)))))))))))) %add: PEI *PAH: kikoo PIW :)))))) %add: PIW *PIW: bonjour PAX ;)))))))))))) xxxxxxxxxxxxxxxxxxxxx %add: PAX *PIW: kikooo PAH ;)))))))))))) %add: PAH *PEI: kikoo PIW xoxxoxoxoxo %add: PIW *PIW: PCK xooxoxooxoxooxoxooxoxooxoxooxoxooxoxo %add: PCK

Ces salutations adressées individuellement sont la règle entre ce type de locuteurs, ainsi que le laisse paraître l'extrait ci-dessous, dans lequel PDW s'excuse de n'avoir pas salué PDT.

Extrait du Corpus F8 *PAS: coucou PDT :))) xxxxx %add: PDT *PDT: allo PBL ;) %add: PBL *PDV: PDT aa mon aaaaaaaaaaaaaammmmmmmmmmmmmmmmmmmooooooooooo oooooooooooouuuuuuuuuuuuuuuuuuuuuuurrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr enfin qq qui va me sortir de cet ennui %add: PDT (2) *PDW: salut PDT je ne t avais pas vue desolee %add: PDT *PDT: aaallo PDV ;))))))))))] mon chiriiiiiiiiiiiiiiiiiiiiiiiiiii %add: PDV *PDT: allo PDW ;)))))))))) %add: PDW

On note également que l'inventaire des salutations est plus grand et que les termes d'adresse sont plus libres. Enfin, les salutations complémentaires sont plus fréquentes entre ce type de participants.

Les salutations complémentaires selon André-Larochebouvy (1984:69) servent "d'amorce à un développement du thème qui servira alors d'exorde avant d'aborder d'autres sujets moins anodins". Dans nos corpus, ces échanges concernent des questions de santé ou de météo.

La littérature [219] fait mention de ce phénomène et affirme que pour des questions de santé, si la réponse est positive, elle permettra l'ouverture d'une autre thématique, tandis qu'une réponse négative entraînera un développement du thème. C'est également le cas dans nos corpus, l'extrait suivant en est une illustration.

Extrait du Corpus F1 *PBW: allo PAY ;)))))))))))))))))))) %add: PAY (2) *PAY: PBW xoxoxoxoxo :)))) %add: PBW (1) *PBW: tu vas bien belle PAY ;)))))))))))))) %add: PAY *PAY: PBW pas mal et toi?? :) %add: PBW *PBW: je suis découragée PAY ;)..... j'en ai marre de la neige ;(%add: PAY (1) *PAY: PBW je comprends mo j'en ai marre de la neige %add: PBW (1) *PAY: mais dans 15 jours on part au soleil :) %add: PBW *PBW: veinarde!!!!!!! tu devais venir ici ... tu aurais vu de la neige à volonté :) %add: PAY (…)

Les participants privilégiés ont également un comportement particulier à l'intérieur des sessions, ils s'absentent et reviennent fréquemment, en le signalant à la communauté profitant le plus souvent de charnières conversationnelles. Les moyens pour ce faire sont assez simples, pour le départ, ils formulent un simple terme : *brb* (*be right back)* accompagné éventuellement d'une justification et à leur retour ils utilisent *re*. Dans le cas où le laps de temps écoulé a été long ou a permis l'arrivée et le départ de participants privilégiés, des salutations sont produites. Dans le cas contraire, en général, aucun

[219] Traverso (1996)

en vertu de la loi du droit d'auteur.

accusé réception n'est formulé. Dans l'extrait ci-après, les circonstances sont particulières, puisque deux participants s'absentent en même temps pour se parler en privé.

Extrait du Corpus F4 (PCC et PAT ont convenu de se joindre par téléphone) *PCC: brb %add: ALL *PAT: LOL %add: PCC *PAT: brb tel %add: ALL (5) *PCC: ret %add: ALL *PAT: re %add: ALL *PCA: re les filles :) %add: ALL *PCC: lol %add: PCA, PAT *PAT: lol %add: PCA, PCC *PAT: ca paraît pas du tout qu'on était au tel :P %add: PCC

PCA formule ainsi une salutation accompagnée d'une frimousse à leur retour marquant le caractère inhabituel de la situation.

Une dernière remarque, dans le domaine de l'entrée en matière sur les IRC et confortant toujours l'hypothèse d'une sorte de frontière entre deux types principaux de participants (privilégiés *vs* occasionnels), concerne les informations individuelles.

Les participants occasionnels formulent des questions sur l'identité de leurs allocutaires et fournissent des informations sur la leur.

Ce sont les mêmes qui également formulent des "petites annonces" pour trouver l'interlocuteur idéal.

Extrait du Corpus F1 *PAE: BONJOUR TOUT LE MONDE !!!! %add: ALL (2) *PAE: Je suis un gentil monsieur de paris %add: ALL (1) *PAE: qui veut parler avec moi? %add: ALL

En ce qui concerne les séquences de clôture, elles sont essentiellement formulées par les participants privilégiés. On trouve des échanges pré-clôturants. Ces derniers concernent essentiellement les raisons du départ du participant, quelques vœux et quelquefois des promesses de revenir.

On trouve donc dans les IRC une relation entre les salutations et l'axe de la relation verticale, c'est-à-dire en fonction de la distance plus ou moins grande qui existerait entre deux individus.

6.3.2. Unités conversationnelles thématiques

Nous avons considéré comme unités conversationnelles thématiques les suites d'échanges cohérentes sur la thématique. Il est en effet difficile de retenir le critère des participants. Aucun signe ne manifeste la participation silencieuse des connectés. L'engagement dans la conversation ne se relève que par les interventions des différents locuteurs, on ne peut donc dire si un participant suit la conversation du début à la fin à moins qu'il n'ait réagi à toutes les interventions.

Dans nos corpus, malgré le nombre important de participants, des unités conversationnelles peuvent se dérouler de bout en bout à deux locuteurs. De telles unités peuvent aussi bien se constituer d'elles-mêmes qu'en parallèle à une unité en cours qui l'a suscitée, ou à sa suite. Toutefois, un très grand nombre d'unités se jouent à plus de deux locuteurs. Le cas où deux ou plusieurs locuteurs se liguent contre un autre est assez rare. Le plus souvent, la conversation est coopérative.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

La longueur des unités est très variable, et elle est en général fonction du nombre de participants.

Dans certaines unités conversationnelles, seule la thématique assure l'unité, et ne présente pas de véritable échange. Ces unités restent courtes. C'est le cas par exemple des unités menées par des locuteurs qui considèrent le canal principal comme un passage vers des canaux privés, et n'interviennent donc que pour quelques messages. Notons également que les locuteurs privilégiés qui semblent ne pas entretenir de relation particulière avec d'autres locuteurs ne s'engagent jamais dans plus d'une unité conversationnelle simultanément.

La thématique d'une unité est rarement remise en cause comme c'est le cas dans l'extrait ci-dessous.

Extrait du Corpus F1 *PBU: PAE c ce we la coupe de francre contre auxerre? %suj: COUPEFRANCE %add: PAE *PAE: demain on se mange de l'auxerrois %suj: COUPEFRANCE %add: PBU (1) *PAE: yep %suj: COUPEFRANCE %add: PBU (2) *PCA: bahhhhhhhhhhhhhhh!!!!!!!!!!!!!!!!!! du foot %suj: COUPEFRANCE-RFUT %add: PBU, PAE *PCA: :(%suj: COUPEFRANCE-RFUT %add: PBU, PAE *PAE: PBU en plus on recoit on peut pas perdre %suj: COUPEFRANCE %add: PBU *(action): PBK aime pas le foot... %suj: COUPEFRANCE-RFUT %add: ALL *PCO: il parait que les joueurs ont eu moins d egueulle de gueule apres le match contre l'Espagne %suj: COUPEFRANCE %add: PBU, PAE (1) *(action): PCA est comme PBK %suj: COUPEFRANCE-RFUT %add: ALL *(action): PBK aime pas les chroniques de la lune noire non plus... %suj: RFUT-JMJMPAS %add: ALL *PBU: le match est televise sur TF1? %suj: COUPEFRANCE %add: PAE, ALL *PAE: ouaip %suj: COUPEFRANCE %add: PBU *PAE: a 17h %suj: COUPEFRANCE %add: PBU *PAE: je crois %suj: COUPEFRANCE %add: PBU

Dans cet extrait, une thématique est l'occasion du départ d'une nouvelle unité, dans laquelle les locuteurs se disent ce qu'ils aiment et ce qu'ils n'aiment pas mais en aucun cas elle n'interdit la poursuite de l'unité dont la thématique est réprouvée.

Les thématiques sont assez diverses et libres. Un seul sujet ligue l'ensemble des participants : lorsqu'un utilisateur occasionnel les questionne au sujet de leur pratique IRC.

Les unités thématiques sont à l'origine de charnières conversationnelles : elles peuvent se diviser en se répartissant les participants, ou générer des dissidences (ainsi qu'on peut le voir dans l'extrait précédent). Nous avons marqué ces liens entre thématiques dans les corpus.

Inversement, des unités thématiques menées en parallèles avec des inventaires de participants différents peuvent se contaminer *i. e.* des lexèmes d'une unité se retrouvent dans une autre.

6.3. Relation interpersonnelle

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Il s'agit là d'observer ce que construit l'échange entre les participants. Signalons tout d'abord que la lecture des sessions révèle une atmosphère détendue. Les manifestations de sympathie sont bien plus nombreuses que les manifestations d'antipathie. En effet, les locuteurs qui d'emblée se présentent avec des propos antipathiques ne reçoivent en général pas de réponse, et sont éjectés du canal. Il arrive que des locuteurs 'malveillants' parviennent à monopoliser la parole, par exemple, dans la session P6, un locuteur répète 17 fois le même énoncé insultant, avant de se faire éjecter du canal. Toutefois, selon la relation que le locuteur entretient avec les autres participants le conflit peut éclater ou non. En d'autres termes, si le locuteur 'agresseur' fait partie de la sphère IRC (s'il a une histoire conversationnelle avec des participants réguliers), les opérateurs auront moins de vélocité à le réprimander, et le champ est ouvert au conflit.

Dans nos corpus, le tutoiement est systématique. Les différences dans la relation horizontale qui se manifestent se relèvent essentiellement sur les appellatifs. En effet, on constate que les participants qui se connaissent modifient les surnoms de leurs allocutaires. Ces modifications prennent plusieurs formes : troncation des syllabes finales, ou troncation des premières syllabes ; changement de voyelle finale (en général pour un i ou un y) ; ajout de suffixes (*-ette, -ou,-ounet(te), -inet(tte), -ouze*) ; réduction à la première syllabe, redoublée ; formation d'un mot valise (par exemple X*chou*). On note aussi que seuls les surnoms des participants réguliers sont affectés par le phénomène de redoublement de caractère exposé au chapitre 5. Des combinaisons de ces procédés se produisent mais sont limitées à la troncation en finale associée à l'ajout d'un suffixe ou au changement de voyelle finale. Notons que les participants privilégiés s'appellent par leur surnom, et pas forcément le surnom courant, mais aussi le surnom utilisé dans une session antérieure, ce qui peut rendre la conversation difficile à comprendre par des tiers.

Les termes d'adresse répondent à des règles, ainsi, il semble que saluer personnellement un utilisateur habitué doive être soumis à la condition de connaître l'utilisateur en question, comme le suggère l'extrait présenté ci-après.

Extrait du Corpus F7 *PAY: salut PJQ!!! :) %add: PJQ (2) *PJQ: PAY on se connait? %add: PAY (1) *PAY: oui %add: PJQ (2) *PHG: PJQ... lui te connait ça s'est sur!!! encore une victimle d'un masskick ;)) %add: PJQ (2) *PJQ: une victime heureuse alors %add: PHG (3) *PHG: ta vision du bonheur est quelque peu aliénée alors PJQ... s'en prendre plein la tronche, du bonheur..... heummm %add: PJQ

De même, les noms d'adresse appellent parfois à négociation explicite, ainsi que dans les extraits ci-dessous :

Extrait du Corpus P2 *PAC: yo PAO? wanikng too much mate %add: PAO (5) *PAO: Mate? %add: PAC *PAO: Ou t'as vu que j'étais ton pote toi? %add: PAC *PAB: mdr %add: PAO (1) *PAB: yo PAO gimme five ! %add: PAO (1) *PAP: PAO aller fait lui payer :P arache lui les cheveu :P %add: PAO

Ce premier extrait se tient entre un locuteur privilégié et un locuteur occasionnel, tandis que le suivant se tient entre deux locuteurs privilégiés.

Extrait du Corpus P2 *PAP: ah ma torbinouche de tannante toi %add: PAK *PAP: ilou ma tite PAK :P %add: PAK *PAP: kin xox avec ca lol %add: PAK (2) *PAK: PAP c't'a moi que tu t'adresses de cette maniere? %suj: ACID %add: PAP (1)

en vertu de la loi du droit d'auteur.

****PAK: 'jour PAP xxxx %add: PAP *PAP: lol PAK bah oui lol %add: PAK *PAP: qui d'autre :P< %add: PAK (PAK se renomme torbinouche de tannante)***

Le premier extrait se poursuit par une série d'échanges houleuse, tandis que le second se termine par des échanges métalinguistiques à propos de la signification des termes employés.

Entre les participants privilégiés, les conflits qui naissent concernent le statut lié au dispositif : avoir ou non le statut d'opérateur et le donner ou non est ce qui pose problème.

Il ne se relève pas véritablement de traces linguistiques de la relation verticale contrairement à ce que pourrait suggérer le fait que les utilisateurs ont des statuts différents.

Les opérateurs sont régulièrement appelés par des participants privilégiés pour ramener l'ordre dans le canal. Ainsi, un certain nombre d'interventions sont des dénonciations de participants ayant des pratiques contraires à la netiquette [220] . Comme beaucoup d'aspects des IRC, ce phénomène est matière à ludicité.

Extrait du Corpus P3 *PAM: [11:51] – P95 - DCC Send LIFE_STAGES.TXT www [221] %add: ALL *PAJ: pas bo ca P95 %add: P95 *PAJ: huhuhuh %add: P95 (PAS éjecte P95) *PAM: PAS mon hero.. lol %add: PAS *PAC: ta été plus rapide ke moi PAS %add: PAS *PAS: ;) %add: PAC *PAC: :))) %add: PAS *PAJ: j'vousémis surLaPiste %add: PAS, PAC, PAM *(action): PAJ DCC GET TotalVirusAnhilation Free :/ %add: ALL *PAM: oouuffff i want it! %add: PAJ *PAM: lol %add: PAJ *PAC: he PAJ tu vas dénoncé tout le monde comme ça? %add: PAJ *PAC: :) %add: PAJ *PAS: ouhais c une balance! %add: PAC (PAJ) *(action): PAJ anti-virus fournit par LifeStage Co Inc %add: ALL *PAS: *** PAJ was kicked by PAC (Balance ! :-)) %add: PAC, PAJ,

Être opérateur c'est avoir le pouvoir de choisir ceux qu'on souhaite exclure. Une forme de respect se manifeste donc à leur égard. Toutefois, ces derniers n'apprécient pas toujours d'être traités de façon particulière. Dans l'extrait ci-après, PCA et PDR sont opérateurs.

Extrait du Corpus F4 *PEI: salut les ops %add: OP *PCA: hey toi on peut jeter que ckya sur le chan :) banane %add: ?PEI (1) *PDR: salut PEI... mais tu devrais aussi saluer les autres.... ils ne font pas tous de la représentation %add: PEI

Les opérateurs en principe réagissent assez souplement, nous l'avons dit. Quelquefois, ils usent de leur pouvoir sans sommation, ainsi qu'on le voit dans l'extrait ci-dessous, mais ces cas restent très marginaux.

Extrait du Corpus F7 *PJP: asv stp %add: ? *** PJP was kicked by PJQ (5 heures de sommeil en 2 nuits... ça fout de sale poil¤) *PJQ: oh le vilain power abuse... %add: ALL *PEI: lol %add: PJQ

Les conflits que nous signalions plus haut sont d'un autre type encore. Parmi les opérateurs, certains le sont automatiquement à leur arrivée dans le canal, d'autres doivent compter sur les autres opérateurs pour leur donner ce statut. C'est essentiellement là que naissent les problèmes, d'autant que tous les opérateurs ne se connaissent pas.

220 Essentiellement des invitations privées à rejoindre d'autres canaux, ou de la publicité pour des sites commerciaux.

221 ***www remplace une adresse IP***

en vertu de la loi du droit d'auteur.

Mutuellement, les utilisateurs ayant le statut d'opérateur ne s'excluent pas les uns les autres. Les utilisateurs privilégiés ne sont en général pas exclus non plus. L'exclusion d'un opérateur peut être une erreur de la part d'un autre ainsi que dans l'extrait ci-dessous.

Extrait du Corpus F7 *** PJQ was kicked by PFN--- (- text flood (scrolling) -¤) *PFN: ah merde %add: PJQ *PAK: mdrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr %add: PFN *PAK: lolololololol %add: PFN *PDS: mouahahah %add: PFN *** PJQ has joined #FCanal (2) *PFN: a koi y sert ce putain de level 75 [222] ? %add: ALL *PJQ: lol PFN %add: PFN *PAK: comme quoi meme dieu n'est a l'abris de rien lololol %add: PFN

Dans ces cas, ce qui pourrait être vécu comme une agression est résolu par un échange de sigles ou de frimousses tenant lieu de rires et sourires.

Il arrive aussi que les opérateurs jouent avec la possibilité d'éjecter un utilisateur, nous en avons déjà présenté quelques exemples.

Extrait du Corpus P9 *** PAA was kicked by PBG (tu flood bcoup mon cher , j'ai mal aux yeux :))¤) (5) *PAA: le kick poli %add: PBG *PAA: :) %add: PBG (1) *PBG: toujours oui :) %add: PAA

Ces phénomènes remarquables sont essentiellement liés à la relation verticale, mais ils ne concernent pas une grande part de l'activité linguistique.

6.4. Conclusion

Ce chapitre a tenté d'exposer quelques propriétés interactionnelles de nos corpus. Il a notamment fait ressortir les dimensions du dispositif qui questionnent le modèle hiérarchique, concernant principalement l'alternance des tours de parole, et le découpage des corpus en unités thématiquement stables.

Nous avons montré qu'il existe différentes catégories de locuteurs dans les IRC, et que ces dernières ne sont pas seulement prédéfinies par le dispositif. Les comportements interactifs nous ont permis de révéler ces différentes catégories, toutefois, nous devons signaler qu'un participant peut passer d'une catégorie à une autre, et que les processus qui le permettent doivent pouvoir être identifiés par une étude plus large. Les membres privilégiés de la communauté d'un canal forment un groupe solidaire qui se définit par des comportements particuliers. Parmi ceux-ci, on a noté les particularités de la gestion des tours de parole, de la mise en place des séquences d'ouverture et de clôture, ainsi que les thématiques conversationnelles. Il faut ajouter à cela un comportement interactif ludique, et l'utilisation des termes relatifs à l'univers de référence des IRC et des procédés de codages que nous avons mentionnés au chapitre 5. Enfin, les locuteurs privilégiés jouent avec de nombreux aspects du dispositif.

[222] ***Le level détermine le nombre et l'importance des commandes que peut exécuter un utilisateur. Level 75 est le niveau de base des opérateurs.***

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Chapitre 7 Matériel lexical

Nous avons vu au chapitre 5 l'écart à la norme que présentent nos corpus afin que le lecteur puisse aborder le chapitre 6, qui tente de mettre en lumière les caractéristiques de nos corpus au plan de l'interaction. Ce chapitre se donne pour objectif l'étude globale des moyens lexicaux mis en œuvre dans les IRC.

Ainsi que le relèvent de nombreux linguistes, le travail sur corpus et plus particulièrement sur des corpus autres que des corpus écrits normés questionne la théorie linguistique à plus d'un titre. Willems (1998:79-87) fait remarquer que la prise en compte des données orales met en question les concepts grammaticaux de base notamment de phrase et de subordination, et les concepts théoriques de norme et de compétence. Au sujet des données d'observation et d'introspection rencontrées en analyse linguistique Willems (1998:81) note également que "leur interrelation n'est par ailleurs pas évidente et [qu'] on pourrait argumenter que les opérations cognitives en jeu dans les jugements d'intuition sont d'un tout autre ordre que celles mises en œuvre lors de la production langagière."

Après un bref questionnement théorique, ce chapitre présentera les principes de la lemmatisation qui ont été mis en œuvre sur le corpus, et les différents choix qu'il a fallu faire. Nous examinons ensuite les propriétés générales du matériel lexical, en utilisant les outils de mesures communément employés (densité et diversité lexicale). Nous considérons ensuite les spécificités du lexique et de quelques phénomènes morphologiques.

en vertu de la loi du droit d'auteur.

7.1. Traitement des corpus.

L'étude des moyens lexicaux mis en œuvre dans un corpus nécessite qu'on soit capable de segmenter le corpus en unités lexicales. Il faut donc procéder à la reconnaissance de ces unités afin d'être en mesure de les classer et de formuler la spécificité de l'inventaire. Pour ce faire, nous avons balisé les unités de notre corpus. Ce travail fastidieux requiert d'effectuer un certain nombre de choix, que nous présentons ci-après.

7.1.1. Questionnement théorique

La théorie linguistique reconnaît parmi les traits structuraux essentiels du langage le fait qu'il possède deux sous-systèmes : grammatical et lexical. On a coutume de distinguer les catégories grammaticales des catégories lexicales en fonction d'un critère de classe ouverte/fermée. Ainsi, se trouveraient dans le lexique, d'une part une catégorie d'unités dites pleines et d'autre part une catégorie d'unités outils. Les disciplines linguistiques se partagent l'étude de ces unités. L'étude des mots outils relève de la morphosyntaxe, tandis que celle des mots pleins relève de la lexicologie [223] . Un des critères retenus pour la distinction entre ces classes est la référence. Les mots pleins (lexicaux) réfèrent (réalité extralinguistique) tandis que les mots outils ou grammaticaux ne réfèrent pas. L'argument traditionnel soutient que les mots qui sont dits ne pas référer ont une autre façon de faire sens et non pas qu'ils n'ont pas de sens [224] . Mais on ne voit pas alors très bien comment ces mots ne réfèrent pas, ils font en tout cas partie des moyens de la référence.

Le critère de la fermeture de la classe est aussi mis en question d'une part en diachronie, et d'autre part en synchronie quand on examine de plus près certaines catégories dites grammaticales telles que les locutions prépositionnelles.

Une catégorie grammaticale se définirait donc par le degré de stabilité des membres de sa classe. Les éléments qui appartiennent ainsi à la grammaire sont des formes liées telles que les flexions verbales, les morphèmes de transcatégorisation ; des formes libres, les catégories grammaticales et les relations grammaticales, l'ordre des mots, certaines formes suprasegmentales et les combinaisons régulières de ces éléments en tournures grammaticales ou structures syntaxiques.

Même si les frontières entre ces deux systèmes, lexical et grammatical, ne sont pas toujours clairement appréhensibles, il semble que ce trait soit partagé par toutes les langues et hypothèse est faite que chacun de ces sous-systèmes présente des fonctions sémantiques particulières. Talmy (2000a:21) propose que lorsqu'un énoncé est utilisé

[223] Béraud *et al.* (1988)

[224] Les études consacrées au sens en grammaire, outre Sapir (1953), Whorf (1970), Boas (1965), ou encore Jakobson (1963) sont de trois types principaux : typologie des solutions grammaticales utilisées pour exprimer une même fonction sémantique, sens et fonction de toutes les unités grammaticales d'une langue, analyse approfondie d'un élément particulier d'une langue.

en vertu de la loi du droit d'auteur.

pour susciter chez l'interlocuteur une "représentation cognitive" les éléments grammaticaux spécifient la structure, la configuration de la représentation cognitive, tandis que les éléments lexicaux construiraient le contenu de la représentation cognitive. "Les spécifications grammaticales fournissent donc à la phrase un cadre conceptuel ou, pour le dire de manière imagée, une ossature, un échafaudage sur lequel s'appuie le matériau conceptuel spécifié par le lexique [225] ". On trouve chez Langacker (1987), une position similaire, bien que ce dernier envisage de façon plus explicite lexique et grammaire comme un continuum : la grammaire se réduit à la structuration et à la symbolisation du contenu conceptuel. Le fonctionnement cognitif est ici assimilé au fonctionnement linguistique. Mais que dire, si les éléments grammaticaux permettent de constituer la structure de la représentation cognitive, du fait que des éléments soient grammaticalisés dans une langue et pas dans une autre ?

Les concepts véhiculés par des éléments grammaticaux à travers les langues sont en nombre restreint. L'hypothèse est que ces notions constituent le système conceptuel fondamental qui structure le langage. Selon Talmy (2000a:21), "cet ensemble de concepts spécifiés par la grammaire, que l'on retrouve dans l'ensemble des langues, fournit à l'organisation conceptuelle son cadre schématique de base, au sein du domaine cognitif propre au langage" [226] .

La notion de catégorie dans la pensée occidentale est fortement déterminée par les écrits d'Aristote. Pour ce dernier, deux types de mots reçoivent un traitement privilégié relativement aux fonctions qu'ils sont susceptibles d'exercer : les fonctions de sujet et de prédicat. Ces types de mots sont les noms et les verbes. Aristote les envisage comme répondant à des catégories, qui sont au nombre de dix : substance, quantité, qualité, relation, la possession, le temps, le lieu, la situation, l'action et la passion. La notion de catégorie, chez Aristote correspond à la manière de prédiquer, c'est-à-dire à la manière d'attribuer une qualité à un sujet.

Une catégorie est en fait un principe de regroupement d'objets en classes selon les propriétés qu'ils partagent. L'appartenance à une catégorie se définit donc par la reconnaissance ou non d'un jeu de propriétés nécessaires et suffisantes. Ce modèle est remis en cause par la théorie du prototype. Les principes de la catégorisation humaine ne reposeraient pas sur le modèle des conditions nécessaires et suffisantes.

La sémantique du prototype, version standard, correspond aux propositions formulées par les psychologues dans les années soixante-dix. La catégorisation, le découpage de la réalité, était conçu comme essentiellement arbitraire. On considère à présent que la catégorisation est issue de l'interaction entre les stimuli et leur traitement, et que cette interaction est dépendante du processeur, qui utilise capacités et stratégies variées pour donner du sens. Rosch (1976), qui est à l'origine de la théorie du prototype en psychologie, estime que la catégorisation, qui découle de l'interaction entre l'homme

[225] "The grammatical specifications in a sentence, thus, provide a conceptual framework or, imagistically, a skeletal structure or scaffolding for the conceptual material that is lexically specified."

[226] "This crosslinguistically select set of grammatically specified concepts provides the basic schematic framework for conceptual organization within the cognitive system of language".

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

processeur et son environnement physique est indépendante des classifications linguistiques. Il y aurait donc une distinction entre d'une part, le sens donné au monde et issu de lui et d'autre part le sens linguistique. La recherche d'invariants à statut d'universaux semble contradictoire dans les travaux de Rosch (1978:28) [227] lorsqu'elle elle précise au sujet de son travail que, "il doit être noté que les thèmes relatifs à la catégorisation que nous traitons prioritairement concernent les principes explicatifs des catégories qui se trouvent dans une culture et codés par la langue de cette culture à un moment donné".

La linguistique s'est réapproprié cette analyse, afin de remédier à une partie des faiblesses inhérentes à l'analyse sémique basée sur le modèle des conditions nécessaires et suffisantes d'Aristote. Le sens se voit attribuer une conception moins rigide, et donc plus adéquate puisque le pouvoir référentiel des mots s'en trouverait mieux expliqué. La notion de prototype se définit comme un ensemble de propriétés typiques. Cet ensemble formerait un modèle mental, auquel le locuteur comparerait de façon globale une représentation (l'image qu'il se fait du référent), et adopte ou non la dénomination correspondante selon qu'elle lui paraît voisine ou non. Les catégories deviennent plus souples et nous font passer d'un monde "manichéen" à un monde plus nuancé.

En ce qui concerne la théorie linguistique même, les principes de classement des unités linguistiques en catégories sont soumis au questionnement. Catégories grammaticales, catégories morphologiques, catégories lexicales sont des termes si souvent utilisés qu'on en vient à se demander quelles réalités ils recouvrent et quelles définitions il convient de leur accorder. Parmi les premières tentatives de catégorisation du matériel linguistique se trouve l'analyse en parties du discours (traduction littérale de *partes orationis*), initiée par les grammairiens latins et grecs et semble-t-il inspirée des catégories d'Aristote, où la substance correspondrait au substantif, la qualité à l'adjectif, la quantité au nombre, la relation aux formes comparatives et relatives, temps et lieu aux adverbes, agir et pâtir aux voix actives et passives… Il s'agit de procéder à une classification des mots éléments de la langue en types. La liste établie par Aelius Donatus [228] , grammairien latin du IV^{e} siècle constitue la base à laquelle se sont référés les travaux ultérieurs. Cette liste n'a subi que de légères modifications jusqu'au XX^{e} siècle et contient 8 ou 9 classes : nom, pronom, verbe, adverbe, adjectif (notion inconnue de l'Antiquité classique), participe, conjonction, préposition, interjection, article.

Les critiques que l'on peut formuler à l'encontre de l'analyse en parties du discours sont nombreuses, la première étant sans doute qu'elle vise à assigner à chaque mot une et une seule place dans la classification, ce qui risque de masquer des comportements différents. Une des faiblesses majeures les plus soulignées est celle de l'hétérogénéité des critères de classification [229] : critère morphologique (nom, verbe), position des mots

[227] Citation traduite par Dubois (1991:32)

[228] Aelius Donatus *De Partibus Orationis Ars Minor* reproduit sur Orbis Latinus © Zdravko Batzarov http://www.orbilat.com/Latin/Texts/05_Late_period/Treatises/Donatus_Aelius-Ars_minor.html

[229] Trask (1999:280-282).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

les uns par rapport aux autres dans la phrase (préposition), fonction syntaxique (conjonction), critères sémantiques (adjectif), et le manque de distinction entre statut lexical et grammatical. Enfin, l'universalité des parties du discours est fortement remise en cause par nombre de linguistes dont Schachter (1985:3) qui affirme que "Tandis que toutes les langues font des distinctions en parties du discours, il y a d'assez importantes différences entre les langues concernant à la fois le type et le nombre des distinctions qu'elles font" [230] . Ainsi, l'inventaire des types fondamentaux d'unités, pour écarter le problème de la question de l'universalité et celui de l'hétérogénéité des critères est souvent diminué de moitié et ne comporte donc plus que les "classes lexicales majeures" : nom, verbe, adjectif, adverbe (Givòn 1984:47-85).

Il est possible, selon Creissels (1995:16), de "rejeter l'idée même d'aborder la question de la définition des notions fondamentales de la syntaxe au niveau du classement syntaxique des unités lexicales ou des "mots"", en acceptant l'idée que toutes les notions syntaxiques sont dérivées de celles de constituant nominal et unité phrastique. Creissels exemplifie cette position pour le verbe qui, en français par exemple, peut être distingué des autres catégories sur la base du système d'affixes qui l'affecte, mais n'est identifiable comme type à travers les langues du monde que par le rôle qu'il tient dans l'unité phrastique. On trouve une position similaire chez Feuillet (1988:72) qui affirme qu' "En partant du mot ou monème, on s'interdit de voir qu'une même relation peut être exprimée par des moyens différents : en disant par exemple que le cas est une catégorie (ou une modalité) nominale, on occulte le fait que la position du groupe ou la préposition peuvent jouer le même rôle. Or, si l'on fait de la préposition une partie du discours, du cas une catégorie et de la position un procédé syntaxique, on obscurcit le lien qui relie l'expression d'une relation identique."

Nous avons vu au chapitre 4 les principes de lemmatisation dans une perspective générale. Il convient maintenant de préciser les choix que nous avons opérés. Si le classement des unités en parties du discours pose problème, il reste une méthode très utilisée en linguistique et la comparaison n’est plus possible si l'on ne se conforme pas aux grands principes d'une discipline. On arguera certes que le renouveau et le progrès sont alors entravés par des positions conformistes.

7.1.2. Dilemmes

7.1.2.1. Classes lexicales majeures

Les unités qui posent le moins de problèmes au niveau de l'identification des classes sont les noms et les verbes. Nous les avons classés selon les méthodes traditionnelles en faisant la distinction entre nom commun et nom propre pour les premiers, et entre verbe auxiliaire et verbe lexical pour les seconds. La difficulté commence à se prononcer lorsque l'on considère les outils de détermination nominale et les adjectifs. Les grammaires proposent des classements forts différents les uns des autres. Nous avons

[230] "While all languages make parts-of-speech distinctions, there are rather striking differences between languages with regard to both the kind and the number of such distinctions that they make".

en vertu de la loi du droit d'auteur.

préféré le classement des quantificateurs ainsi que celui des "adjectifs possessifs" dans la catégorie des déterminants nominaux plutôt que de les ranger avec les adjectifs. Ce facteur réduit considérablement, par rapport à d'autres études, la proportion des adjectifs comptabilisés dans notre corpus puisque n'y subsistent que les adjectifs qualificatifs.

La classe des adverbes est très hétérogène, c'est un fait reconnu, et elle le reste dans nos corpus, puisque nous n'avons pas opéré de distinctions particulières au sein de celle-ci. Seuls les morphèmes de négation, traditionnellement étiquetés comme adverbes, ont connu un traitement particulier.

7.1.2.2. Prépositions et conjonctions

La définition de la classe des prépositions repose sur un principe syntaxique, ce sont des unités antéposées au nom dont la propriété caractéristique serait d'intervenir au niveau de la relation entre le prédicat verbal et ses arguments nominaux (un groupe prépositionnel apparaît comme un complément indirect du verbe). Dans le passage du latin au français, les prépositions ont remplacé les désinences casuelles des substantifs et de ses déterminants. Les prépositions ont pris en charge certains de leurs rôles.

Certains linguistes classent les prépositions en deux groupes : les prépositions clitiques et les prépositions non clitiques. La liste des prépositions dites clitiques est : *à, de, chez, par, pour.* Et celle des prépositions non clitiques est : *depuis, pour, après*… c'est-à-dire les prépositions qui reçoivent traditionnellement le double statut de préposition et adverbe.

(1) Pierre est parti après (ton départ) (2) Il est revenu depuis (ce temps là) (3) En ce qui concerne l'intégration du CDLS au SCD, les élus ont voté pour (l'intégration…)

En première analyse, avec Creissels (1995:185) on considèrera "comme prépositions les morphèmes qui s'antéposent au nom et dont l'occurrence est liée à l'insertion du nom dans une construction de rang supérieur quelle que soit la nature exacte de cette construction."

La préposition aurait alors deux statuts. Si elle est inscrite dans le schème de prédication, son rôle est de préciser le statut argumental d'un constituant nominal. Sinon, elle permet à un constituant nominal d'occuper une position autre que strictement nominale. Nous avons marqué les cas où la préposition *de* appartient au système de la négation (4) et ceux où elle entre en construction génitivale (5).

(4) Je ne bois pas de café (5) Pierre déchiré le dessin de la jeune fille

Préposition et conjonction font partie des parties du discours dites pauvres de sens et invariables. Elles ont dans certains cas des comportements similaires qui ont amené certains linguistes, comme Tesnière (1969) à postuler une catégorie plus large, comprenant conjonctions et prépositions. Cette catégorie est celle des opérateurs de translations. La fonction de la conjonction est d'unir des éléments. On divise en général cette classe en deux sous classes : coordination et subordination. Mais les critères que donne la grammaire traditionnelle pour les distinguer sont vite mis en défaut. Le premier critère examine la nature des éléments qui sont mis en relation. On observe cependant que les éléments donnés pour conjonction de coordination peuvent coordonner soit des

en vertu de la loi du droit d'auteur.

éléments qui ne sont pas des unités phrastiques, soit des éléments qui sont des unités phrastiques. De même, les conjonctions de subordination, qui sont censées relier seulement des éléments phrastiques relient aussi parfois des éléments non phrastiques comme *sa conduite est prudente quoique un peu rapide*.

On donne aussi comme critère le fait qu'une conjonction de coordination peut commuter avec un adverbe tandis que la conjonction de subordination ne le pourrait pas. Mais ce critère est sémantiquement trop vague pour qu'on puisse l'utiliser de façon fiable d'autant plus qu'on ne détermine pas dans la classe hétérogène des adverbes lesquels permettent le test. La zone qui pose problème est celle où un élément permet l'introduction d'une unité dans une unité de rang supérieur. Ainsi la conjonction de subordination joue devant la proposition qu'elle introduit le même rôle qu'une préposition devant un groupe nominal :

(6) avant que nous n'arrivions (7) avant notre arrivée

La distinction entre prépositions et conjonctions peut sembler claire au premier abord si on considère que les prépositions permettent l'insertion d'un nom dans une unité de rang supérieur, tandis que les conjonctions interviendraient au niveau de la relation entre deux unités phrastiques.

Si on observe les choses d'un peu plus près, certaines conjonctions ont des comportements syntaxiques caractéristiques des compléments prépositionnels du verbe, c'est le cas par exemple de :

(a) Extrait du Corpus F3 *(action): PAE fait une excuse comme johnny dep ds matrix

Cependant des cas tels que le suivant justifient qu'on rétablisse le caractère prédicatif de l'élément introduit par *comme* dans l'exemple précédent.

(b) Extrait du Corpus F3 *PCA: ca ce trouve, t'es vraiment comem je dis :)

qui amènerait à considérer qu'un mécanisme de réduction discursive est à l'œuvre et que la structure complète se donnerait comme suit :

PAE fait une excuse comme la fait johnny dep ds matrix

Nous avons cependant considéré les cas correspondant à (a) comme des prépositions et les cas (b) comme présentant des conjonctions.

D'autre part, on remarque que certains éléments considérés comme des prépositions ont dans certaines circonstances des comportements analogues à ceux d'une conjonction. C'est le cas par exemple de *de* suivi d'un infinitif.

7.1.2.3. Pronoms

La catégorie des pronoms est l'une des plus problématiques en grammaire traditionnelle. Si l'on s'en tient à l'étymologie du terme, elle rassemblerait les unités autres que des noms ayant la particularité de prendre la place d'un nom. Cette définition peut aussi suggérer que les unités en question fonctionnent comme anaphoriques c'est-à-dire représentant un constituant nominal présent dans le co-texte. Or un examen même superficiel des unités rangées dans cette catégorie par les grammaires traditionnelles montre que non seulement certaines ne peuvent pas prendre la place d'un nom (certains

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

pronoms personnels en particulier), mais aussi que des unités qui ne sont pas répertoriées comme pronom peuvent manifester le comportement anaphorique qui est à la base de la définition traditionnelle du pronom.

Nous avons pour cette catégorie utilisé les critères proposés par Creissels (1995:108-128) qui amènent en particulier à considérer d'une part les pronoms personnels conjoints comme des éléments affixés à la base verbale dénommés indices pronominaux, et d'autre part les pronoms personnels disjoints comme une catégorie de noms autodéterminés appelés noms déictiques. Les pronoms interrogatifs de la grammaire traditionnelle sont scindés en deux catégories. D'une part on considèrera ceux qui fonctionnent accompagnés d'un substantif comme des déterminants interrogatifs qui peuvent présenter des phénomènes de réduction discursive. D'autre part, les pronoms et adverbes interrogatifs de la grammaire traditionnelle seront rangés dans une même catégorie.

Enfin, nous avons conservé la catégorie traditionnelle des pronoms relatifs telle quelle dans notre corpus.

7.1.2.4. Interjections

Les interjections forment une catégorie qui reste en marge des préoccupations des linguistes, à l'exception de ceux qui adoptent des perspectives énonciatives [231] . Ainsi Givòn (1984:84) s'en débarrasse en quelques lignes : "la plupart des langues présente cette catégorie fourre-tout avec des expressions telles que *yes, no, hey, oh, uh, hi, wow, ouch*, etc. ou leur équivalent fonctionnel. Il ne s'agit pas d'une catégorie unifiée fonctionnellement, morphologiquement ou syntaxiquement, et elle est hautement spécifique à la langue." [232]

Nous avons rangé dans cette catégorie les représentations de rire, de cris, d'hésitation d'une part et d'autre part les formules rituelles de salutation, remerciement, excuse… Cette catégorie contient aussi ce que Luzzati (1991) nomme des appuis du discours. Elle reste donc un "fourre-tout" que nous avons toutefois organisé en attribuant des marques spécifiques aux trois dimensions envisagées.

Nous avons donc présenté ici les choix que nous avons effectués pour la lemmatisation. Ainsi que nous l'avons mentionné au chapitre 4, nous l'avons accompagnée d'un codage de la flexion. Cette opération a été réalisée par des procédures semi-automatiques avec contrôle, selon le système proposé pour CHILDES [233] et les spécifications de Parisse et Lenormand (2000) auxquelles nous avons apporté quelques aménagements. Cette étape de traitement des corpus permet d'obtenir des

[231] Culioli (1991), Roulet *et al.* (1991) entre autres.

[232] "Most languages display this mixed-bag category, with expressions such as 'yes', 'no', 'hey', 'oh', 'uh', 'hi', 'wow', 'ouch', etc. or their functionnal equivalent. It is not a unified category functionnally, morphologically or syntactically, and it is highly language specific."

[233] Cf. MacWhinney (1995).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

informations sur le lexique et la morphologie. Nous examinerons dans un premier temps les indices de diversité lexicale et de densité lexicale, puis nous présenterons les spécificités lexicales rencontrées dans les corpus.

7.2. Listes de fréquences

Après la lemmatisation d'un texte, chaque vocable comporte une fréquence c'est-à-dire le nombre de ses occurrences. Nombre d'analyses de fréquences ne portent pas sur des textes lemmatisés. Ainsi, les listes de fréquences issues d'un corpus de français parlé d'un million de mots (Corpaix, version mai 2000) et celles extraites du CD-ROM du Monde Diplomatique (1987-1997) [234] sont comparables entre elles dans la mesure où ce sont des listes de fréquences de formes non lemmatisées. En revanche, elles ne comportent pas des populations comparables : un million de mots pour le corpus oral et 11 139 376 unités pour le corpus écrit. Nous disons unités pour souligner une différence supplémentaire, qui entrave la comparaison, issue du fait que les signes de ponctuation du texte écrit ont été comptabilisés de la même manière que les unités lexicales. De plus, la question de la norme lexicale est cruciale. On voit dans le tableau ci-dessous que la norme lexicale n'est pas la même pour le corpus IRC que pour les deux autres : *c'est* est considéré comme une unité lexicale dans nos corpus, alors qu'il en compte deux dans les deux autres corpus.

Plus le texte considéré est grand, plus le nombre de vocables [235] est grand. Comparer des listes de fréquences pourrait s'envisager de deux manières : ne prendre en compte que le rang des unités revient à gommer la proportion relative des unités dans le texte et travailler sur des proportions oblige à considérer l'ensemble des listes ou à n'en conserver qu'une proportion égale.

Tableau (7-1) – Comparaison de fréquences Corpaix / Le monde Diplomatique

[234] Véronis, J. Université de Provence Aix-Marseille Page personnelle http://www.up.univ-mrs.fr/~veronis

[235] Nous employons la terminologie de Muller (1992) : "élément du vocabulaire d'un texte" qui dépend de la norme lexicologique adoptée.

en vertu de la loi du droit d'auteur.

	Corpaix	Monde Diplomatique	Corpus IRC
1	EST	,	de
2	DE	DE	je
3	C'	.	pas
4	EUH	LA	salut
5	IL	L'	c'est
6	ET	LES	le
7	QUE	ET	tu
8	LE	LE	ça
9	ON	DES	j'
10	LA	À	à
11	À	1	la
12	JE	D'	les
13	PAS	-	et
14	LES	EN	un
15	A	DU	que
16	UN	UN	en
17	ÇA	UNE	moi
18	DES	EST	est
19	EN	DANS	une
20	QUI	QUE	t'
21	L'	)	pour
22	Y	(	non
23	QU'	QUI	il
24	UNE	PAR	l'
25	MAIS	IL	mais
26	CE	POUR	ai
27	D'	A	vous
28	OUI	AU	oui
29	TU	PLUS	qui
30	J'	:	on
31	DANS	SUR	des
32	POUR	PAS	tout+le+monde
33	VOUS	SE	va
34	ILS	S'	toi
35	FAIT	NE	me
36	TOUT	QU'	suis
37	ÉTAIT	CE	bien
38	PLUS	N'	d'
39	LÀ	AUX	ce
40	ALORS	SONT	mon
41	ELLE	MAIS	si
42	DONC	AVEC	es
43	PARCE	ONT	avec
44	DU	OU	a

en vertu de la loi du droit d'auteur.

45	NON	ON	sur

Ce qui apparaît dans les listes de fréquences, c'est que les items les plus fréquents sont des items grammaticaux [236] . L'unité 'de' est fréquente à rang égal dans les deux corpus. Mais on se heurte très vite au défaut de lemmatisation [237] . *'est'* peut aussi bien représenter le point cardinal que le verbe et s'il est permis de supposer que les corpus présentent pour cette unité une égale répartition des différents lemmes attachés à *'est'*, rien ne permet de comparer *'que'* entre les deux listes sans informations sur les fonctions qu'il prend dans les textes. Et l'observation peut se faire dans l'autre sens, puisqu'une même unité, présentant des variantes contextuelles, ou des flexions différentes, se trouve scindée en deux unités dans les listes de fréquences.

Les listes de fréquences qui ne sont pas lemmatisées ne permettent guère que l'observation d'unités qui ne seraient ni polyvalentes ni invariables. L'examen des listes ci-dessus rappelle que la situation de monovalence est rare dans les langues et qu'on ne peut la reconnaître à la limite que pour quelques unités : 9 pour les deux corpus confondus, dont 3 pour le corpus le monde diplomatique et 6 pour le corpus Corpaix.

Les listes de fréquences de formes lemmatisées sont plus rares, par le simple fait qu'elles requièrent un investissement temporel plus important. Pour le français, une des listes de fréquences les plus utilisées pour les observations de l'oral est celle de Gougenheim (1964) [238] qui porte sur des formes lemmatisées et sous-catégorisées pour la classe lexicale et a servi à l'établissement du vocabulaire du français fondamental. Il existe de nombreuses autres listes de fréquences [239] de référence, réalisées principalement sur des corpus littéraires. Si nous appliquons les mêmes critères de lemmatisation que Gougenheim (1964), nous obtenons pour nos corpus la liste de fréquences suivante.

Tableau (7-2) – Fréquences selon les principes de lemmatisation de Gougenheim (1964)

[236] Et ceci est également valable pour nos corpus, ce qui écarte l'hypothèse d'un style télégraphique (ce dernier étant reconnu pour faire l'économie des mots grammaticaux).

[237] Cf. chapitre 4.

[238] L'initiative de l'établissement d'un dictionnaire et d'une grammaire du français fondamental revient à l'UNESCO (organisation des Nations Unies pour la science, la culture et l'éducation) qui créa en 1947, un comité linguistique dans le but de susciter " la diffusion des langues de communication comme l'un des moyens les plus efficaces de répandre largement l'éducation de base ". Le Centre de recherche et d'étude pour la diffusion du français (CREDIF) dont Georges Gougenheim, professeur à la Sorbonne fut le directeur, naquit à cette occasion en 1959. Malgré les nombreuses critiques qui ont pu être adressées à ces travaux, des outils de traitement automatique du français utilisent leurs résultats. C'est le cas, par exemple, de Cordial (http://www.synapse-fr.com/).

[239] INALF (1971) ; Juilland, Brodin, Davidovitch (1970) ; Engwall (1984)

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

numéro d'ordre	lemmes	fréquence
1	être	2001
2	je	1663
3	de	1260
4	pas (neg)	1006
5	salut	957
6	ce	952
7	tu	912
8	avoir	822
9	ça	646
10	le (art)	614
11	il(s)	605
12	à	546
13	la	527
14	et	500
15	les	500
16	un (det)	483
17	que (conj)	423
18	moi	399
19	une (det)	353
20	me	334
21	non	331
22	pour	327
23	mais	308
24	vous	295
25	oui	287
26	on	276
27	tout le monde	270
28	ne	252
29	dire	247
30	des	240
31	en (prep)	231
32	toi	215
33	se	188
34	qui	185
35	bien (adv)	184
36	mon	184
37	sur (prep)	171
38	alors	166
39	bon (interj)	163
40	avec	161

Cette liste a été obtenue à partir de nos principes de lemmatisation en regroupant certaines unités, de sorte qu'elle soit comparable à celle que donne Gougenheim (1964).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Tableau (7-3) – Fréquences du français fondamental

numéro d'ordre	lemme	répartition	Fréquence
1	être	163	14083
2	avoir	163	11552
3	de	163	10503
4	je	162	7905
5	il(s)	160	7515
6	ce (pro)	163	6846
7	la (art)	163	5374
8	pas (neg)	158	5308
9	à (prép)	163	5236
10	et	161	5082
11	le (art)	163	4957
12	on	128	4266
13	vous	154	4202
14	un (art)	162	4188
15	ça (pro dem)	159	3972
16	les (art)	162	3815
17	que (conj)	162	3537
18	ne	150	3283
19	faire	162	3174
20	qui (relatif)	160	3096
21	oui	154	2935
22	alors	155	2854
23	une (art)	163	2780
24	mais	159	2768
25	des (art indef)	161	2646
26	elle (s)	134	2462
27	en (prep)	161	2405
28	dire	160	2391
29	y	143	2391
30	pour	161	2076
31	dans	162	2066
32	me	152	2014
33	se	161	1993
34	aller	161	1876
35	bien (adv)	163	1697
36	du	154	1658
37	tu	105	1536
38	en (pronom adverbe)	153	1501
39	au	161	1490
40	là	153	1468

en vertu de la loi du droit d'auteur.

Trente et un items sont communs aux deux listes. Il s'agit exclusivement de mots grammaticaux à l'exception des vocables *être* et *avoir* pour lesquels il n'y a pas eu de distinction entre les occurrences des auxiliaires et les occurrences des verbes pleins, conformément aux décisions prises dans le travail de Gougenheim (1964). Les unités communes sont, par ordre alphabétique, les suivantes :

à (prép)*, alors, avoir, bien* (adv)*, ça, ce* (pro)*, de, des* (art indef)*, dire, en* (prep)*, et, être, il(s), je, la* (art)*, le* (art)*, les* (art)*, mais, me, ne, on, oui, pas* (neg)*, pour, que* (conj)*, qui* (relatif)*, se, tu, un* (art)*, une* (art)., *vous.*

Les items présents dans notre liste et absents de la liste pour le français fondamental sont *avec, bon* (interj)*, moi, mon, non, salut, sur* (prep)*, toi, tout le monde*.Pour les unités *salut* et *tout le monde*, leur fréquence provient de la fréquence des séquences d'ouverture de nos corpus et de leur grande uniformité. Les pronoms disjoints et déterminants possessifs, appartenant à la catégorie des déictiques, ainsi que *bon* ne sont pas surprenants dans ce type d'usage linguistique.

L'écart entre les ordres des items communs aux deux listes est parfois important. La moitié des items ne présente pas de gros écart, c'est le cas de *être, de, ce, un, une, se, pas, oui, les, le, je, et, en (prep), dire, bien.* Cependant l'appréciation de l'écart entre les ordres ne va pas sans prendre en compte les courbes de fréquences. On remarque en effet qu'il survient des chutes importantes de la fréquence entre les plus hauts rangs. Dans nos corpus, les 10 unités les plus fréquentes représentent 22,18 % de l'effectif total et les 20 unités les plus fréquentes représentent près du tiers de l'effectif total soit 31,74 %. Dans le corpus du français fondamental on obtient pour les mêmes observations respectivement 25,44 % et 37,77 %.

Il convient dans ce cas de préférer la comparaison des fréquences relatives. Nous nous reportons à une technique courante en statistique lexicale et dénommée normalisation par Biber *et al.* (1998:263) :

"La normalisation est une façon d'ajuster les fréquences de textes de différentes longueurs de sorte qu'elles puissent être comparées de façon précise. Le nombre total de mots dans chaque texte doit être pris en considération lorsqu'on norme les fréquences. Particulièrement, la fréquence brute doit être divisée part le nombre de mots dans le texte, puis multiplié par la base choisie pour la normalisation.". [240]

On note une fréquence relative plus importante pour les unités *être, avoir, de, il,* et *je* dans le corpus du français fondamental et une proportion plus importante pour les unités *dans elle(s) ne,* et *pour* dans notre corpus.

7.2.1. Diversité lexicale

La mesure de la diversité lexicale est souvent utilisée pour caractériser des types de

[240] "Normalization is a way to adjust raw frequency counts from texts of different lengths so that they can be compared accurately. The total number of words in each text must be taken into consideration when norming frequency counts. Specifically, the raw frequency count should be divided by the number of words in the text and then multiplied by whatever basis is chosen for norming".

en vertu de la loi du droit d'auteur.

textes et elle est particulièrement invoquée dans la recherche des différences linguistiques liées au mode de production oral/écrit. Ainsi M. Fayol (1997:15) souligne que "À l'écrit, la moindre contrainte temporelle et l'absence de destinataire présent autoriserait une sélection plus lente et plus réfléchie des mots, ainsi qu'une recherche d'explication afin de pallier le caractère décontextualisé de la production". Nos corpus se présentent bien évidemment comme n'ayant de l'écrit que la propriété matérielle. Nous nous attendons donc plus probablement à observer des mesures de diversité qui soient proches de celles obtenues à partir de corpus oraux. Cependant, lorsqu'on fait artificiellement produire des narrations orales à des locuteurs de laboratoire on observe des variations liées au mode de production. Les dimensions mises en avant par Fayol (*ibid*.) sont la contrainte temporelle et la présence/absence d'un destinataire contribuant au caractère décontextualisé de la production. La communication IRC présente un degré de décontextualisation bien moindre que l'écrit idéalisé.

La diversité, ou richesse lexicale selon la terminologie de Muller (1992:115-117), est définie par le nombre de vocables d'un texte, c'est-à-dire le nombre de mots différents. Pour un texte, la mesure est donc d'établir un rapport entre le nombre de vocables (*types*) et le nombre de mots (*tokens*). Cependant ce calcul est fortement influencé par la taille de échantillon considéré. Si l'étendue d'un texte dans l'analyse des fréquences est neutralisée par la normalisation telle que la présente Biber (1988:75), il n'en est rien pour les mesures de rapport entre type et occurrences, ce qui a été ignoré longtemps dans les études de psychologie. Afin de neutraliser l'influence de la longueur du texte, Biber (1988:224-225) propose de calculer la diversité lexicale en comptant le nombre d'items lexicaux différents qui apparaissent dans les 400 premiers mots d'un texte. Cette mesure fastidieuse dans son application, outre qu'elle ne présente pas de pertinence pour nos corpus à moins de l'appliquer sur les unités conversationnelles établies sur le thème (cf. chapitre 6), n'a pas été retenue pour cette étude. Nous lui avons préféré celle de Herdan (1966) [241] , qui utilise le rapport entre le logarithme du nombre de types et le logarithme du nombre de *tokens*.

Tableau (7-4) – Indices de diversité lexicale

[241] Cité dans Richards & Malvern, (1999) *Type-token and type-type measures of vocabulary diversity and lexical style : an annotated bibliography,* The University of Reading http://www.rdg.ac.uk/~ehsrichb/home1.html, révision du 16/06/1999 – Herdan, G, *The Advanced Theory of Language as Choice and Chance*, Springer Verlag, Berlin, 1966, pp. 75-77.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

corpus	types	tokens	ratio	herdan
F1	580	1630	0.356	0,860
F2	248	877	0.283	0,814
F3	1032	3506	0.294	0,850
F4	770	2378	0.324	0,855
F5	937	3008	0.312	0,854
F6	1068	3469	0.308	0,855
F7	2405	11010	0.218	0,837
F8	1436	6478	0.222	0,828
P1	178	312	0.571	0,902
P2	755	2216	0.341	0,860
P3	1080	3553	0.304	0,854
P4	486	1282	0.379	0,864
P5	342	789	0.433	0,875
P6	862	3154	0.273	0,839
P7	547	1455	0.376	0,866
P8	517	1312	0.394	0,870
P9	633	1558	0.406	0,877
	6568	47987	0,137	0,815

Cette mesure montre sur l'ensemble de nos corpus un indice de diversité lexicale assez fort puisque tendant vers 1 tous les mots sont différents. Cet indice conforte notre impression générale selon laquelle les thématiques de conversation à travers le corpus sont très diverses et peu redondantes. On pourrait s'attendre, ce dispositif de communication supprimant un certain nombre des indices que comporte une situation de communication synchrone habituelle, à un certain type de redondance, permettant par exemple d'expliciter à quelle intervention initiative se rattache une intervention réactive. Cela n'est que très peu présent dans nos corpus, ces indications se réalisant essentiellement par le biais de vocatifs.

7.2.2. Densité Lexicale

La densité lexicale est une mesure du rapport qui existe dans un texte entre mots grammaticaux et mots lexicaux. Cette notion suppose qu'on ait clairement déterminé à quel endroit se trouve la frontière entre mots lexicaux et mots grammaticaux. Le critère de l'ouverture fermeture de la classe permet de poser une ligne de démarcation bien que celle-ci soit souvent qualifiée de floue, voire impossible à placer ainsi que le stipule Langacker (1999:18) "les éléments traditionnellement assignés à la grammaire tendent à être assez schématiques (sémantiquement et/ou phonologiquement) tandis que ceux assignés au lexique tendent vers une plus grande spécificité. La différence est clairement une question de degré, et toute ligne de démarcation serait arbitraire" [242] . Tout en

[242] "the elements traditionnally ascribed to grammar tend to be quite schematic (semantically and/or phonologically), whereas those assigned to the lexicon tend toward greater specificity. Yet the difference is clearly one of degree, and any particular line of demarcation would be arbitrary".

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

admettant cette restriction, on peut reconnaître que la part des items intermédiaires reste faible et considérer comme classes lexicales avec Givòn (1984:47-84) les noms, les verbes lexicaux, les adjectifs et une partie de la classe des adverbes [243] .

La part des mots grammaticaux dans un texte est toujours importante, ne retenir que les 20 items les plus fréquents aboutit en général à une liste d'items grammaticaux. Sur l'ensemble d'un texte, on recueille environ 20% de mots grammaticaux [244] . Contrairement à ce qu'on trouve dans les textes écrits littéraires ou plus caricaturalement dans le style télégraphique, nous relevons dans nos corpus un taux de 53,34 % d'occurrences de mots grammaticaux. De plus, dans les 46,66% de mots lexicaux restants on relève une forte proportion de noms propres, soit 21,5 %. Cette dernière est liée à ce mode de communication particulier, dans lequel les participants n'ont pas d'autre moyen que de recourir au nom propre vocatif pour sélectionner un allocutaire parmi les autres. Ainsi pour les classes lexicales majeures, le tableau (7-5) présente la proportion qu'elles manifestent dans nos corpus. On remarquera la très faible place des adjectifs et adverbes et la part importante des verbes.

Tableau (7-5) – Classes lexicales majeures

catégorie	proportion
noms communs	11,85
verbes	15,60
noms propres	10,03
adverbes	0,32
adjectifs	0,04

En ce qui concerne les éléments grammaticaux, les indices sujets sont les items les plus fréquents. La fréquence des pronoms personnels disjoints est presque aussi importante que celle des indices sujets, et celle des pronoms qui entrent dans les structures phrastiques en positions argumentales autres que sujet est légèrement en dessous de celle des disjoints. Cette observation est à corréler avec la fréquence de structures disloquées que nous présenterons au chapitre 8.

Cette caractérisation des textes en fonction des mots lexicaux *vs* grammaticaux qui les composent, si on s'intéresse à la part d'information transmise par le lexique *vs* par la grammaire n'est qu'une vision partielle puisque ne sont pris en compte ni les marques morphologiques flexionnelles et dérivationnelles, ni les phénomènes syntaxiques et macro-syntaxiques.

Les mesures de fréquence, de densité et de diversité lexicale, si elles permettent d'obtenir des indices intéressants d'un point de vue global et permettant la comparaison entre des types de textes, ne donne pas une information qualitative sur la question des moyens lexicaux. Dans la section suivante, nous nous penchons plus précisément sur les

[243] Dans nos comptes, seuls les adverbes en *–ment* ont été retenus comme lexicaux.

[244] Picoche et Honeste (1992).

en vertu de la loi du droit d'auteur.

spécificités que présentent nos corpus du point de vue lexical.

7.3. Spécificités lexicales

L'émergence des Nouvelles Technologies de l'Information et de la Communication s'accompagne sur le plan linguistique de faits de création lexicale. De nouveaux référents, de nouveaux concepts appellent à la désignation. Les phénomènes de néologie lexicale sont assez bien connus, on peut en repérer deux grands types. Le premier type, qui passe inaperçu dans les dictionnaires dans la mesure où il n'augmente pas la nomenclature, utilise un signifiant existant auquel il ajoute un sémème. C'est le cas des unités polysémiques. Ainsi récemment, *souris* ("animal", "jeune fille", "partie charnue d'un membre") dénote en supplément par métaphore perceptuelle un objet permettant d'interagir avec un ordinateur, emploi attesté depuis 1983 selon le *Petit Robert*. Ce type est donc néologisme de signifié et non de signifiant. Le second type crée aussi du neuf avec du vieux, et l'on rencontre là nombre de procédés pour le français [245] . La création primitive, phénomène extrêmement rare et qui concerne essentiellement des produits industriels, construirait des mots entièrement nouveaux, ne respectant que la structure phonologique et syllabique de la langue. La création conventionnelle utilise des éléments de la langue et des procédés de formation connus. On distingue la création conventionnelle par composition, dérivation, abréviation, siglaison, réduplication de syllabe, substitution, mots-valises. Il faut aussi considérer les cas d'emprunts des mots d'une autre langue, pour lesquels il serait d'ailleurs plus approprié de parler de copie. Ainsi que le remarque Crowley (1997:155) "Quand une langue copie un mot d'une autre langue, elle prend la forme et la re-modèle de façon à l'adapter à sa propre structure phonologique" [246] .

Nous présenterons ici les cas qui apparaissent dans nos corpus et ce sont essentiellement des cas d'emprunt. Nous avons marqué dans nos corpus les items qui n'appartiennent pas au français de référence. Il s'agit essentiellement du point de vue de la dénotation, ainsi que le laissait supposer notre phrase introductive, de termes se rapportant aux NTIC et à l'informatique.

7.3.1. Emprunts

Pour ce qui concerne strictement le vocabulaire de l'informatique et des multimédia, une très large part est, dans un premier temps en tout cas, un emprunt à l'anglais, langue de la première puissance économique de la planète. Des stratégies commerciales de dénomination des produits en anglais sont en partie à l'origine de ce phénomène reconnu

[245] Pour cette présentation rapide nous nous inspirons de Niklas-Salminen (1997) et Mortureux (1997), ainsi que de Walter (1984).

[246] "When a language copies a lexical item, it takes the form of a word in one language and it generally reshapes that word to fit its own phonological structure".

en vertu de la loi du droit d'auteur.

dans de nombreuses langues. Cette situation perturbant les défenseurs de la langue française, une commission de terminologie [247] a été créée pour instaurer des remplaçants officiels à ces mots xénolectaux [248] . Les équivalents proposés sont des traductions lorsque cela est possible en reprenant les métaphores le cas échéant :

World wide web	toile d'araignée mondiale
Information highway	autoroutes de l'information
access provider	fournisseur d'accès

la francisation,

hypertext	hypertexte

ou la réutilisation d'items existants,

cookie	témoin (de connexion), mouchard
emoticon, smiley	frimousse
newsgroup	forum

Il est notable que quelques équivalents proposés utilisent des unités lexicales très peu courantes : l'adjectif *réticulaire* correspondant à *réseau* dans *adresse réticulaire universelle* [249] risque d'intriguer la grande majorité des francophones.

Le succès des propositions lexicales, – ou devrait-on dire prescriptions lexicales, puisqu'elles paraissent au Journal Officiel de la République [250] – est mitigé. 42% des termes pour lesquels au moins un équivalent français a été proposé par la Commission Générale de terminologie et de néologie sont employés couramment par les internautes, c'est le cas de *forum* (pour *news*) *autoroutes de l'information* (pour *information highway*), *internaute* pour *cybernaute*. 9% sont d'emplois courants mais ne présentant pas ou peu de différence de matériel signifiant avec l'équivalent étranger : par exemple *Internet*, *intranet*, *domaine* (pour *domain*). 36% connaissent un échec, c'est le cas du terme proposé pour *DVD* (*Digital Versatile Disk*) : **"disque numérique polyvalent"** . Et enfin

247 Commission générale de terminologie et de néologie.

248 La Commission générale de terminologie et de néologie ne s'occupe pas seulement du vocabulaire de l'informatique et de l'Internet, elle publie aussi des listes de termes, expressions et définitions (adoptés et publiés au Journal officiel de la République française en application du décret N°96602 du 3 juillet 1996 relatif à l'enrichissement de la langue française concernant la défense, la chimie, la stéréochimie, les techniques spatiales, l'ingénierie nucléaire, l'économie et les finances, l'automobile, la "Zone euro", les carburants, l'architecture , la monnaie européenne, le courrier électronique, le sport, les transports…

249 Équivalent proposé pour *URL (Uniform Resource Locator).*

250 Dès leur parution au *Journal officiel,* l'emploi des termes recommandés par la Commission générale de terminologie et de néologie est obligatoire pour les administrations et services de l'État (article 11 du décret du 3 juillet 1996 relatif à l'enrichissement de la langue française).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

restent 13% de termes qui connaissent des emplois techniques et pour lesquels nous ne nous prononçons pas, les professionnels interrogés, chez lesquels cet usage devrait se manifester, n'étant sûrement pas représentatifs de l'ensemble.

Il faut également préciser que certains termes reçoivent plusieurs équivalents français. Pour *URL* par exemple, la commission de terminologie propose *adresse réticulaire et adresse universelle*. On trouve sur le net et dans la presse spécialisée de nombreux articles critiques au sujet de ces propositions.

Dans nos corpus, la plupart des vocables relatifs à l'informatique et aux réseaux sont directement empruntés à l'anglais [251], certains sont francisés, et d'autres sont des traductions conformes à ce que propose la commission de terminologie.

Outre des mots, les unités empruntées à l'anglais sont de différents types. On repère quelques mots-valises tels que *codec* (*compression-decompression*) et *modem* (*modulator-demodulator*), des abréviations, des acronymes et sigles. Cet emprunt se fait selon diverses modalités. Soit les unités sont empruntées directement, et c'est en particulier le cas de quelques abréviations, des sigles et acronymes, pour lesquels la forme graphique est inchangée, et la prononciation sans doute adaptée au système phonologique du français (réinterprétation phonétique) ; soit les unités empruntées sont francisées dans la graphie et la prononciation. Un autre type d'emprunt se fait par traduction du terme utilisé en anglais.

Liste des emprunts à l'anglais apparaissant dans le corpus

[251] Les termes qui sont récemment entrés dans les dictionnaires, tels que *souris, cliquer, clic,* ne sont pas présentés ici.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Termes empruntés directement	
Informatique et réseau	
burner	nom équivalent de *graveur*, périphérique permettant de graver des CD.
download	verbe équivalent de *télécharger*
hack	nom désignant l'activité de passionnés de l'informatique
net	nom, équivalent de réseau
scan	nom équivalent de *scanner*, périphérique permettant de digitaliser des images papier et le terme désigne aussi l'opération de digitalisation et son résultat.
shell	nom désignant un interpréteur de commandes
IRC	
ban	verbe ou nom dénotant l'opération de bannir un utilisateur d'un canal.
channel	nom désignant un canal IRC
chat	verbe équivalent de *discuter* sur les réseaux
disconnect	verbe équivalent de *déconnecter*
flood	verbe ou nom dénotant l'activité d'envoi de messages répétés, importants en volume ou successifs.
founder	nom désignant le fondateur d'un canal IRC
guest	nom dénotant un utilisateur occasionnel d'un canal IRC ayant pour surnom "Guest"
join	verbe ou nom désignant l'action d'entrer dans un canal
kick	verbe ou nom désignant l'action de virer un utilisateur d'un canal
lag	verbe ou nom dénotant une lenteur de connexion, un délai dans l'acheminement des messages.
log	nom désignant un fichier d'activité d'un canal
moderate	verbe désignant l'activité de modération d'un espace public réticulaire
nuke	nom désignant un logiciel permettant de causer une panne temporaire sur une machine distante, employé comme verbe, il désigne l'action de causer une panne.
part	verbe ou nom désignant l'action de sortir volontairement d'un canal
peer	nom, équivalent de *pair* utilisé pour désigner un ordinateur de statut équivalent sur le réseau.
ping	nom ou verbe désignant l'action de vérifier le délai d'acheminement des messages
private	adjectif équivalent de *privé*
provider	abréviation de *Internet Provider* équivalent de *fournisseur d'accès Internet*

en vertu de la loi du droit d'auteur.

quit	nom désignant un fichier obtenu lors d'une sortie impromptue d'un canal
register	verbe désignant l'action de déposer un surnom auprès d'un logiciel de gestion des surnoms
script	nom désignant un type de programme informatique. Employé comme verbe, il désigne l'action de programmer.
smiley	nom désignant des entités formées de caractères représentant le plus souvent des visages.
spam	verbe ou nom désignant l'action de diffuser des messages publicitaires.
user	nom désignant les simples utilisateurs d'un canal IRC
voice	nom désignant les utilisateurs ayant le premier statut de privilèges sur un canal IRC
Composés empruntés directement	
banlist	nom désignant une liste d'utilisateurs interdits d'entrée temporairement sur un canal IRC.
eggdropp	nom désignant un type de logiciel robot permettant de gérer un canal IRC
masskick	nom désignant l'opération de sortir par force tous les utilisateurs d'un canal.
massnuke	nom désignant l'opération de causer une panne temporaire à toutes les machines des utilisateurs d'un canal.
netsplit	nom désignant un événement réticulaire résultant dans la déconnexion d'un ou plusieurs serveurs formant un réseau IRC.
nickname	nom équivalent de *surnom* ou *pseudonyme*
online	adjectif équivalent de *en ligne*
password	nom équivalent de *mot de passe*
shitlist	nom désignant une liste d'utilisateurs interdits d'entrée sur un canal de façon permanente.
webcam	nom désignant une caméra conçue pour diffuser des images mobiles en temps réel sur le web.
Mots-valises empruntés directement	
codec	*compression-decompression* – nom – élément logiciel permettant la compression et la décompression de données numériques.
modem	*modulator-demodulator* – nom – périphérique permettant la connexion à un point d'entrée sur Internet par le biais du RTC.
notilist	*notify list* – nom – liste d'utilisateurs intégrée dans un logiciel client IRC et modifiable permettant à son possesseur d'être averti de l'entrée d'un membre de la liste sur le canal).

en vertu de la loi du droit d'auteur.

spambot	*spam robot* – nom – logiciel permettant la diffusion d'un message vers n destinataires.
Abréviations empruntées directement	
chan	*channel* – nom – équivalent de *canal*
nick	*nickname* – nom – équivalent de *surnom*
op	*operator* – nom – équivalent de *opérateur*
serv	*server* – nom – équivalent de *serveur*
quit	*quit message* – message reçu à la déconnexion
pic	*picture* – nom – équivalent de *photo*.
bot	*robot* – nom – équivalent de *robot logiciel*
Acronymes et sigles empruntés directement	
Informatique	
aspi	*Advanced SCSI Programming Interface*
avi	*Audio Video Interleave* (format de fichier)
cd	*compact disc*
dvd	*digital versatile disc*
hd	*hard disc*
divx	*Digital Video Express*
ide	*Integrated Device Electronics*
mp3	*Motion Picture Experts Group 1 Audi Level 3* (format de fichier)
mpg	*Motion Picture Experts Group* (format de fichier)
pc	*personal computer*
ram	*random access memory*
rip	*Raster Image Processor*
scsi	*Small Computer Systems Interface*
Internet	
adsl	*Asymetric Digital Subscribe Line*
ip	*internet protocol*
irc	*internet relay chat*
lan	*Local Area Network*
pppoe	*Point to Point Protocol over ethernet*
smtp	*simple mail transfert protocol*
url	*uniform ressource locator*
Spécifiques IRC	
ml	*mailing list*
gt	*go together*
brb	*be right back*
gtg	*got to go*
dcc	*Direct Client Connection*
kb	*kick ban*
lmao	*loughing my ass off*
lol	*loughing out loud*
rofl	*rolling on the floor laughing*
Unités de formation mixte empruntées directement	

aop	*automatic operator*
sop	*super operator*
akick	*automatic kick*
mkick	*mass kick*
mnuke	*mass nuke*
email	*electronic mail*
Termes traduits	
canal	*(channel)*
pseudo	*(nick)*
privé	*(private)*
disque dur	*(hard disk)*
navigateur	*(navigator, browser)*
copier/coller	*(copy and paste)*
Acronymes et sigles traduits	
mdr	*mort de rire (*équivalent de *lol loughing out loud)*
asv, asl	*âge, sexe, ville ou lieu (*équivalent de *age sex location)*
Termes adaptés	
fondeur	*(founder)*
disquette	*(diskette)*
icône	*(icon)*
Termes similaires	
clone	*(clone)*
mode	*(mode)*
quitter	*(quit)* comme verbe en emploi intransitif, dénote l'action d'arrêter un logiciel. Employé comme nom il s'agit de l'événement que constitue la sortie d'un canal

On notera la très faible part dans cette liste des termes traduits ou adaptés. L'essentiel des termes spécifiques à l'IRC est représenté dans la catégorie des emprunts directs. Toutefois, on note certains phénomènes qui sont signes d'intégration dans la langue : à ces unités empruntées, les locuteurs ajoutent en effet des affixes dérivationnels [252] .

7.3.2. Intégration des emprunts

Lorsqu'une unité lexicale est empruntée, elle subit des transformations pour correspondre au patron de la langue cible. Elle prend ensuite son autonomie qui se manifeste par intégration à un niveau plus profond de la langue qui lui fait recevoir les marques de flexion puis de dérivation.

À partir de bases verbales anglaises, sont formés des verbes :

[252] Procédé couramment répertorié pour le français québécois.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

loggé log(g)-é

C'est le cas de *ban*, *kick, chat, download, flood, join, part, spam*

L'onomatopée anglaise ping est aussi utilisée comme nom ou comme verbe

pinger (v) ping-er

Les affixes et suffixes présents dans notre corpus sur les termes empruntés à l'anglais sont des morphèmes qui ont une fonction sémantique ou une fonction de transcatégorisation. Nous les présentons ci-après.

Le premier type est indécis sur la réelle dérivation française, dans la mesure où le procédé est commun à l'anglais et au français. L'absence d'accent sur le morphème *dé-* n'est cependant pas le signe de l'anglais ainsi que cela a été expliqué au chapitre 5.

autodeop	a(uto)-de-op	suppression automatique du statut d'opérateur
deop	de-op	suppression du statut d'opérateur
deaop	de-aop	suppression du statut d'opérateur automatique
deaopage	de-aop-age	suppression du statut d'opérateur automatique
deban	de-ban	suppression du bannissement d'utilisateurs

Envisageons les différents cas présents.

- *op* est un nom. Il devient un verbe par transcatégorisation sans marque [253] .

Extrait du Corpus F7 *PBP: j'op personne PBS :o((

Lorsqu'il lui est ajouté le morphème *de-* , c'est un verbe

Extrait du Corpus P9 *PBG: a celui qui deop le plus vite lolol

- *aop* est un nom formé sur *op*. On ne trouve pas de cas dans notre corpus où la forme *aop* est un verbe bien que rien ne l'interdise.

Extrait du Corpus F1 *PAD: 'tain tu as fais quoi aux zautre aop? Lol

avec le morphème *de-* il se comporte tantôt comme un nom tantôt comme un verbe

Extrait du Corpus F2 *PAL: oui PBT je suis tout a fait d'accord avec ce deaop
Extrait du Corpus F2 *PAL: et P99 a ete deaop

avec les morphèmes *de-* et -age, enfin il se comporte comme un nom.

Extrait du Corpus F7 *PHQ: PBS non moi j' ai pas le stade deaopage c akick direct

- *deban* est formé sur la base verbale *ban*. Celle-ci se comporte tantôt comme un verbe, tantôt comme un nom sur le modèle inverse de *op* :

Extrait du Corpus F7 *PBS: on les ban ou pas? Extrait du Corpus P9 *PBB: moi j'aime pas les ban!

[253] Dérivation impropre

deban se comporte dans tous les cas de notre corpus comme un verbe.

en vertu de la loi du droit d'auteur.

Extrait du Corpus P3 *PAC: PBX tu deban

L'exemple suivant, montre clairement l'intégration de l'item copié dans la langue cible :

réoper ré-op-er

Extrait du Corpus F7 *PBO: bah.. y'a bot pour vous réoper non ? ;=)

- Le suffixe *-eur* est propre au français (le correspondant anglais étant tout de même proche). On lui reconnaît une fonction sémantique (une base à laquelle *-eur* est ajouté dénote l'agent de l'action dénotée par la base) et est en même temps un indicateur de classe des noms et adjectifs.

Un seul item fonctionne sur le modèle de *théâtre théâtreux ; violon violoneux ; rebouter rebouteux.*

- Quelques rares emplois du suffixe *-en, -enne* permettent de transformer un sigle anglais emprunté tel quel pour former un adjectif :
- Quelques items présentent une formation particulière en *-ouze.*

Ce phénomène se présente aussi sur des surnoms.

7.3.3. Morphologie ludique

Le titre de cette section peut laisser songeur. Il ne s'agit ici pas à proprement parler de création lexicale, mais d'ingéniosité linguistique ludique. Nous avons vu au chapitre 5, section 5.1.4.6, que divers types de variations phonétiques sont représentés dans la graphie. Les locuteurs jouent à imiter à l'écrit les productions phonétiquement déviantes (par rapport au français standard). L'identification de ces situations comme ludiques, se base sur un faisceau d'indices (changement de style d'écriture d'un locuteur, rires des destinataires ou des destinataires secondaires, clins d'œil sous forme d'émoticones de la part de l'émetteur…). Le plan de la représentation de variations phonétiques n'est pas le seul en jeu dans ces situations. Certaines unités, par exemple, sont le résultat de la transformation d'unités sur des modèles fautifs.

cochonneté	*conchonceté*
reviendu	*revenu*
faisez	*faites*

D'autres sont le résultat d'utilisation de suffixes de la langue de façon inattendue

délinker	agir en délinquant
bafrade	de se bâfrer
retentage	de retenter

On trouve encore des redoublements de syllabes initiales

en vertu de la loi du droit d'auteur.

mumuse	jeu, de s'amuser [254]
poupougner	se faire du bien

Enfin, un certain nombre d'unités du corpus sont formées sur le modèle anglais de *coffee time*.

- boulot time
- mumuse time
- kebab time
- prout time
- café time
- dodo time
- miam miam time

Ces unités fonctionnent comme des unités phrastiques.

Tous ces phénomènes ne sont certainement pas spécifiques aux IRC, on peut constater qu'ils se produisent dans des conversations en face à face non formelles. On aura remarqué donc jusqu'ici que les spécificités lexicales viennent d'autres sources (de l'anglais pour la terminologie relative à l'informatique, au réseau et aux IRC, du français employé dans des situations non formelles pour le « ludique »). Nous allons voir que l'utilisation d'unités lexicales en marge du lexique commun standard est une autre dimension de la spécificité du vocabulaire de nos corpus.

7.3.4. Français relâché et verlan

Il est assez difficile de donner des qualificatifs pour des phénomènes linguistiques qui peuvent se présenter chez tout un chacun mais pour lesquels on sait qu'ils paraîtraient incongrus dans beaucoup de situations d'énonciations. La liberté d'expression qui se manifeste sur le net et son caractère informel se répercute dans le vocabulaire. Ainsi, nombre de vocables de nos corpus présentent un très faible degré de formalité. On y rencontre jurons en tous genres et termes équivalents de termes plus formels (doublons) :

- jurons

merde, merdouille, putain, zut

- noms de qualité

chiant, chien, chieur, con, conne, pédé, poufiasse, pute, tafiole, tarlouze…

[254] Ces termes sont sans doute issus de topolectes ou de sociolectes, ils ne sont cependant pas dans nos dictionnaires de références.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

· doublons

crade	sale
déconne	plaisanterie
populasse	peuple
petiote	petite
calbut	caleçon
godiot	chaussures
bouquin	livre
beauf	beau-frère
pif, ou paf	nez
gnon	coup
chicaner	ergoter
entuber	duper
niquer	mettre hors d'usage, détruire

Certains tabous, notamment d'ordre sexuel ou scatologiques sont levés, des unités conversationnelles entières se consacrent à ces thèmes. Certains y voient la conséquence d'une fonction de défouloir qui pourrait être associés aux espaces virtuels du fait de la forme d'anonymat qu'ils permettent. La fréquence de ces termes ne se manifeste pourtant pas dans le vocabulaire global de nos corpus.

On note aussi une présence quoique minime d'éléments de verlan :

zoubis	bisous
caifran	français
ramps	parents
chelou	louche (adj)
relou	lourd
reum	mère
meuf	femme
teuf	fête
ouf	fou (adj)
iech	chier

c'est-à-dire des termes qui ont subi une inversion de syllabes, ou pour les mots qui n'en comportent qu'une (en CV) une inversion de sons, et éventuellement une troncation en finale.

Enfin, quelques éléments de parlers jeunes se présentent, dont certains sont d'émergence assez ancienne (ex. *cool*) :

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

taf	employé comme nom ou verbe, désigne le travail ou l'action de travailler
paf	employé comme verbe, désigne l'action de se battre ou de se chamailler

Reste à apprécier la présence d'éléments topolectaux.

7.3.5. Éléments de variétés régionales du français.

On trouve dans la littérature quantité de listes des spécificités lexicale des variétés régionales en France et hors de France du français [255] . Les termes qui n'appartiennent pas au français standard présents dans nos corpus sont presque exclusivement des termes de français du Québec.

platte	mort, inerte(adj)
capoté	fou (adj)
pasteuse	?
obstineuse	obstinée
tannante	pénible
criss	juron blasphématoire
maudit	juron
torbinouche	juron
brake à bras	frein à main
char	voiture
chum	ami [256]
date	rendez-vous galant[257]
skidoo	véhicule des neiges
achaler	embêter, énerver
s'astiner à	s'entêter à
checker	vérifier, regarder[257]
typer	taper, dactylographier[257]
jaser	parler
icitte	ici
pentoute	absolument pas
sucres	fête liée à la production du sirop d'érable
tempête des corneilles	grosse tempête de neige au printemps

Nous disons presque exclusivement, puisque certains termes bien plus rares apparaissent tels que l'abréviation de *université* en *uni* qui est une abréviation courante dans le français de Suisse.

[255] Walter, H. (1988) présente des listes lacunaires.

[256] De l'anglais

en vertu de la loi du droit d'auteur.

Ces termes représentent une très faible part du vocabulaire de nos corpus. On remarque d'autre part qu'ils ne sont pas utilisés exclusivement par les participants revendiquant une identité québécoise. Des internautes qui se présentent comme français de métropole les utilisent créant une sorte de complicité avec les locuteurs Québécois.

Quelques autres termes absents de la variété standard du français métropolitain apparaissent.

chuquer	*sucer* (nord de la France)

7.4. Particularités d'ordre morphologique

La littérature reconnaît depuis longtemps une différence d'ordre morphologique entre l'oral et l'écrit. Le décompte des morphèmes, à l'écrit, est plus important qu'à l'oral. Cela se manifeste pour les marques de genre et de nombre, ainsi que pour la flexion verbale.

Nous avons observé dans le chapitre 5 l'écart à la norme que présentent nos corpus sur ce point, puisque nombre de morphèmes qu'il y aurait lieu de trouver à l'écrit selon la norme ne sont pas représentés. Le sur-marquage morphologique de l'écrit ne s'explique que par sa fonction de désambiguïsation et par le fait que les situations d'écrit traditionnel ne permettent pas au récepteur de questionner l'émetteur du message en cas d'ambiguïté sans perturber fortement le déroulement de l'interaction. Cette dernière remarque est emportée par le dispositif de communication dont sont issus nos corpus. L'orthographe, et le marquage morphologique redondant en sont fortement atteints.

Toutefois des phénomènes spécifiques méritent attention. C'est le cas de l'omission d'indices de sujet d'une part, et d'autre part d'un phénomène qui affecte les termes spécifiques aux IRC et copiés de l'anglais.

7.4.1. Indices pronominaux

La notion d'indice pronominal établie par Creissels (1995:22-31) permet de reconnaître le comportement particulier d'une partie des unités traditionnellement désignées comme pronoms que sont les pronoms personnels conjoints. La remise en cause du statut de ces unités se base sur l'observation du fait qu'elles ne se présentent pas dans la phrase en qualité de constituants nominaux ainsi que le remarquent de nombreux auteurs. Creissels (1995:24) propose de les traiter comme des "affixes du mot verbal au même titre que les morphèmes traditionnellement reconnus comme désinences personnelles du verbe". Ses arguments sont que ces unités sont à la fois référentiellement équivalentes à des éléments ayant le statut de constituant nominal et ne peuvent occuper des positions assimilables à des positions occupées par des constituants nominaux et que ces positions montrent un degré de cohésion de la séquence important. Pour le français, les indices pronominaux sont donc les pronoms personnels conjoints et les désinences personnelles

en vertu de la loi du droit d'auteur.

du verbe.

Pierre mange la pomme	Il la mange
[Nsuj Pierre] $ø_{is}$-$ø_{io}$-mange-ø [CNobj la pomme]	[Nsuj ø] il_{is}-la_{io}-mange-ø [Nobj ø]

En l'absence de constituant nominal sujet, le verbe reçoit un indice pronominal :

(8) Pierre est parti (9) *est parti (10) il est parti

Il survient que ces indices peuvent être effacés dans le cas d'enchaînements de propositions dont les sujets grammaticaux sont coréférentiels :

(11) Il a ramassé ses affaires, les a mises dans son cartable, a ouvert la porte et s'en est allé.

Nous remarquons dans nos corpus une tendance à la disparition de certains indices sujets dans des cas autres que ceux présentés ici. Nous examinerons dans quels cas cette disparition a lieu et tenterons d'en proposer une explication.

Les indices de sujet concernés sont ceux de première (a) et troisième (b) personne du singulier, et troisième personne du pluriel (c).

Exemples :

(a) Extrait du Corpus F7 *PEI: PCJ m' en fou tuer les tous!!!! :) %add: PCJ (b) Extrait du Corpus F3 *PBQ: en effet... faut de tout pour faire un monde de merde (c) Extrait du Corpus P6 *PA1: sfoute dma gueule

Les indices de première personne sont le plus souvent représentés. Les cas où ils disparaissent sont assez rares. Ils disparaissent préférentiellement avec des verbes pronominaux ou dans les temps composés (d) du verbe, mais pas exclusivement.

(d) Extrait du Corpus P9 *PAA: PBF emploi pas des mots que tu connais pas :/ %add: PBF *PAA: te les deja dis %add: PBF (e) Extrait du Corpus F5 *PBX: reconnais plus rien moi

Les indices de troisième personne qui disparaissent sont presque exclusivement des indices non personnels représentant un sujet grammatical auquel ne correspond aucun agent référentiel. Ce phénomène est relevé par la littérature Gadet (1989) comme étant le seul cas où l'indice de sujet peut être facilement omis.

Extrait du Corpus F5 *PBU: faut qu'ellesarretent Extrait du Corpus P2 *PAS: bah faudrait etre sur Extrait du Corpus P9 *PAA: fais froid ici

Les indices de troisième personne du pluriel disparaissent essentiellement sur le verbe *être.*

Extrait du Corpus F7 *PJQ: sont fatiguants avec leurs inepties et leurs discours vains Extrait du Corpus F3 *PCO: sont bons hien PCK

7.4.2. Flexions des termes IRC

Nous avons vu à la section précédente que nos corpus contiennent une partie de vocabulaire spécifique au dispositif de communication dont ils sont issus. Nous avons tenté de montrer l'intégration de ces termes dont la très grande majorité est directement

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

copiée de l'anglais à travers le fait qu'ils se conforment aux principes de dérivation du français. Cependant, ces termes présentent des particularités quant à leur flexion, notamment verbale.

Nous avons déjà souligné que le code écrit utilisé par les internautes dans ce dispositif de communication se soucie peu de la norme orthographique et manifeste tous les morphèmes qui apparaissent à l'oral, les cas de liaison ou d'élision n'étant pas systématiquement représentés. En revanche les morphèmes écrits connaissent une tendance à disparaître.

Le cas des termes spécifiques à l'IRC révèle que des morphèmes qui devraient apparaître à l'oral ne sont pas représentés.

Sur tout le corpus, nous comptons cinq cent trente-six termes spécifiques à l'IRC régulièrement intégrés à un énoncé. Les sigles IRC tels que *lol, mdr*, n'ont pas été comptabilisés dans la mesure où ils constituent des énoncés autonomes dont la fonction est essentiellement la régulation de l'activité de communication.

Il faut noter avant toute chose que certaines formes non marquées ne se prêtent pas à l'étude dans la mesure où elles se présenteraient sans distinction particulière dans le compte de morphèmes entre l'oral et l'écrit, le masculin singulier d'un nom, par exemple, se présentant sans morphème particulier à l'oral comme à l'écrit.

Restent donc les cas où une forme reçoit les morphèmes qui apparaissent à l'oral, mais pas ceux de l'écrit, ceux où une forme reçoit les morphèmes de l'écrit qui n'ont pas de contrepartie à l'oral, ceux où une forme ne présente pas les morphèmes qui devraient apparaître pour une unité de ce type à l'oral.

Pour le premier type présenté, il s'agit de noms qui devraient à l'écrit prendre la marque de pluriel 's',

Extrait du Corpus P3 *PBQ: ca amuse les op.

et de verbes qui devraient prendre une marque de flexion :

Extrait du Corpus P7 *PAC: car tu flood

Le second type présenté, est celui pour lequel la norme écrite est respectée. Il concerne des noms au masculin pluriel et des verbes à l'infinitif ou au passé composé, un verbe au conditionnel et deux impératifs.

Extrait du Corpus F2 *PAD: alors va til y a avoir une autre poussée de demission chez les aops de france? Extrait du Corpus F3 *PCO: et apres on asperge de wavs peteux sur le channel Extrait du Corpus P9 *PAA: j ai oublié de flooder Extrait du Corpus F8 *PDO: mééééé opez moi pas chuis sous java lol je peux rien faire

Enfin le troisième type ne présente aucune flexion. Les items concernés sont employés comme verbes qui auraient dû porter la marque de l'infinitif, ou du participe passé.

Extrait du Corpus P6 *PI7: je vois vraiment pas l'interet de me kick Extrait du Corpus P8 *PDJ: dites vous pouvez pas le kick le PDU Extrait du Corpus P3 *PBZ: ton script ne sais pas ban avec une commande???? Extrait du Corpus F2 *PAL: et P99 a ete deaop

Ces trois catégories apparaissent dans nos corpus selon des proportions différentes. La

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

première représente 27 % des phénomènes ici considérés. La seconde est la plus importante et représente 56 %. Enfin, la dernière catégorie représente 17 %.

On ne peut donc conclure à un type d'intégration particulier que manifesterait l'usage spécifique aux IRC. On note une tendance à représenter l'oral sans égard à la norme écrite et un type d'intégration sans adaptation au patron de la langue qui représente une proportion non négligeable. Toutefois, pour donner une mesure de ce dernier phénomène qui soit fiable, il faut rapporter le compte au cas où il apparaît effectivement une adaptation au patron de la langue. Nous avons pour ce faire pris en considération les items spécifiques aux IRC employés comme verbe. 62,71% reçoivent les flexions qui ont une contrepartie orale, et 37,29 % ne les reçoivent pas.

7.5. Conclusion

La première section de ce chapitre nous a permis, par des mesures quantitatives, de montrer que le type de texte (au point de vue de la densité et de la diversité lexicale) ressort davantage de ce qu'on considère comme de l'oral prototypique que de l'écrit prototypique. Elle a permis également d'établir que la contrainte temporelle des IRC n'a pas le même effet que la contrainte sur le nombre d'unités du télégramme puisque la part des mots grammaticaux est très importante dans nos corpus.

Cependant ces indices ne permettent pas d'étudier le lexique d'un point de vue qualitatif. Nous avons alors entrepris d'observer plus spécifiquement les unités présentes dans le corpus, qui n'appartiennent pas au lexique commun standard. En effet, nos listes de fréquences ne montrent que peu d'unités du lexique standard de faible fréquence (seuls quelques items reconnus comme rares ou vieillis apparaissent – ex. *fléau d'armes*).

Nous avons pu alors observer que le français utilisé dans les IRC d'une part, est plus anglicisé que la variété standard, précisément à cause des éléments qui concernent l'univers de référence de l'Internet, des réseaux et de l'IRC. On remarque, entre autres, que les termes empruntés ne sont pas complètement intégrés, puisqu'ils ne s'attachent pas systématiquement les morphèmes grammaticaux qu'ils devraient manifester. Il faut aussi noter que la création lexicale spécifique aux IRC en français est de très faible incidence (quelques unités seulement – ex. *ptdr, robopub…)*

Nous avons également observé qu'il peut se dégager plusieurs catégories d'items parmi ceux qui n'appartiennent pas au lexique commun standard et/ou sont encore absents des dictionnaires. Le vocabulaire de nos corpus comporte en effet des éléments qu'on attribue à la jeunesse, tant au niveau des bases que morphèmes dérivationnels. Nous avons également trouvé dans nos corpus des éléments de la variété québécoise, et de quelques autres variétés tandis qu'aucun terme du français du sud n'a été repéré.

Ainsi, si les contraintes imposées par le dispositif font que code est fortement touché, le niveau lexical l'est beaucoup moins – et l'on s'y attendait. On peut toutefois assurer que le profil du vocabulaire est particulier à plusieurs points de vue : la plus grande part du vocabulaire est constitué d'unités très courantes ; quelques îlots de termes peu fréquents

en vertu de la loi du droit d'auteur.

se manifestent (en particulier le champ sémantique des armes) ; un nombre important de termes relatifs à l'informatique et aux IRC peut être dégagé et enfin des unités issues de sociolectes de jeunes générations ainsi que des unités provenant de variétés régionales du français ponctuent les corpus. Cet ensemble permet de faire l'hypothèse que les locuteurs sur les IRC ne renient pas ce qui fait leur identité propre pour constituer une communauté linguistique nouvelle.

en vertu de la loi du droit d'auteur.

Chapitre 8 Structures syntaxiques

Nous avons observé jusqu'ici un certain nombre de particularités que présente le matériel linguistique de nos corpus. Les recherches de Biber (1988) montrent que les différents types de situations de communication influent sur les stratégies de mise en discours entre autres au niveau de la syntaxe de la phrase.

La littérature relève un certain nombre de paramètres au plan syntaxique relatifs à la spécificité des modes de production oral/écrit. Sont établies des listes de phénomènes relevant des stratégies de mise en discours typiques de l'oral ou de l'écrit. Blanche-Benveniste (1994:18), note ainsi par exemple que "Le récit oral emploie beaucoup plus de 'verbes principaux', alors que l'écrit, qui en comporte toujours moins, fait davantage recours aux participes, aux infinitifs, aux nominalisations et aux propositions relatives." On peut lire chez Koch (1995:19) relevant l'attitude de nombres de linguistes que "l'oralité est le domaine de la juxtaposition tandis que la scripturalité favorise l'intégration." Il y aurait donc entre l'oral et l'écrit une différence de complexité syntaxique.

Nous essaierons ci-après de décrire les caractéristiques au plan syntaxique qui apparaissent dans le corpus afin de déterminer s'il y a lieu de considérer que le dispositif de communication joue sur la structure des messages et dans quelle mesure.

Nous n'aborderons que sporadiquement la syntaxe des constituants pour concentrer nos investigations sur les structures phrastiques.

Dans un premier temps, nous exposons les principes du codage auquel nous avons

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

procédé sur les corpus. Nous formulons ensuite les particularités rencontrées pour les structures complexes puis pour les différentes modalités de phrases, et les dispositifs syntaxiques.

8.1. Unité de segmentation.

8.1.1. Problématique

Notre corpus présente l'utilisation d'un code particulier. Certes le mode de production est écrit, et un écrit dactylographié, il ne répond cependant pas aux normes traditionnelles de production écrite par bien des aspects ainsi que nous en avons discuté en partie dans les chapitres précédents. Nous cherchons à présent à caractériser la syntaxe de la phrase pour nos corpus, afin d'en déterminer les spécificités.

Il est de tradition de discuter le statut théorique de la phrase, et ce de façon plus polémique lorsque les analyses portent sur la langue parlée. Le terme est ancien et la linguistique contemporaine ne reçoit ce concept qu'en héritage. Les grammaires traditionnelles s'appuient pour définir la notion de phrase sur des critères hétérogènes et sujets à critique. Comme le mot, la phrase reçoit une délimitation graphique : la phrase commence par une lettre majuscule et se termine par un point. Cette délimitation graphique est appliquée comme un calque sur l'oral, dans lequel les marqueurs de délimitation correspondants sont des pauses. Enfin les grammaires traditionnelles confèrent à la phrase une qualité d'autonomie sémantique. La phrase serait au plan sémantique une unité de pensée. À cela s'ajoutent dans les grammaires récentes des définitions qui se veulent purement syntaxiques et insistent, pour reconnaître une phrase, sur la notion de cohérence syntaxique.

Nombre de critiques ont été exprimées à l'égard de la notion de phrase ainsi formulée. Le premier point est qu'elle est donnée pour un produit à reconnaître comme tel et non pas du point de vue de sa construction. D'autre part, et faut-il le souligner encore, elle se base pour la délimitation des unités sur des caractéristiques écrites qui n'entretiennent pas, malgré les tentatives d'identification des signes de ponctuation à des pauses ou à des courbes intonatives, un rapport simple avec l'oral. Un cas extrême pour illustrer la fausseté de l'identification du point à une pause ou une intonation de l'oral est l'intonation de style journalistique radiophonique ou télévisée, qui, si on devait la ponctuer selon ces critères présenterait, dans bien des cas, de graves incohérences. La notion de phrase est aussi envisagée en situation monologale [257] sans doute sous le prétexte qu'il s'agit d'une unité de langue plutôt que de parole. C'est ainsi qu'on distinguera la notion de phrase de celle d'énoncé, l'énoncé s'inscrivant dans une production effective.

Enfin, la notion de complétude sémantique engagée dans la définition ne permet absolument pas de déterminer si une séquence a le statut de phrase. Quant à la notion de

[257] Fox, B. et R. Jasperson (1995), Mondada (2001).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

complétude syntaxique, elle questionne le domaine d'œuvre de la syntaxe.

L'unité de segmentation qu'est la phrase est inutilisable telle qu'elle est définie par les grammaires traditionnelles pour nos corpus. D'une part, nos locuteurs n'utilisent pas ou utilisent peu les marqueurs de délimitation de l'écrit, ce qui ne laisse donc pas la possibilité de savoir ce qu'ils considèrent comme une phrase. D'autre part, le dispositif de production ne laisse pas non plus de traces exploitables pour une segmentation cohérente ainsi que le montrent les exemples suivants, dans lesquels les ruptures interviennent au beau milieu d'une unité syntaxique (extraits de Corpus F3 et F6), voire d'un mot (extrait du Corpus F7) ; cependant elles n'interviennent jamais, dans nos corpus, en laissant apparaître ensemble des unités totalement indépendantes les unes des autres [258] .

Extrait du Corpus F3 *PCO: c pas moi PDL regardes %add: PDL *PEJ: saklut %add: ALL *PCO: ton quit [259] %add: PDL Extrait du Corpus F6 *PEP: PAN non j'ai n5 machines %add: PAN *PEP: sous win2000 advanced server %add: PAN Extrait du Corpus F7 *PAO: s *PAO: a *PAO: l *PAO: u *PAO: t

Il convient de se donner une unité d'analyse précise qui ait un statut théorique stable et valide, aussi, nous référerons nous à la notion d'unité phrastique, telle qu'elle est discutée par Creissels (1995:31-40) et qui fait appel à la notion de contenu propositionnel et d'opération énonciative.

"Une unité phrastique est un énoncé ou fragment d'énoncé que sa structure interne permet de mettre en relation systématique avec un ensemble d'énoncés qui ont en commun de pouvoir représenter les mêmes événements conceptualisés de manière identique et qui diffèrent entre eux par la façon dont les événements en question sont envisagés dans le cadre d'une énonciation". (Creissels 1995:35)

Nous retenons également qu'une unité phrastique est une séquence d'unités linguistiques, dans laquelle se reconnaît un prédicat qui structure un ensemble de constituants nominaux [260] .

8.1.2. Solution adoptée

La clause (ou unité phrastique) constitue ainsi une unité de segmentation inférieure ou égale à la phrase traditionnellement définie [261] .

Telle que conçue par Berman et Slobin (1994:657), elle est

[258] Nous entendons là qu'une intervention n'est jamais utilisée pour adresser des messages différents à des locuteurs différents. En revanche, des unités entre lesquelles une rupture de thématique intervient peuvent apparaître dans la même intervention, mais le cas est peu fréquent.

[259] Un quit est un message apparaissant à un utilisateur après déconnexion.

[260] Selon Creissels (1995:16) il est possible de "rejeter l'idée même d'aborder la question de la définition des notions fondamentales de la syntaxe au niveau du classement syntaxique des unités lexicales ou des "mots" en acceptant l'idée que toutes les notions syntaxiques sont dérivées de celles de constituant nominal et unité phrastique".

en vertu de la loi du droit d'auteur.

…" "toute unité qui contient un prédicat unifié". Par unifié, nous entendons "un prédicat qui exprime une seule situation (activité, événement, état) incluant aussi bien des verbes finis et non finis que des adjectifs attributs." " [262]

Ces auteurs ont le mérite d'établir une méthodologie du découpage en clauses. Nous avons utilisé en grande partie leurs indications pour analyser nos corpus sur le plan syntaxique. La définition méthodologique à laquelle nous référons nécessite toutefois quelques précisions. Berman et Slobin (1994:660) indiquent en effet que le cas général présente un verbe fini par clause. Cependant des aménagements à ce cas général doivent être présentés. Seront donc considérés également comme des clauses les fragments d'énoncé correspondant aux cas présentés dans les paragraphes suivants. Le premier questionnement concerne le cas où deux formes verbales se suivent. Nous envisageons ensuite le cas de formes verbales non finies.

8.1.2.1. Auxiliaires, semi-auxiliaires, et fusion de prédicats

Lorsque deux formes verbales se suivent strictement, on est appelé à se demander s'il faut analyser ces formes comme conservant chacune son propre fonctionnement prédicatif ou si les deux formes ne constituent ensemble qu'un seul prédicat.

Pour les formes verbales composées traditionnellement reconnues, la question ne se pose même pas de tenter de les dissocier, nous verrons plus loin les tests qui permettent de l'affirmer. Cependant les phénomènes d'auxiliation reconnus à travers la littérature ne se présentent pas tous aussi clairement.

Pour le français, Benveniste (1974:177) donne, à la suite de Guillaume (1938) et Tesnière (1969), de *verbe auxiliaire* la définition de base suivante : "il s'agit d'une forme linguistique unitaire qui se réalise, à travers des paradigmes entiers, en deux éléments, dont chacun assume une partie des fonctions grammaticales, et qui sont à la fois liés et autonomes, distincts et complémentaires".

Les verbes auxiliaires seraient donc des verbes qui serviraient d'éléments morphologiques en se combinant avec d'autres formes verbales. Dans cette construction, ils perdraient tout ou partie de leur signification propre.L'auxiliation "consiste en la jonction syntagmatique d'une forme auxiliante et d'une forme auxiliée, ou plus brièvement, d'un auxiliant et d'un auxilié" [263] . L'auxiliation se caractériserait ainsi par une structure en binôme.

Benveniste (1974:179) propose que se cachent sous cette structure trois phénomènes différents et cette caractérisation est à l'origine de beaucoup d'analyses ultérieures. Les trois types d'auxiliation [264] distingués par Benveniste sont : auxiliation de

[261] De nombreux auteurs (Longacre 1999, Chafe 1982, Halliday 1985…) reconnaissent ou utilisent cette unité de segmentation, en admettant ses limites. Nous utilisons Berman et Slobin (1994) pour la documentation abondante qu'ils apportent à leur définition de base.

[262] " "any unit that contains a unified predicate" by which we mean "a predicate that expresses a single situation (activity, event or state), including finite and non finite verbs as well as predicate adjectives"".

[263] Benveniste (1974:179-193).

en vertu de la loi du droit d'auteur.

temporalité (1), auxiliation de diathèse (2), auxiliation de modalité (3) [265] , types illustrés respectivement par les exemples (1), (2) et (3) en tant qu'ils s'opposent chacun d'une façon spécifique à *il frappe* [266] :

(1) il a frappé (2) il est frappé (3) il peut frapper

Ainsi se caractérise la notion de forme verbale composée : forme qui comporte deux éléments verbaux, l'auxiliant et l'auxilié. L'auxiliant présente à lui seul les caractéristiques morphologiques d'une forme verbale (il porte les marques de temps et de personne [267]). La forme de l'auxiliation ainsi caractérisée se présente telle que dans le tableau (8-1) :

Tableau (8-1) – Forme de l'auxiliation

Auxiliant	Auxilié
forme fléchie d'un verbe	forme verbale nominale [268] d'un verbe (infinitif, participe)

L'observation avancée par Benveniste (1974:177), est que cette forme "se réalise, à travers des paradigmes entiers" et nous comprenons là également que les verbes auxiliants ne présélectionnent pas une classe de verbes particulière. Pour la forme auxiliante, des listes assez importantes se présentent dans les grammaires modernes du français, qui, de plus, procèdent couramment à une distinction entre auxiliaires et semi-auxiliaires. Cette distinction trace une ligne de démarcation entre *avoir* et *être* d'une part, et les autres verbes [269] susceptibles d'entrer dans le même type de construction binomiale d'autre part. Ce classement se superpose ainsi à la classification de Benveniste :

Tableau (8-2) – Auxiliaires et semi-auxiliaires

264 La littérature retient souvent seulement deux de ces types : auxiliation de temporalité, et de modalité.

265 Benveniste (1974:187) la définit comme "assertion complémentaire portant sur l'énoncé d'une relation". Elle n'aurait que deux modes au point de vue linguistique : possibilité et nécessité, représentés respectivement par *pouvoir* et *devoir*.

266 Benveniste (1974:179).

267 Benveniste (1974:183) insiste sur le fait que "l'auxilié […] n'est pas seulement le sémantème ; il est aussi porteur d'une partie de la fonction grammaticale. Car enfin le rôle de sémantème dans le syntagme n'exigeait pas la forme spécifique du participe passé."

268 Nous avons choisi de combiner *verbale* et *nominale* pour bien marquer qu'il s'agit d'une forme verbale et non d'une simple forme nominale d'un verbe.

269 La liste des verbes tenus pour des semi-auxiliaires varie selon les auteurs.

271 L'auxiliation temporelle marque toujours un passé.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Auxiliaires [270] *être, avoir*		Semi-auxiliaires
Auxiliation temporelle [271]	Auxiliation de diathèse	Auxiliation modale [272]
Tu as chanté	*Il est conduit*	*Il peut conduire*
Il sera parti	*Vous êtes pris*	*Tu dois dormir*
Vous aviez terminé	*Nous sommes dévorés*	*Je veux jouer*

Les auxiliaires sont distingués des semi-auxiliaires selon la structure de la forme binomiale, ainsi que le présente le tableau (8-3) :

Tableau (8-3) – Structures d'auxiliation

Auxiliaires *(être* ou *avoir)* forme fléchie + participe passé		Semi-auxiliaires forme fléchie + infinitif
J'ai fini mon travail. Nous avons perdu les clefs. Vous aviez perdu vos clefs. Il est monté à l'étage. Il a été tué sur le coup. (...)		*Il doit être sur le chemin du retour Il peut partir demain Pierre va finir son travail ce soir Je viens de voir une étoile filante Je commence à réfléchir Pierre a cessé de fumer (...)*

Parmi les verbes qu'on nomme semi-auxiliaires, d'un point de vue sémantique, certains apportent une nuance d'aspect, certains apportent une nuance de temps, et enfin, d'autres apportent une nuance modale.

Tableau (8-4) – Types d'auxiliation

temps	aspect	modalité
Sophie va rester chez elle Sophie vient de partir	*Il commence à manger Il finit de laver la vaisselle*	*Il doit prendre le train Il peut partir*

On remarquera que certaines constructions peuvent avoir plusieurs lectures. En effet, en français, l'auxiliation temporelle de futur périphrastique est parfois ambiguë, ce qui témoigne de son origine. Le verbe *aller*, dans *Sophie va chercher le pain* peut aussi bien s'interpréter comme un verbe de mouvement que comme auxiliaire du futur périphrastique.

(4) Sophie part chercher le pain. (5) Sophie va chercher le pain. Mais elle ne va pas le trouver.

[270] Il faut noter, que si dans certaines langues, il est possible de considérer l'auxiliaire comme un type particulier de verbe à montée du sujet, ce n'est pas le cas des verbes *êtres* et *avoir* en français si on maintient le principe selon lequel "les clitiques pronominaux objet et datif s'affixent au verbe dont ils représentent un complément" Creissels (2000c:12;9).

[272] Benveniste (1974:187-192) distingue les modalisants de fonction *devoir* et *pouvoir*, et les modalisants d'assomption *aller, falloir, désirer, espérer, savoir, faire* "selon qu'ils ont exclusivement ou non la construction avec l'infinitif auxilié."

[271] L'auxiliation temporelle marque toujours un passé.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Cette ambiguïté est prégnante lorsque l'action dénotée par l'auxilié implique un déplacement de la part de l'agent. En revanche si l'action dénotée est en fait un déplacement de l'agent, il semble que l'interprétation évite le pléonasme et se concentre sur une lecture strictement temporelle.

(6) Elle va partir. (7) Elle va entrer. (8) Elle va voyager. (9) Elle va arriver.

La lecture temporelle est plus probablement levée lorsqu'une spécification de localisation/destination est présente :

(10) Elle va chercher le pain à la boulangerie. (11) Elle va chercher du travail à Paris.

Pour le second auxiliaire de temporalité l'ambiguïté ne se présente pas naturellement.

(12) ? Elle vient de chercher le pain à la boulangerie. (13) ? Elle vient de chercher du travail à Paris.

Si maintenant on considère les semi-auxiliaires de modalité, le premier écueil réside dans la conception de la notion même de modalité. La conception restreinte, dans la lignée d'Aristote, pose qu'il y a modalité dans le cas où le contenu d'une proposition se trouve modifié par une idée de nécessité, d'impossibilité, de possibilité ou de contingence – position adoptée par Benveniste (1974:187-188), mais à laquelle toutefois ce dernier déroge en présentant *aller,* et *faire,* entre autres, au chapitre de l'auxiliation de modalité. La conséquence d'une telle modification est le renforcement ou l'affaiblissement de l'assertion d'une proposition. La conception plus large étend le terme de modalité à tous les cas où le contenu d'une proposition se trouve transformé dans un sens quelconque.

Von Wright (1951:1-2) donne le tableau des modalités suivant :

Tableau (8-5) – Modalités selon Von Wright (1951:1-2)

Alethic	Epistemic	Deontic	Existential
necessary	verified	obligatory	universal
possible	-	permitted	existing
contingent	undecided	indifferent	-
impossible	falsified	forbidden	empty

Les modalités aléthiques sont les modes de la vérité, les épistémiques sont les modes du savoir, les déontiques ceux de l'obligation et les existentielles les modes d'existence.

Un examen rapide de la littérature [273] concernant les modalités permet d'apprécier l'étendue des notions qui sont considérées comme relevant de l'attitude du sujet parlant à l'égard de son énoncé [274] . Le tableau (8-6) recense les principales d'entre elles :

Tableau (8-6) – Diversité des modalités [275]

[273] Brunot (1922), Confais (1995), Guillaume (1929), Bally (1965), Le Querler (1996), David, J. & G. Kleiber (eds) (1983), Martin (1987), Palmer (1986), etc.

[274] D'après la définition de Bally (1965).

en vertu de la loi du droit d'auteur.

dénominations	domaines	exemples d'opérateurs
aléthiques ou ontiques	ce qui est, ontologie	*nécessaire, possible, contingent, impossible*
déontiques	ce qui doit être, conformément à une règle morale, sociale, conventionnelle	*obligatoire, permis, interdit*
épistémiques	le savoir, les croyances	*savoir, croire*
temporelles	le temps	*toujours, parfois, jamais*
axiologiques [276]	évaluation du contenu propositionnel	*bien, mal*
bouliques [277]	la volonté du sujet	*espérer, craindre, douter...*
érotériques [278]	questions	

Les listes présentées semblent bien hétérogènes et varient d'un auteur à l'autre. Les semi-auxiliaires modaux les moins contestés sont *devoir* et *pouvoir* [279] dans leurs acceptions épistémiques (14b, 15b) et déontiques (14a, 15a).

(14) a. Tu peux partir Il t'est permis de partir. b. Il se peut que tu partes. (15) a. Il doit partir il est obligé de partir. b. Il se peut qu'il parte.

Cet aperçu des analyses originelles de suites de formes verbales présente l'intérêt essentiel d'un inventaire. Les distinctions sémantiques entre les différents types d'auxiliations étant posées, la question reste de savoir s'il faut dans tous les cas considérer qu'effectivement les deux formes verbales qui se suivent constituent deux prédicats ou si les deux formes n'en constituent ensemble qu'un seul.

Parmi les séquences recensées, pour une partie, l'ensemble auxiliant-auxilié ne présente pas les caractéristiques syntaxiques qu'on pourrait attendre d'une combinaison de deux termes prédicatifs. Cet ensemble s'analyse alors globalement comme un terme prédicatif unique équivalent à une forme verbale simple issue du lexème identifiable dans l'auxilié.

Ainsi, l'auxiliaire (*être* ou *avoir*) n'a aucune valence propre, il est dit servir de support au verbe recteur (auxilié). Les pronoms clitiques du verbe recteur se portent en effet sur l'auxiliaire en (16b) et (17b), qui leur sert de support morphologique. La séquence de verbes forme un seul domaine pour la sous-catégorisation.

(16) a. Paul a perdu ses livres. b. Paul les a perdus. (17) a. Il est devenu méchant.

[275] d'après Gardies (1979).

[276] du grec ἀξία,-ας, valeur, prix – d'après Bailly (1935).

[277] du grec βούλομαι, vouloir, désirer – d'après Bailly (1935).

[278] du grec ἐρωτάω-ῶ, interroger – d'après Bailly (1935).

[279] Sueur (1979), Fuchs (1989).

en vertu de la loi du droit d'auteur.

b. Il l'est devenu.

On remarque d'autre part que dans l'inventaire de verbes qui se manifestent en première position de la structure binomiale, seuls *être* et *avoir* présentent cette dernière caractéristique, c'est à dire seuls les verbes entrant dans la construction manifestant un auxiliant fléchi suivi forme verbale au participe passé [280] , phénomène qu'on dénommera après Creissels (2000c:12;1-10) fusion de prédicats [281] .

Le deuxième type de construction pose problème, le test des clitiques en effet ne permet pas toujours de procéder à la même analyse. De plus, d'autres verbes que ceux mentionnés jusqu'ici apparaissent en tant que premier membre d'une suite verbe fléchi suivi d'un infinitif.

(18) Pierre tente de répondre à Marie. (19) Pierre semble apprécier tes conseils. (20) Il faut courir plus vite. (21) Sophie court acheter le pain. (22) Pierre fait réparer la moto à Paul. (23) Pierre laisse réparer la moto à Paul.

Les divers tests exposés ci-après permettent de former des groupes de verbes.

Les clitiques ne se placent pas de la même façon que pour les constructions identifiées plus haut comme cas de fusion de prédicats, excepté pour les suites de verbes dont *laisser*, *faire* sont les premiers éléments.

(24) a. Pierre vient de terminer ses devoirs. b. Pierre vient de les terminer. (25) a. Pierre commence à comprendre la leçon. b. Pierre commence à la comprendre. (26) a. Jacques doit acheter la maison. b. Jacques doit l'acheter. (27) a. Pierre espère acheter des cigarettes. b. Pierre espère les acheter. (28) a. Pierre vient acheter des cigarettes. b. Pierre vient les acheter. (29) a. Pierre pense réparer la moto demain. b. Pierre pense la réparer demain. (30) a. Pierre tente de répondre à Marie. b. Pierre tente de lui répondre. (31) a. Pierre fait réparer la moto à Paul. b. Pierre la lui fait réparer. (32) a. Pierre laisse réparer la moto à Paul. b. Pierre la lui laisse réparer. (33) a. Pierre semble apprécier tes conseils. b. Pierre semble les apprécier. (34) a. Il faut réparer la moto. b. Il faut la réparer.

Le premier groupe, pour lequel nous avons choisi comme représentant *venir de*, ainsi que le deuxième, le cinquième le huitième et le neuvième groupes, présentent un fonctionnement particulier qui ne permet pas d'identifier l'infinitif comme tête d'une unité phrastique en position de complément [282] alors que cela pourrait être le cas des infinitifs des autres groupes.

[280] Des constructions présentant un verbe fléchi suivi d'un infinitif que nous exposons plus loin entrent dans cette catégorie (*voir, laisser, faire)*.

[281] *Clause union*, ou *clause reduction* sont les termes anglais proposés par la grammaire relationnelle -initiée par Perlmutter et Postal dans les années 70 – pour ce phénomène cf. Aissen & Perlmutter (1983), Blake (1990:104-130). La littérature – Edmonds (1978), Abeillé, Godard & Sag (1998), Abeillé, Godard (à paraître) – utilise également le terme de prédicat complexe. On trouvera une présentation claire de ce phénomène dans différentes langues dans Givón (2001:39-90) qui postule que cette construction présente un isomorphiste entre les niveaux syntaxiques et sémantiques qui respectivement opèrent à une intégration phrastique et à une intégration d'événements.

[282] Argument du premier verbe.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

(35) a. Pierre vient de terminer ses devoirs. b. *Pierre en/le vient. (36) a. Pierre commence à comprendre la leçon. b. *Pierre y/le commence. (37) a. Jacques doit acheter la maison. b. Jacques le doit. (38) a. Pierre espère acheter des cigarettes. b. Pierre l'espère. (39) a. Pierre vient acheter des cigarettes. b. *Pierre le vient. (40) a. Pierre pense réparer la moto demain. b. ?Pierre le pense. (41) a. Pierre tente de répondre à Marie. b. ?Pierre le tente. (42) a. Pierre fait réparer la moto à Paul. b. *Pierre le fait. (43) a. Pierre laisse réparer la moto à Paul. b. *Pierre le laisse. (44) a. Pierre semble apprécier tes conseils. b. Pierre le semble. (45) a. Il faut réparer la moto. b. Il le faut.

Les possibilités d'extraction vont également dans ce sens, ainsi que le présente la série d'exemples ci-dessous.

(46) a. Pierre vient de terminer ses devoirs. b. *c'est de terminer ses devoirs que Pierre vient. (47) a. Pierre commence à comprendre la leçon. b. *c'est à comprendre la leçon que Pierre commence. (48) a. Jacques doit acheter la maison. b. c'est acheter la maison que Jacques doit. (49) a. Pierre espère acheter des cigarettes. b. C'est acheter des cigarettes que Pierre espère. (50) a. Pierre vient acheter des cigarettes. b. *C'est acheter des cigarettes que Pierre vient. (51) a. Pierre pense réparer la moto demain. b. *C'est réparer la moto demain que Pierre pense. (52) a. Pierre tente de répondre à Marie. b. ?C'est de répondre à Marie que Pierre tente. (53) a. Pierre fait réparer la moto à Paul. b. *c'est réparer la moto que Pierre fait. (54) a. Pierre laisse réparer la moto à Paul. b. *c'est réparer la moto que Pierre laisse. (55) a. Pierre semble apprécier tes conseils. b. ? c'est apprécier tes conseils que Pierre semble. (56) a. Il faut réparer la moto. b. c'est réparer la moto qu'il faut.

Il en est le cas aussi de la possibilité de faire commuter l'infinitif avec une complétive dont le sujet est [283] ou n'est pas coréférentiel à celui de la phrase supposée matrice.

(57) *Pierre vient de que je termine mes devoirs. (59) *Pierre commence à que je comprenne la leçon. (60) *Jacques doit que j'achète la maison. (61) Pierre espère que j'achète la maison. (62) *Pierre vient que j'achète des cigarettes. (63) Pierre pense que je répare la moto demain. (64) *Pierre tente que je réponde à Marie. (65) *Pierre fait que je répare la moto. (66) *Pierre laisse que je répare la moto. (67) *Pierre semble que j'apprécie les conseils. (68) Il faut que je répare la moto.

On note également que la négation n'est pas également possible en deux sites dans tous les groupes et présente des différences sémantiques variables selon les séries.

(69) a. Pierre ne vient pas de terminer ses devoirs. b. ? Pierre vient de ne pas terminer ses devoirs. (70) a. ? Pierre ne commence pas à comprendre la leçon. b. Pierre commence à ne pas comprendre la leçon. (71) a. Jacques ne doit pas acheter la maison. b. Jacques doit ne pas acheter la maison. (72) a. Pierre n'espère pas acheter des cigarettes. b. Pierre espère ne pas acheter des cigarettes. (73) a. Pierre ne vient pas acheter des cigarettes. b. ?Pierre vient ne pas acheter des cigarettes. (74) a. Pierre ne pense pas réparer la moto demain. b. Pierre pense ne pas réparer la moto demain. (75) a. Pierre ne tente pas de répondre à Marie. b. Pierre tente de ne pas répondre à Marie. (76) a. Pierre ne fait

[283] Afin de ne pas encore alourdir cette série d'exemples, la forme présentant des sujets coréférentiels n'est pas présentée. Elle devrait en effet être étoilée dans les mêmes cas.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

pas réparer la moto à Paul. b. *Pierre fait ne pas réparer la moto à Paul. (77) a. Pierre ne laisse pas réparer la moto à Paul. b. *Pierre laisse ne pas réparer la moto à Paul. (78) a. Pierre ne semble pas apprécier tes conseils. b. Pierre semble ne pas apprécier tes conseils. (79) a. Il ne faut pas réparer la moto. b. Il faut ne pas réparer la moto.

D'un point de vue syntaxique, si deux formes verbales répondent aux critères mentionnés ci-dessus pour reconnaître le phénomène de fusion de prédicats, bien qu'en structure sous-jacente deux unités phrastiques puissent être postulées, on n'en comptera pour notre part qu'une seule. En dehors des formes verbales composées, nous l'avons vu, sont également concernées les structures causatives et certains autres verbes, qui peuvent ou non présenter la construction tels que *laisser*, *voir*. Toutefois des différences de comportements se manifestent par rapport aux auxiliaires de temps [284] . Sont également analysés comme des cas de prédicat complexe la série de verbes qualifiés de semi-auxiliaire de temps et d'aspect, ainsi que les semi-auxiliaire modaux *devoir* et *pouvoir* [285] , *penser* [286] et *savoir* ; les verbes de mouvement *aller*, *venir*, ainsi que *entendre* et *envoyer.*

D'un strict point de vue syntaxique, les autres cas mentionnés ici devraient faire l'objet de deux clauses.

8.1.2.2. Gérondifs, participes et infinitifs (avec ou sans sujets propres)

Nous avons également marqué en tant que clauses les cas traditionnellement identifiés comme propositions participiales, et infinitives avec sujets coréférentiels ou non, tels que les exemples ci-dessous.

(80) J'ai aperçu Pierre faisant ses courses. (81) J'ai aperçu Pierre en faisant mes courses. (82) Faisant mes courses, j'ai aperçu Pierre. (83) Arrivé à la gare, Pierre cherche son train. (84) Pierre cherche son train arrivé à la gare. (85) Pierre achète des planches pour faire une étagère. (86) Faire les courses est épuisant.

Talmy (2000b:215) propose que des processus cognitifs très généraux qu'il dénomme "découpage conceptuel" [287] et "attribution du statut d'entité" [288] , opèrent dans l'esprit humain en perception ou en conception pour poser une limite autour d'une portion de ce qui pourrait être autrement un continuum. Il examine qu'un événement peut être représenté syntaxiquement par une phrase simple, tandis que d'autres peuvent être représentés par une phrase complexe. La structure de l'information dans la représentation syntaxique d'un même événement pouvant se réaliser sous forme simple ou complexe,

284 Cf. Abeillé, Godard & Sag (1998).

285 *vouloir* sera considéré comme élément prédicat complexe dans des cas particuliers seulement – lorsqu'il est combiné avec *dire* par exemple.

286 Lorsqu'il est construit directement.

287 "conceptual partitioning".

288 "ascription of entityhood".

en vertu de la loi du droit d'auteur.

elle dépend donc de son mode de conceptualisation synthétique ou analytique.

Talmy (2000b:213) remarque que les langues du monde semblent se diviser en deux groupes sur la base du patron caractéristique dans lequel la structure conceptuelle d'un complexe événementiel est mappée sur la structure syntaxique. Dans certaines langues le noyau serait exprimé par le verbe principal tandis qu'il le serait par le satellite [289] dans les autres. Le français compte parmi les langues qui font porter l'événement principal par le verbe, tandis que le co-événement intervient typiquement sur un syntagme adpositionnel ou un constituant de type gérondif. Il en est effectivement le cas de certaines structures de nos corpus.

Extrait du Corpus F8 *PZZ: PBO tourne autour d PDT en lui tirant la langue

Toutefois notre découpage concerne ici les unités syntaxiques, nous compterons donc deux clauses.

8.1.2.3. Dispositifs syntaxiques

Nous avons considéré comme constituant deux clauses certains dispositifs syntaxiques [290] impliquant un verbe qui, selon Blanche-Benveniste (1997:92-94), n'est pas recteur mais construit un dispositif, "auxiliaire de la détermination nominale".

Extrait du Corpus F6 *PDH: yen ah qui s'en sont foutu plein les poches avec sa

Cela nous amène cependant à ne pas compter au rang des relatives propres la seconde partie de la construction puisque nous considérons cette partie comme élément d'une construction plus large.

On peut ici admettre que l'on a affaire à une double assertion : assertion d'existence couplée à une assertion à propos de la chose assertée exister. Selon une première définition de travail de Cadiot et Furukawa (2000:3), "La prédication seconde c'est celle qui est réalisée par un type de séquence qui malgré son statut syntaxiquement intégré, exprime sémantiquement un contenu phrastique à l'intérieur même de la phrase" [291] . Cette position serait confortée par la négation possible en deux sites :

(87) a. Il y en a qui ne font pas de pub. (88) b. Il n'y en a pas qui font de la pub. (89) c. Il n'y en a pas qui ne font pas de pub.

Toutefois dans le cas présent (87) l'impossibilité de considérer *en* comme un anaphorique, est une entrave à une telle considération.

D'autres types de structures se présentent avec une relative et nous les traitons de la même façon pour le découpage.

Extrait du Corpus P2 *PZZ: PAR a les mains qui fument

[289] Pour Talmy (2000a et 2000b), un satellite du verbe est la catégorie grammaticale de tout constituant autre qu'un complément nominal qui est en relation sœur avec la racine du verbe. Un satellite peut être un affixe lié ou un mot libre.

[290] Blanche-Benveniste (1991) et (1997).

[291] Il faut se garder des attitudes excessives manifestées par les premières versions de la grammaire transformationnelle de Chomsky (1968) dans laquelle toutes sortes de constituants nominaux étaient analysés comme le résultat d'une transformation à partir d'une structure profonde de nature phrastique.

en vertu de la loi du droit d'auteur.

Cela amène évidemment à perdre dans une certaine mesure l'unité de la construction, qui se manifeste, et à découper de même les attributs de l'objet [292] . Les relatives ainsi comptabilisées seront marquées en fonction de la structure plus large à laquelle elles appartiennent.

8.1.2.4. Clauses sans verbe

Les clauses sans verbe sont nombreuses dans notre corpus. En effet, le mode dialogique implique des réductions discursives [293] . Ces clauses font l'objet d'un marquage particulier. Nous avons également marqué d'une façon particulière les interventions comportant des rires ou smileys.

Enfin, certaines phrases ne présentent pas de prédicat explicite

(90) Cool, ce chan.

Cependant, ainsi que l'indique Creissels (1995:51) au sujet du hongrois, si on souhaite exprimer ce contenu propositionnel avec une référence temporelle de passé, une forme prédicative est forcément rétablie. Nous considérerons ce cas comme une clause.

Toujours en dehors des cas de réduction discursive, Le Goffic (1993:509-523) présente plusieurs types d'énoncés sans verbe : phrase nominale attributive, locative, ou existentielle [294] . Ainsi l'exemple (90) s'analyse comme un prédicat qui prend la forme d'un adjectif, tandis que le thème ou sujet est un syntagme nominal. Le prédicat peut prendre la forme d'un SN, et le sujet celle d'un SN complexe ou d'unité phrastique.

Ainsi, ces clauses peuvent évidemment être matrices, les exemples suivants l'illustrent :

(91) Cool, ce chan que tu m'as indiqué. (92) Cool que tu sois venu.

Certaines phrases nominales peuvent également ne manifester que le prédicat, le sujet restant implicite :

(93) Cool.

On remarquera que l'ordre est inversé par rapport à l'ordre canonique Sujet-prédicat, toutefois, dans des cas plus rares, ce dernier peut-être respecté :

(94) Chose promise, chose due. (Le Goffic, 1993:519)

Enfin, certaines unités phrastiques ont été marquées comme s'écartant de la norme. En effet, si à l'oral on peut repérer des stratégies particulières de formulation [295] et reformulation, l'écrit digital permet l'effacement "on-line", les phénomènes – si on peut supposer qu'ils ne sont pas tous du même type [296] – ne sont donc pas visibles. Certains énoncés se manifestent pourtant comme mal formés et sont l'objet ou non de corrections auto ou hétéro-effectuées. Ces énoncés déviants ont également fait l'objet d'un balisage

[292] On compte 66 adjectifs attributs de l'objet dans nos corpus.

[293] Blanche-Benveniste (1992).

[294] Selon leur sémantisme puisque des constructions similaires se trouvent dans les trois catégories.

[295] Blanche-Benveniste (1997).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

distinctif [297] .

Les unités phrastiques ainsi dégagées, il convient d'examiner la façon dont elles s'assemblent pour former le texte conversationnel. Nous essaierons ci-après de définir les limites dans lesquelles ces unités peuvent être envisagées d'un point de vue syntaxique.

8.2. Phrase simple vs complexe

Il existe un lien implicite dans la théorie syntaxique entre intégration syntaxique et complexité. La littérature [298] présente plus ou moins ouvertement les équations "parataxe = simplicité", "hypotaxe = complexité". Il y paraît implicitement ou explicitement l'hypothèse que l'emploi de la parataxe serait l'apanage de l'oral tandis que les phénomènes d'intégration se présenteraient essentiellement à l'écrit.

La phrase complexe est définie comme une unité phrastique intégrant plusieurs unités phrastiques [299] . Dès lors se pose le problème des limites de l'unité supérieure ainsi considérée et des types d'intégration.

La grammaire traditionnelle distingue deux types d'intégration phrastique : subordination et coordination. D'une part ces notions sont remises en cause (Givón 2001:327-387) [300] et d'autre part la linguistique textuelle renouvelle les problématiques en ajoutant à un niveau supérieur l'étude des marqueurs d'organisation entre les phrases [301] .

[296] Il est sûrement des phénomènes relatifs spécifiquement à l'écrit dactylographié, tels que la correction orthographique, le déplacement de mots ou groupes de mots…– qui se manifestent dans le secret du locuteur.

[297] On note que souvent, les corrections, quand elles existent, ne surviennent qu'après avoir été pleinement intégrées sans mention métalinguistique au corps de conversation en cours. Il se produit donc un rétablissement de la forme correcte en fonction des informations disponibles dans le co-texte et le contexte élargi aux connaissances partagées et encyclopédiques, au système d'inférence sur les intentions d'autrui, par le(s) destinataire(s) du message déviant.

[298] Beaman (1984), Kroll (1977), O'Donnell (1974), Chafe, (1982), Lakoff (1979), Ochs (1979), Koch (1995).

[299] La question de la complexité n'est pas triviale. On associe souvent simplicité avec facilité et complexité avec difficulté de traitement. Pourtant, des tâches complexes peuvent être réalisées facilement après entraînement. Faut-il alors considérer que la complexité se définit par le nombre de modules de traitements impliqués et les relations qu'ils entretiennent ? La complexité est-elle du coté de l'objet ou du côté de son traitement ? Nous envisagerons ici cette question de façon traditionnelle, on se reportera à Gayraud (en préparation)pour une discussion approfondie.

[300] En effet, l'opposition se formule en termes de dépendance. Le caractère d'indépendance des coordonnées est, selon Givón (2001:327-328), illusoire qu'on le considère d'un point de vue fonctionnel ou syntaxique, la dépendance serait plutôt une question de degrés puisque "aucune clause dans un discours connecté, cohérent n'est fonctionnellement 100% indépendante de son contexte local ('linéaire') ou même global ('hiérarchique').", en anglais dans le texte : "no clause in connected coherent discourse is functionnaly 100% independant of its local ('linear)' or even global ('hierarchic') context."

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Il convient de déterminer ce que la grammaire entend par phrase complexe étant entendu que l'étude de ces phénomènes est fortement influencée par la culture de l'écrit. On n'aura pas de peine à reconnaître, pour la notion de subordination qu'il s'agit d'une notion de dépendance (comme le subordonné prend ses ordres de son supérieur hiérarchique), mais on s'interroge au vu des faits présentés sous cette section par la grammaire sur la nature des unités en relation de dépendance et sur la nature même de la dépendance. D'autre part on se demandera quels critères permettent d'opposer et d'identifier systématiquement coordination et subordination.

Andersen (1995:43-56) tente d'établir une typologie des relations de dépendance propositionnelle. Elle distingue trois types, conformément à la tendance générale : subordonnée, coordonnée et juxtaposée, les deux premières étant typiquement introduites respectivement par un élément subordonnant ou coordonnant, tandis que les dernières ne seraient introduites par aucun morphème.

Il est intéressant d'affiner ces distinctions et de considérer, avec Koch (1995:13-42) qu'il existe différents niveaux d'imbrication. Ainsi, on serait à même de rendre compte d'énoncés tels que les suivants :

(95) Si tu écoutais les disques que je te prête, tu reconnaîtrais l'air qui passe.

Koch (1995:13-42) distingue ainsi six types d'intégration, en référence à l'étude du phénomène de jonction présenté par Raible (1992) [302] . Le premier type, celui de la parataxe asyndétique, correspond aux cas où les clauses sont simplement juxtaposées.

Extrait du Corpus F3 *PCA: et parlez pas aux chiens inconnus *PCA: on sait jamais

Le second type est celui de la parataxe à reprise (anaphorique). Koch ne donne pas de précision à ce sujet, on peut cependant envisager plusieurs types : soit l'élément anaphorique reprend toute une unité phrastique (c'est le cas de l'exemple de Koch), soit il n'en reprend qu'un élément. Le premier type ne se trouve dans nos corpus que lorsque *ça* reprend une unité phrastique précédente ; tandis que le second type est plus fréquent.

Extrait du Corpus F3 *PBD: moi non plus, et en plus, ce soir grenoble joue a domicile... ca va faire du barouf ds les rues :(

Le troisième type concerne les mises en relation par des connecteurs

Extrait du Corpus F3 *PAY: PBW je comprends mo j'en ai marre de la neige (1) *PAY: mais dans 15 jours on part au soleil :)

Koch (1995:16) lui donne le nom de parataxe syndétique.

Le quatrième type dégagé comprend deux sous types et concerne les cas d'hypotaxe (subordination) à verbe conjugué (relatives) et non conjugué (infinitives).

Extrait du Corpus F3 *PAA: PBF emploi pas des mots que tu connais pas :/
Extrait du Corpus P8 *(action): PAN s echauffe les doigts pour clik plus vite :)

Le second sous-type serait plus intégratif que le premier dans la mesure où un élément de la phrase matrice se représente en tant que trace dans l'unité intégrée (ce qui est le cas

[301] Roulet *et al. (*1987).

[302] Cité par Koch (1995).

en vertu de la loi du droit d'auteur.

également de la relative) et où l'absence de marques temporelles et personnelles sur le verbe de cette unité induirait une dépendance plus importante, la référence temporelle du procès dénoté par le verbe intégré s'appuyant sur les marques portées par le verbe de l'unité phrastique matrice.

Le cinquième type d'intégration relevé par Koch (1995:18) regroupe les cas d'hypotaxe à verbe non conjugué participe et infinitif. Koch (1995:18) propose d'inclure ici des "adjectifs, voire des substantifs employés à la manière d'un participe".

Extrait du Corpus F3 *PAY: 'vais prendre une douche en écoutant du metal a fond, ca devrait me detendre...

Enfin, le type de jonction le plus intégré selon Koch (1995) – donc le sixième type – serait la nominalisation, dans laquelle la représentation d'un état de choses du monde apparaîtrait sous forme d'un nom [303] .

On peut critiquer cette hiérarchie à plusieurs points de vue [304] , mais notamment sur le fait qu'elle ne tient pas compte de la distinction entre les trois types de relations de subordination.

Il est distingué en effet deux types de subordonnées, selon la fonction de l'unité phrastique enchâssée dans l'unité phrastique matrice. Foley & Van Valin (1984) analysent ainsi une unité phrastique en trois éléments : le nucleus, le *core* (coeur, *i. e.* comprenant les arguments valenciels du verbe), et la périphérie (comprenant les autres éléments, *i. e*. circonstants et participants secondaires). La proposition est donc de distinguer entre les unités phrastiques enchâssées faisant partie du noyau et celles relevant de la périphérie. Les unités phrastiques "subordonnées" faisant partie du noyau ou de la périphérie ont la propriété de connaître des possibilités de modulation énonciative bloquées, et les celles qui font partie de la périphérie fonctionnent comme des constituants nominaux en fonctions d'obliques. On peut noter également que les unités phrastiques faisant partie du noyau ou de la périphérie peuvent être modifieur d'un constituant nominal appartenant à un constituant argument central ; ainsi les complétives peuvent saturer la valence d'un verbe et faire alors partie des arguments 'core', ou encore saturer la valence d'un nom tandis que les relatives se trouvent essentiellement en fonction de modifieurs de noms.

Toutefois, Koch (1995) introduit un autre critère : celui des degrés de subordination. En effet, les types 4 à 6 peuvent être intégrés à divers degrés, ainsi que nous l'annoncions. Sont distingués 4 degrés. L'énoncé (96) présente deux subordonnées de premier degré et une subordonnée de deuxième degré.

(96) Si tu écoutais les disques que je te prête, tu reconnaîtrais l'air qui passe.

En effet, la phrase matrice intègre en position de modifieur du nom une relative (premier degré) et en position d'adverbial une conditionnelle (premier degré) qui elle-même intègre en position de modifieur du nom une relative (deuxième degré).

[303] Une position similaire est adoptée dans Halliday (1985), Chafe (1982).

[304] En particulier, les critères permettant d'établir la hiérarchie de sont pas clairement problématisés. Il serait par exemple intéressant d'étudier si l'interprétation des relations sémantiques entre phrases est facilitée par la présence d'un élément morphologique ou non. D'autre part les critères précis d'identification des structures n'apparaissent pas dans le texte.

en vertu de la loi du droit d'auteur.

Koch (1995) examine également la position de la subordonnée par rapport à la régissante. La subordonnée peut se trouver à droite, c'est-à-dire à la suite de la régissante. Dans notre exemple, la relative *qui passe* a cette propriété. Elle peut se trouver à gauche, c'est-à-dire qu'elle précède strictement la proposition régissante, ainsi que le fait la conditionnelle dans notre exemple. Enfin, la subordonnée peut être emboîtée dans de la proposition régissante ainsi que l'est la relative de deuxième degré de notre exemple.

Nous avons donc codé les structures enchâssées selon ces critères et en distinguant les différents types de clauses enchâssées selon leur fonction syntaxique : compléments argumentaux, modifieurs du nom, adverbiaux (circonstancielles), ainsi que le degré de leur imbrication.

Quant aux unités phrastiques coordonnées, nous ne les considérons comme relevant de la syntaxe de la phrase que dans deux cas présents dans nos corpus. Le premier concerne la coordination d'unités phrastiques subordonnées (exemple 97), et le second la coordination d'unités phrastiques modifiant des propriétés morphosyntaxiques des énoncés indépendants correspondant (exemple 98). Les autres cas de coordination d'unités phrastiques sont relégués au niveau discursif.

(97)

Extrait du Corpus F2 [305] *PZZ: PAS cherche une nana moche, conne, bossue avec des poils, aimant les nouilles et faire de la mobylette

(98)

Extrait du Corpus F7 *PEI: PAK tu n' a rien lu, rien vu lol

Nous avons également codé le type sémantique des enchâssées adverbiales et relatives, ainsi que leur position dans la phrase matrice. À ce sujet, il convient d'apporter quelques précisions.

On distingue en effet trois positions pour les unités phrastiques subordonnées : enchâssée au centre, à droite ou à gauche. Ces critères ne s'imposent pas d'eux même. On considère la position à droite dans la mesure où l'unité phrastique subordonnée apparaît en fin de phrase.

(99) Je mangerai le gâteau qui est sur la table.

La position à gauche est celle tenue par une unité qui se place avant l'unité phrastique matrice :

(100) Si tu es d'accord, je mangerai le gâteau.

On peut noter d'emblée que les relatives dans cette position seront d'un type particulier et peu fréquent (cf. infra).

La position au centre, est celle d'une unité phrastique qui interrompt l'unité phrastique matrice :

(101) La jeune fille qui est venue a mangé le gâteau.

Dans notre codage, la position de l'unité phrastique subordonnée est toujours observée par rapport à l'unité phrastique matrice du niveau immédiatement supérieur. Ainsi, dans

[305] Cet exemple contrevient au principe de "coordination of likes".

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

l'exemple (95) qui apparaît au début de notre exposé,

(95) Si tu écoutais plus souvent les disques que je te prête, tu reconnaîtrais l'air qui passe.

la relative de niveau 2 *que je te prête* est dite à droite, bien qu'elle se trouve à gauche de la principale *tu reconnaîtrais l'air*.

Ce codage permet d'obtenir une première information sur les types de structures en usage dans les IRC, et de procéder à un rapprochement avec d'autres types de textes étudiés.

8.2.1. Relatives

La relativisation est une procédure d'enchâssement syntaxique reconnu dans les langues du monde. Selon Creissels (2000c:11;1) la relativisation est "une construction employée **typiquement** pour restreindre le référent d'un terme nominal d'une unité phrastique (la principale) en lui attribuant une propriété construite en prenant une variable comme terme d'une autre unité phrastique" [306] .

La structure de la relativisation en français standard est postnominale. La relative est enchâssée dans le constituant relativisant de la structure phrastique principale, ce dernier ayant pour tête lexicale un nom auquel succède la relative qui a les caractéristiques syntaxiques d'un modifieur et dont la fonction est de restreindre l'extension du nom.

Une propriété majeure de la relative est l'impossibilité de rétablir dans sa forme et sa position canonique le terme de l'unité phrastique indépendante correspondante. Elle constitue ainsi un trou syntaxique [307] comblé par le relativiseur.

(102) Je connais la personne qui _ parle

L'élément relateur dans la relativisation, que les grammaires traditionnelles nomment pronom relatif, marque la frontière d'une clause enchâssée, il est le représentant du syntagme nominal relativisé et marque la fonction du terme relativisé dans la clause enchâssée, c'est ainsi qu'on lui accorde depuis Chomsky (1968) une double fonction. Les différentes fonctions syntaxiques de constituants nominaux auxquelles accèdent les relatives sont des inventaires variables selon les langues.

Pour le français, la liste des morphèmes relativiseurs se décompose en deux catégories [308] . D'une part *qui, que, quoi, dont, où* et d'autre part les dérivés du pronom relatif – *lequel, laquelle, lesquels, lesquelles, auquel, auxquels, auxquelles, duquel,*… – qui résultent de l'assemblage de l'article défini et de l'interrogatif *quel*. La seconde catégorie varie en genre et en nombre en fonction du syntagme nominal relativisé. La terminologie traditionnelle présente les pronoms relatifs (*i. e. qui, que, quoi, dont, où*) mais les dérivés de *lequel* sont dénommés adjectifs relatifs. Dans nos corpus cette deuxième

[306] C'est nous qui soulignons.

[307] Sag & Wasow (1999:335-354).

[308] Godard (1988:75-84).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

catégorie n'est pas représentée, c'est pourquoi nous n'en discuterons pas plus avant.

Certains pronoms relatifs peuvent être introduits par une préposition, c'est le cas de *qui,* et *quoi.*

(103) La personne à qui je confie mon chat ne répond pas.

dont s'emploie pour représenter syntagme nominal prépositionnel (la préposition est comprise dans sa forme). Notons qu'il est le seul "pronom relatif" qui ne puisse pas s'employer aussi comme "pronom interrogatif".

(104) Je connais la personne dont tu parles.

Après ce premier tour d'horizon, des cas problématiques se présentent en l'absence de pivot de la relativisation. Ainsi, Blanche-Benveniste (1997:94-96) relève une différence de stabilité du /i/ de qui lorsqu'il est pronom relatif sans antécédent, pronom relatif précédé d'une préposition ou interrogatif et lorsqu'il est pronom relatif sujet, ce qui l'amène à considérer deux formes [ki] homographes mais non homophones le premier étant marqué du trait [+ humain] à l'inverse du second.

Dans certains cas, il se présente un substitut du pivot de la relativisation, soit en relation d'anaphore (105) avec un autre élément de l'unité phrastique matrice, soit indéfini (106).

(105) Cette confiture n'est pas celle que nous avons faite hier. (106) Ceux qui l'ont goûtée l'ont aimé.

Le second type (106) présente des structures parfois qualifiées de fausses relatives [309] ou de relatives périphrastiques, n'ayant pas de véritable antécédent.

On trouve encore, dans la littérature, des relatives sans antécédent dénommées substantives pour les différencier de celles qui se trouvent en position de modifieur du nom.

(107) Qui m'aime me suive. (108) Sauve qui peut. Extrait du Corpus P8 *PBA: Qui péri jeune ne pourri

Les cas de pronoms résomptifs tels qu'ils sont identifiés par Chomsky (1987:89-91), sont ceux dans lesquels le terme relativisé apparaît comme dans la phrase indépendante correspondante en relation d'anaphore discursive avec l'antécédent.

(109) J'ai rencontré un mec que quand il te parle tu le comprends pas. (110) Il faut que je lui trouve un disque dont je suis sûr qu'il l'a pas déjà.

Les pronoms résomptifs se rencontrent dans la variété standard [310] dans des relatives de second degré, ou lorsque une unité phrastique subordonnée à la relative éloigne le relativiseur de la trace. Dans ces cas se trouve un décumul des fonctions, le relativiseur ne portant plus que celle de démarcation.

Sont également identifiées comme des cas de relativisation certaines structures participiales [311] . Ces cas ne sont pas reconnus par les grammaires traditionnelles

[309] Riegel *et al.* (1994).

[310] Gadet (1989:115) identifie plusieurs types de relatives résomptives (*i. e*. qui comportent un pronom résomptif *l'homme que j'en parle,* un groupe prépositionnel *l'homme que je parle de lui*, ou un possessif *l'homme que je parle de sa femme*, dans une relative de premier degré). Ces types ne se trouvent pas dans nos corpus.

en vertu de la loi du droit d'auteur.

(111) Un chasseur sachant chasser doit savoir chasser sans son chien.

Du point de vue du rôle joué par les relatives, la littérature recense principalement trois cas : relatives restrictives (déterminatives), explicatives (descriptives, appositives), narratives – les deux premiers étant les plus établis. Ce qui justifie de considérer les relatives restrictives comme prototypiques est le fait que les autres types ne se présentent pas dans toutes les langues.

La définition présentée au début de cette section se résume en deux points : elle revient à considérer comme relatives des structures marquées par un "pronom relatif" qui a pour fonction syntaxique de déterminer l'"antécédent" du "pronom relatif".

Si on examine les exemples suivants, on est forcé de reconnaître que la relative ne vient pas en détermination de l'antécédent :

Extrait du Corpus P2 *PAG: non mais j' y suis revenu en taxi qui m' a fait faire 2 fois le tour du boulevard circulaire parce que lui aussi savait pas aller a l'afnor

(112) Je suis passé sous le pot de fleurs qui m'est tombé dessus au moment où Marie m'appelait.

L'exemple (112) permet trois lectures. La première établit que les événements représentés par les verbes successifs se déroulent dans l'ordre où ils apparaissent dans la phrase. La deuxième suppose que le pot de fleur sous lequel passe le locuteur est déterminé et référentiellement repérable par le fait qu'en un moment passé ce pot de fleur était tombé à un moment où Marie appelait le locuteur. La troisième lecture suppose que l'événement présenté par le premier verbe survient chronologiquement au même moment que l'événement présenté par le dernier verbe et que le pot de fleur est référentiellement repérable par l'allocutaire par le fait qu'il était tombé sur le locuteur en un moment antérieur à celui impliqué pour le premier et le dernier verbe.

Ces différentes lectures ne sont plus possibles dans le cas où l'événement dénoté par la "relative" n'est pas réversible.

(113) J'ai fais une tarte que j'ai mangée à midi. (114) J'ai donné le rôti à Paul qui l'a mangé. (115) Marie a fabriqué ce lustre pour Paul qui l'a cassé.

De même certaines relatives présentent des lectures ambiguës entre détermination et explication :

(116) Le chat qui était sur le piano a renversé le pot de fleur. (117) Le chat que j'ai trouvé ne m'appartient pas. (118) Marie qui est très intelligente a réussi ses examens. (119) J'ai confié les comptes de la société à Mina qui est très forte en comptabilité.

On notera que le cas où la relative s'applique à un nom propre et plus sûrement à un nom propre de personne n'est pas ambigu. En effet, il est rare qu'on ait à restreindre l'extension d'un nom propre [312] par une détermination et si cette éventualité se présente, le nom propre présente un fonctionnement proche de celui d'un nom commun.

(120) J'ai donné ce papier à la Sophie qui est dans ton bureau. (121) Tu parles de Paris ? –Non, le Paris qui est au Texas.

[311] Foley & Van Valin (1985:304) ; Creissels (2000b).

[312] Certains auteurs excluent cette possibilité, cf. Lambrecht (1994:23).

en vertu de la loi du droit d'auteur.

Il convient donc de considérer que ces mécanismes d'intégration de structures phrastiques ne se distinguent pas de ceux relevés pour les relatives restrictives, et reconnaître avec Creissels (1995:316) la relativisation comme "tout mécanisme d'intégration de structures phrastiques qui a comme condition nécessaire l'identification de deux termes nominaux mais qui n'impose pas que les deux termes nominaux concernés occupent chacun une fonction précise dans les deux structures phrastiques donnant lieu au mécanisme d'intégration en question".

Notons que la première seulement des catégories de relatives jusqu'ici envisagées reçoit une caractérisation en terme de fonction syntaxique et que pour les deux types restant ce sont des considérations sémantico-discursives qui permettent de les distinguer.

Les relatives de nos corpus ont été codées en fonctions de ces différents critères, ainsi que des critères établis pour les degrés de subordination et la position de la subordination.

On notera la très faible quantité de relatives dans nos corpus : 288 structures seulement pouvant être analysées comme appartenant à cette catégorie.

L'inventaire du type de pronoms relatifs se trouve restreint dans nos corpus par rapport aux possibilités théoriques.

Tableau (8-7) – Inventaire des pronoms relatifs du corpus

type	nombre
qui	182
que	90
quoi	1
dont	1
où	15

On remarque que près de 95% des emplois présentent *qui* ou *que*. Les trois dernières formes relatives sont très peu fréquentes dans nos corpus, conformément à la tendance que présentent les corpus d'oral.

Les types de relatives que nous avons codés sont au nombre de 10. Il s'agit : du cas prototypique, des relatives sans pivot de relativisation, des relatives sans pivot de la relativisation qui apparaissent dans des structures interrogatives directes, des relatives avec substitut de pivot de relativisation et des relatives intégrées dans des dispositifs syntaxiques. Les dispositifs [313] en question sont les structures pseudoclivées (122) et clivées (123) [314] ; les structures du type *il y a SN qu…* (124), et *avoir SN qu*… (125), les constructions sans verbe (126),

(122) Extrait du Corpus F8 *PDE: Ah bon ce que tu fais là avec nous c'est quoi alors? (123) Extrait du Corpus F5 *PCZ: c'est PAC qui va qque part (124) Extrait du Corpus P2 *PAY: y en a qui sont pas exclusif (125) Extrait du Corpus P4

[313] Koenig & Lambrecht (1999).

[314] Un questionnement sera présenté plus loin à ce sujet.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

(action): P03 tel un poulpe geant a ses bras et ses jambes qui repoussent

Le tableau suivant (8-8) présente les proportions pour chaque type rencontré.

Tableau (8-8) – Inventaire des types de relatives

Type	Proportion (%)
prototypique	44,79
dans une structure clivée	19,79
avec substitut	14,58
il y a N qu…	9,03
sans antécédent	3,82
avoir X qu…	3,13
sans antécédent dans structures interrogatives	1,74
dans une construction sans verbe	1,74
dans une structure pseudo clivée	1,04
Total	100

Les structures prototypiques sont en plus grand nombre, alors qu'on pourrait s'attendre, comme on l'attendrait d'une conversation orale, que les dispositifs syntaxiques absorbent la majorité des emplois.

Quant à la nature de l'antécédent, elle dépend évidemment du type de relative envisagé. Nous avons distingué s'il s'agissait d'un nom commun ou d'un nom propre, d'un nom déictique, d'un substitut, d'une réduction discursive (extrait ci-dessous), etc.

Extrait du Corpus F7 *PBN: PBP pour la pub? fallu jlenleve mais on en remet un mieux scripté et tout cette semaine qui va les classé auto

Tableau (8-9) – Inventaire des types d'antécédents

Type	Proportion (%)
nom	44,79
nom propre	3,82
noms déictiques	19,44
substitut	19,44
sans support	3,82
réduction discursive	3,82
là	1,04

En ce qui concerne la position de l'antécédent, conformément à ce qu'on pouvait attendre, les cas où il est dissocié du relatif sont rares et représentent 6,25%.

La position de l'unité enchâssée par rapport à la phrase matrice est également bien plus occurrente à droite, seuls 14,26% des cas présentent un enchâssement au centre, c'est-à-dire interrompant l'unité phrastique matrice. Un seul cas de position à gauche étant

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

répertorié (relative sans support).

La fonction syntaxique occupée par le syntagme auquel se rattache la relative dans la phrase matrice peut être sujet, objet, attribut, datif, oblique, un syntagme nominal génitival. Les positions d'objet, datif et oblique rassemblent 50,7% des occurrences de relatives, la position sujet n'en comptant que 5,2%. Les autres cas sont les dispositifs syntaxiques, ainsi que les structures sans verbe (15,97%)

La fonction de la trace peut se déduire des formes relatives. Elle est sujet dans 61,7% des cas, objet, ou oblique dans 33,44% des cas.

Quant à la fonction discursive de la relative, elle est le plus souvent déterminative 55,6%, les relatives explicatives sont les moins nombreuses (7,69%) enfin, les narratives sont en nombre assez important (18,18%). Les autres cas correspondent aux dispositifs syntaxiques auxquels les relatives sont intégrées.

8.2.2. Complétives

En ce qui concerne les unités phrastiques subordonnées qualifiées de complétives, on en distinguera deux types : celles qui sont en position d'argument central, et celles qui se trouvent saturer la valence d'un nom. L'inventaire des formes pour ce type de subordination donne *Que* comme vedette. On pourrait avec Attal (1999:15) affirmer que "la seule constante qui s'impose est le rôle de frontière de phrase des mots de la famille de *que*" [315] . Il est en tout cas neutre quant à la sémantique, puisque tout type d'adverbiale subordonnée accepte un conjoint en *que*.

Les complétives objet en *que, si…* [316] ou infinitives sont toutes à droite de l'unité phrastique à laquelle elles s'intègrent.

Les complétives sujets sont plus rares. Celles qui contiennent un marqueur de complémentation sont en plus petit nombre que les infinitives. Il faut noter qu'un nombre important de ces dernières entre dans une structure *ça V de Vinf*. Cela présente une tendance identifiée par Blanche-Benveniste (1997) à l'évitement de certains types de constituants sujets à gauche.

Les complétives qui saturent la valence d'un nom sont peu nombreuses.

8.2.3. Adverbiales

Les adverbiales sont proportionnellement les structures d'intégration phrastique les plus nombreuses dans notre corpus. Il nous faut toutefois préciser ici que certaines unités, souvent considérées comme des subordonnants ne l'ont pas été systématiquement. En effet, *parce que* est généralement recensé comme un subordonnant, et il l'est dans la plupart des cas dans nos corpus. Des énoncés tels que les exemples (127a) et (128a)

[315] Subordonnées relatives, complétives, circonstancielles, marque de la valeur énonciative d'un énoncé.

[316] Le cas particulier des complétives apparaissant dans les "interrogatives indirectes" – Martin (1987:48-52).

en vertu de la loi du droit d'auteur.

s'analysent comme tels.

(127a) Extrait du Corpus F5 *PCS: tu p pas PEN pcq je suis le gardien (128a) Extrait du Corpus P2 *PAO: Ca sent la fraise parce que tu passes après moi =D

La focalisation est en effet possible, si on peut se permettre la transformation sur ces énoncés authentiques en (127b) et (128b) :

(127b) C'est parce que je suis le gardien que tu peux pas. (128b) C'est parce que tu passes après moi que ça sent la fraise.

Toutefois, dans certains cas, *parce que* introduit une unité phrastique qui n'a pas de rapport syntaxique avec une autre unité phrastique. Nous ne parlons là évidemment pas des cas de réponses elliptiques à des questions en *pourquoi*.

Extrait du Corpus F5 *PEF: c une simple question... %add: PEE (2) *PEF: Parce qu'un fichier powerpoint de 300 Mo.. ça s'ouvre pas n'importe où %add: PEE

L'unité phrastique qui devrait être le support de celle introduite par parce que n'apparaît pas, elle pourrait se formuler comme suit :

(129) Je pose cette question parce qu'un fichier powerpoint de 300 Mo.. ça s'ouvre pas n'importe où.

De même, *si* ne s'analyse pas toujours comme subordonnant. Dans l'extrait ci-dessous, PCA propose une reformulation d'un énoncé dans un français approximatif, puis produit les interventions suivantes.

Extrait du Corpus F4 *PCA: je traduis :) %add: PBZ (1) *PCA: enfin si j'ai bien compris :))) %add: PBZ

On se trouve dans une situation similaire à celle ci-dessus, le support pouvant être rétabli :

(130) Ma reformulation est correcte, si j'ai bien compris (l'énoncé).

En ce qui concerne la sémantique de ces unités phrastiques, nous référons à Givòn (1984:827-852). Nous avons distingué ainsi celles qui expriment la condition (et l'hypothèse), le but, la cause, la conséquence, le temps, la manière, la concession.

Dans notre corpus, ces différents types sémantiques se répartissent comme suit :

Tableau (8-10) – Types sémantiques des adverbiales (%)

Type	proportion (%)
but	32,77
cause	9,57
condition	30,627
hypothèse	0,72
manière	2,39
conséquence	0,48
concession	2,87
temps	20,57

Quant à la position des adverbiales, elle est représentée dans le tableau (8-11).

en vertu de la loi du droit d'auteur.

Tableau (8-11) – Types positionnels des adverbiales (%)

Type	Proportion (%)
centre	2,88
droite	61,66
gauche	26,86
sans support	5,51

Les adverbiales, conformément à ce qu'on attendait, manifestent plus de diversité dans leur position. La position à droite est toutefois préférée.

Pour les adverbiales indiquant le but, 95,62% sont à droite. Les adverbiales exprimant la cause, sont préférentiellement enchâssées à droite également, mais à 66,66% seulement, 7,69% se présentant à gauche et 23% s'insérant dans des énoncés réponses ne reprenant pas le verbe de la matrice.

Les temporelles sont préférentiellement enchâssées à droite (59,3%), on relève une proportion toutefois importante de l'enchâssement à gauche (33,72%).

La condition et l'hypothèse sont préférentiellement enchâssées à gauche (60,30% des occurrences). 31,3% sont toutefois enchâssées à droite, et 5,34% s'insèrent dans des énoncés réponse ne reprenant pas le verbe de la matrice.

Enfin, la concession et la manière apparaissent exclusivement à droite.

Ces résultats vont dans le même sens que ceux obtenus pas Koch (1995). Les préférences pour la position d'enchâssement selon les types sémantiques découleraient d'un principe d'iconicité, qui par exemple favoriserait la présentation de la condition avant sa conséquence.

8.2.4. Bilan

Nous noterons que les unités phrastiques enchâssées faisant appel à des participes sont en très petit nombre dans nos corpus, et n'apparaissent ainsi pas dans le détail des observations présentées ici.

En ce qui concerne la répartition des différents types d'intégration, nous pouvons dresser le tableau suivant (8-12).

Tableau (8-12) – Intégration d'unités phrastiques

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Type	Proportion (%)
relatives	19,38
complétives nominales	1,86
adverbiales	44,87
complétives argumentales	33,88

Les unités clausales de notre corpus montrent bien sûr plus de parataxe asyndétique, syndétique et à reprise. Toutefois, un nombre important d'interventions manifeste des phénomènes d'hypotaxe.

En ce qui concerne les degrés d'intégration, on notera que, si on trouve plusieurs exemples d'intégration de niveau 3, il s'agit là d'une limite pour nos corpus. On notera cependant qu'il n'est pas rare de trouver plusieurs unités enchâssées dans une même structure phrastique, ces dernières n'étant pas le plus souvent emboîtées les unes dans les autres, mais occupant diverses positions d'intégration de niveau 1 éventuellement coordonnées. Les intégrations de niveau 2 se trouvent dans un peu moins de 1% des phrases complexes.

8.3. Types de phrases

La grammaire traditionnelle dégage des grands types de phrases : assertive (déclarative), exclamative, injonctive (ou impérative, exhortative), interrogative en se basant sur les buts communicatifs qui les induisent. La distinction des types de phrases se fonde à l'origine sur la notion logique de modalité. Aristote, fondateur de la logique en tant que science formelle, pose que pour qu'un discours soit susceptible de vérité ou de fausseté, une simple juxtaposition de mots ne suffit pas. Pour toucher le vrai ou le faux, il faut affirmer ou nier quelque chose, c'est à dire formuler un jugement. Pour Aristote, toute proposition affirme que quelque chose appartient, appartient nécessairement ou peut appartenir à quelque chose d'autre. "Toute prémisse pose soit une attribution pure, soit une attribution nécessaire, soit une attribution contingente" [317] . Les foncteurs modaux ne qualifient pas la proposition, mais les faits eux-mêmes, l'attribution, la relation. Ces opérateurs caractérisent des modes d'attribution. Notons que l'attribution pure, ou assertoricité n'est, pour Aristote, pas véritablement une modalité, bien qu'étant un mode d'attribution. Elle ne reçoit pas le statut de fonction de modalisation.

Pour Aristote, il y a donc le nécessaire, l'impossible (*nécessairement pas*) et le possible. Ce dernier peut être scindé entre contingent (*pas nécessairement*), et possible (*pas nécessairement pas*). L'assertion simple, non modifiée est dite assertorique et on distingue quatre modes classiques, aristotéliciens [318] , qui donnent lieu à 2 types de

[317] Aristote (1983) *Les premiers analytiques (ORGANON. III.),* Tr. J. Tricot, Paris, Vrin.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

propositions : les propositions apodictiques, et propositions problématiques [319] :

propositions assertoriques [320] :	non modifiées
propositions apodictiques :	nécessaire – impossible
propositions problématiques :	possible – contingent

Aristote [321] a minutieusement étudié les relations logiques entre les quatre foncteurs des modalités propres [322] : possible, impossible, contingent et nécessaire.

La théorie grammaticale s'est centrée sur l'étude des déclaratives, et ce, sans doute à l'origine parce qu'elles sont les seules à pouvoir être déterminées en termes de valeurs de vérité. Cependant pour Bally, le *dictum* (du lat., *le dit*) peut-être énoncé de différentes façons : phrases affirmatives, négatives, interrogatives, impératives, exclamatives. Ces manières de présenter le *dictum* sont ce que Bally appelle le *modus* (du latin, *manière*). Mais considérer la négation au rang des modalités de phrases, ce qui est induit par les origines de l'étude des types de phrases, revient à occulter le fait que la négation peut se combiner avec les autres modalités tandis que les autres ne se combineraient pas toutes entre elles. On aboutit donc à trois types de phrases majeurs qui se confondent avec des familles d'actes de langage [323] .

La phrase déclarative apporte des informations qui se qualifient en termes de valeur de vérité.

La phrase interrogative est indéterminée au regard de la vérité, elle suspend la valeur de vérité d'une assertion sous jacente.

La phrase injonctive vise à provoquer l'accomplissement d'un acte ou d'un événement.

Il s'agit, si nous suivons Givón (2001:318-325) de prototypes syntaxiques qui manifestent un continuum du point de vue fonctionnel [324] .

[318] Au sens d'Aristote, ne sont que trois modes : nécessité, impossibilité, possibilité. Le contingent pose problème. Il existe deux conceptions du possible chez Aristote, qui ont été reprises par les classiques en possible et contingent.

[319] Selon la terminologie de Kant. "Les jugements sont *problématiques*, lorsque l'on admet l'affirmation ou la négation comme simplement possible (arbitraire); *assertoriques*, lorsqu'elle est considérée comme réelle (vraie); apodictiques, quand on la regarde comme *nécessaire*" (Kant, *Critique de la raison pure*, deuxième partie, 1ère division, livre I, chap. 1).

[320] ou attributives.

[321] Aristote (1984) *De l'interprétation (Organon II),* tr. J. Tricot, Paris, Vrin.

[322] Les logiciens font une distinction entre modalités propres et modalités impropres (affirmation, négation).

[323] Austin (1970), Searle (1979).

[324] Par le jeu des actes de langage indirects.

en vertu de la loi du droit d'auteur.

8.3.1. Structures interrogatives

Givón (2001:318) reconnaît ainsi deux types d'interrogatives selon le but communicatif : les questions oui/non dont le but est de faire confirmer ou infirmer la vérité d'une proposition et les questions Wh- dont le but est d'obtenir l'identité d'un terme manquant. Il est pourtant des usages de tournures interrogatives qui ne sont pas des questions, de déclaratives qui ne sont pas des assertions…

Pour ce qui concerne les phrases interrogatives, on distingue traditionnellement les questions totales des questions partielles. Les interrogations totales (131) appellent une réponse oui/non, le contenu propositionnel intégral de l'interrogative est en question.

(131) Tu sais que le québec peut aussi être très ensoleillé ?

On peut les distinguer de celles (132, 133) pour lesquelles se présentent plusieurs assertions sous jacentes mutuellement exclusives [325] .

(132) Tu viens ou tu viens pas ? (133) T'es au boulot ou à la maison ?

Les interrogatives partielles (134) portent sur un élément du contenu propositionnel.

(134) Qui veut parler avec moi ?

Les interrogatives qui portent sur l'intégralité du contenu propositionnel dans les langues sont souvent marquées par la seule intonation. En français, on distingue les questions par intonation (135), par *est-ce que (*136), par inversion du sujet (137) [326] , et, la variante alternative (138) dans laquelle le dernier membre n'est pas explicité, en *ou pas, ou quoi, ou bien, ou non* [327] .

(135) Marie est partie ? (136) Est-ce que Marie est partie ? (137) Marie est-elle partie ? (138) Marie est partie ou quoi ?

Cependant, ainsi que le remarque Givón (2001:293-294) les questions oui/non qui sont sous un processus de focalisation ne sont plus "neutres concernant la source exacte de l'ignorance du locuteur" [328] .

(139) C'est Marie qui est partie ?

Givón (2001:231) remarque encore que "lorsqu'un constituant optionnel est présent, il attire le focus de la question oui/non" [329]

(140) Est-ce que Marie est partie hier ?

Les questions partielles sont construites au moyen de proformes interrogatives

[325] Givón (2001:292-294).

[326] Il s'agit parfois d'une inversion complexe, présentant deux sujets.

[327] Essentiellement dans le parler suisse.

[328] "neutral regarding the exact source of the speaker's ignorance". Dans l'exemple, on présuppose que Marie est effectivement partie, mais la source d'ignorance du locuteur concerne *hier*.

[329] "Once an optional constituent is present, it attracts the focus of the yes/no question".

en vertu de la loi du droit d'auteur.

(cf. similarités répertoriées par la grammaire générative entre le pronom relatif et le pronom interrogatif) ou de combinaisons de déterminants interrogatifs et de noms. En français, les proformes interrogatives peuvent se placer au même endroit que le constituant correspondant dans l'assertion correspondante, dans une position spéciale à la marge gauche de l'unité phrastique. L'interrogation partielle peut être en proie à un procès de focalisation et dans ce cas, la proforme a un traitement identique à celui observé pour un constituant dans une assertion. Notons encore que dans une interrogation partielle, il est possible de rencontrer plusieurs proformes interrogatives. Ce cas (141) est rare dans nos corpus.

(141) Extrait du Corpus P2 *PAP: PAO qui qui es dasn la poche de ki la? :P

En ce qui concerne la formulation des questions, nous verrons selon les paramètres exposés ici quelles sont les tendances dans nos corpus.

Le premier point, trivial, concerne le marquage de la question par un point d'interrogation. Les IRC dispositif basé sur le texte écrit, ne permettent pas le type d'interrogation "par intonation". On suppose donc que le recours au point d'interrogation permettra de désambiguïser les cas qu'aucune autre marque ne présente comme interrogation (interrogatif, inversion…).

En effet, les interrogatives non marquées par un point d'interrogation représentent seulement 17,63%. 50% exactement de ces interrogatives présentent dans leur structure au moins un élément permettant de les interpréter comme des questions, la ponctuation se trouve donc redondante dans ces cas. Les autres structures interrogatives sans ponctuation, ne sont donc interprétées comme telles qu'en fonction du contexte, des connaissances extralinguistiques, des inférences sur les intentions d'autrui.

(142) Extrait du Corpus F1 *PBW: tu vas bien belle PAY ;))))))))))))) %add: PAY *PAY: PBW pas mal et toi?? :) %add: PBW

Les questions alternatives représentent moins de 2% de l'ensemble. Les questions partielles représentent 31,8% tandis que les questions totales représentent 66,3%.

Les questions partielles qui comportent un déterminant interrogatif (*quel* x) forment 5,75%, celles qui comportent un pronom interrogatif 88,73%. Les autres questions partielles (5,51%) combinent un pronom interrogatif et "*est-ce que*" (143).

(143) Extrait du Corpus F8 *PDV: PDT bon ben j'attends! kes tu me racontes?

Les interrogations sans verbe (144) représentent 25% de l'ensemble et les interrogatives complexes (145) 7,76%.

(144) Extrait du Corpus P8 *PAN: tiens j ai une question %add: PBB *PBB: laquel? %add: PAN (145) Extrait du Corpus P2 *PAO: Ou t'as vu que j'étais ton pote toi?

Les inversions concernent 5,35% des interrogatives. Cela se produit sur des questions partielles à 65,75% et sur des questions totales à 35,61%.

Les structures interrogatives les plus fréquentes sont les questions totales se présentant comme des déclaratives (24,46% – exemple 146) et les questions sans verbe (16,41% – exemple 147), conformément à la tendance présentée par l'oral spontané. [330]

[330] Gadet (1989), Blanche-Benveniste (1997).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

(146) Extrait du Corpus F7 *PBP: on mets une limite a 2 PJQ :o)) (147) Extrait du Corpus F3 *PCO: au boulot ou a la maison %add: PBU (1) *PBU: a la maison %add: PCO

Un petit nombre de structures non standard (148) se trouve dans les corpus.

(148) Extrait du Corpus P8 *PAN: qui qui le veu mon program

On repère également quelques occurrences de variétés d'interrogatives fréquentes au Québec, ainsi que l'illustre l'extrait suivant.

(149) Extrait du Corpus P2 *PCM: bon je peux tu fumer une clope la pfffffffffff

Pour ce qui est des structures interrogatives, la tendance est donc similaire à celle rencontrée pour l'oral telle que l'a définie Gadet (1989).

8.3.2. Structures injonctives

Les structures injonctives dans nos corpus se conforment au modèle classique. On en compte 400.

Extrait du Corpus F4 *PAJ: attendez je mets la musique Extrait du Corpus P7 *PAL: boude pas va %add: PBB

Une singularité cependant concerne les énoncés produits pour déclencher des actions des opérateurs de canaux sur des participants indisciplinés. Il s'agit d'énoncés sans verbe. Ces énoncés sont pour la plupart très rudimentaires et sont composés comme suit :

Surnom_du_participant_indiscipliné nom_de_l'action_réprouvée

Il s'agit donc de "dénonciations". Ces énoncés se distinguent des autres par le fait qu'ils ne s'intègrent à aucune unité conversationnelle et surviennent à tout moment, sans précision supplémentaire, et sont interprétées par les opérateurs comme des demandes d'interventions de type kick.

Extrait du Corpus F6 *PH5: P97 pub de merde Extrait du Corpus F1 *PAD: P99 DCC

Il ne s'agit donc pas d'injonction au sens syntaxique.

Il se produit qu'un même locuteur formule plusieurs dénonciations à la suite et dans ce cas des mécanismes de réduction discursive apparaissent.

Extrait du Corpus F6 *PCN: P99 pub *PCN: P98 idem Extrait du Corpus F7 *PBD: P99 spam (1) *PBD: P98 aussi

8.3.3. Structures exclamatives

Le statut mal défini de ce type de phrase, ses similarités tant avec les structures déclaratives que les structures interrogatives semble former un continuum entre les types de phrases [331] . Les critères de reconnaissance de cette catégorie sont divers tant à l'oral qu'à l'écrit. Ces énoncés sont dits exprimer en plus du contenu propositionnel une affectivité, un sentiment du locuteur à l'égard de son énoncé. Il s'agirait donc d'un type

[331] Voir toutefois les réflexions de Martin (1987) à propos des limites du parallèle avec l'interrogative.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

mettant en jeu l'émotivité du locuteur, et force est de reconnaître que le domaine est vaste et difficilement discrétisable. Toutefois, certaines structures sans verbe et comportant à l'initiale *que* ou *quel (*variable en genre et en nombre) peuvent être qualifiées de structures exclamatives prototypiques. Ces structures sont peu présentes dans notre corpus. La signification de l'état émotif (surprise, admirativité…) du locuteur est préférentiellement portée par d'autres marques (ponctuations déviantes, interjections, lexique, orthographe déviante).

8.3.4. Structures déclaratives

Les structures déclaratives sont les plus nombreuses dans nos corpus. Elles ne répondent pourtant pas, pour la plupart au cas canonique présentant un sujet, un verbe éventuellement suivi d'un objet. Un certain nombre de dispositifs syntaxiques sont à l'œuvre dans nos corpus, que nous allons examiner successivement.

8.3.4.1. Focalisation

Les procédés de focalisation sont réputés avoir la fonction de mettre en relief la valeur informative d'un élément de l'unité phrastique. On rattache aussi à ce phénomène ceux qui manifestent une mise en relief de la valeur assertive de la phrase ou d'autres éléments de sens.

CONSTRUCTIONS PSEUDO CLIVÉES ET CLIVÉES

L'identification de telles constructions est issu de la linguistique anglo-saxonne qui utilise pour les dénommer les termes *cleft sentences* et *pseudo-cleft sentences*. Ces constructions seraient caractéristiques de l'oral spontané [332] .

Les constructions dites pseudo-clivées utilisent une prédication d'identification assortie d'une détermination.

(150) Celui que j'aime pas c'est Paul. (151) Celui à qui je parlerai pas c'est Paul. (152) Celle dont nous parlons c'est Anne.

Cette construction présente son premier membre en position de thème, le second prenant un statut de rhème (information nouvelle, ou focalisée).

Dans ces constructions entre parfois dans nos corpus *il y a,* utilisé comme verbe d'existence, sur le modèle de l'exemple suivant :

Extrait du Corpus F6 *PEP: ce qu'il y a de bien avec le thé chaud %add: ALL (1) *PEP: c'est que ça brûle %add: ALL Extrait du Corpus F6 *PAW: mais PDH... keskiya? %add: PDH *PDH: ya que je supporte plus ce nick %add: PAW

Blanche-Benveniste (1997:96-100) remarque que les constructions pseudo-clivées autorisent un branchement aisé des compléments et sujets en *que* suivi d'une construction verbale. C'est effectivement le cas dans nos corpus, ainsi que l'illustre l'exemple suivant dans lequel une unité phrastique intégrée en fonction de sujet introduite par *que* appelant un subjonctif est évitée :

[332] Blanche-Benveniste (1997).

en vertu de la loi du droit d'auteur.

Extrait du Corpus F8 *PDY: PDT d'accord, mais ce qui est marrant c'est que personne peut dire exactement pourquoi il m'en veut

Ce type de construction est toutefois peu représenté dans notre corpus.

Le procédé de clivage consiste également en l'extraction d'un terme de la phrase pour le placer dans une position rhématique. Cette construction connaît des restrictions sur le type d'élément extractible, ainsi que l'illustrent les exemples suivants pour l'extraction du sujet (153), extraction de l'objet (154), extraction d'une subordonnée (155), extraction d'oblique (156), extraction d'adverbe (157, 158, 159).

(153) C'est Pierre qui aime la confiture. (154) C'est la confiture que Pierre aime. (155) C'est tant que sa mère ne l'appelle pas qu'elle reste dans la cour. (156) C'est dans la cour qu'elle reste tant que sa mère ne l'appelle pas. (157) ?C'est vraiment pas qu'elle aime le saumon. (158) C'est injustement qu'il a pris cette décision. (159) C'est franchement qu'elle est bête.

On reconnaît que cette structure (160) présente une tendance pour la déviance par rapport au système d'accord.

(160) C'est les jeunes qu'elle aime pas.

Cependant, le cas ne se présente pas dans nos corpus, sans doute à cause du petit nombre d'occurrences de la construction ; les structures présentatives simples avec ou sans détachement en revanche sont des occurrences de ce phénomène.

Il faut noter que la forme canonique de la construction pseudo-clivée, s'illustrant dans l'énoncé *ce que j'aime pas c'est la confiture*, dans laquelle *que j'aime pas* peut parfaitement s'analyser comme relative, trouve un correspondant dans la construction dite clivée dans laquelle *que j'aime pas* semble difficile à analyser comme relative dans la mesure où le relativisant ne présente pas un fonctionnement identique à celui dégagé pour la relative (cf. supra) :

(161) C'est Paul que j'aime pas. (162) C'est à Paul que je parlerai pas [333] . (163) C'est d'Anne que nous parlons.

Ainsi la focalisation du datif s'oppose à la relativisation du datif.

(164) Je regarde Paul à qui je parlerai pas.

Ces structures, quoique plus fréquentes que les structures pseudo-clivées, ne sont pas très utilisées dans nos corpus. On relève quelques rares cas s'écartant de la variété standard.

Extrait du Corpus F8 *PAV: mais tu n' es pas con... et tu ne fous pas la merde... c nous qu'on comprend rien :(

Les structures présentatives avec détachement se présentent plus fréquemment.

Extrait du Corpus F6 *P66: ba oui mais coser tout seul c pas fun

8.3.4.2. Thématisation

La terminologie qui concerne ce domaine syntaxique est multiple. On parle de thème en l'opposant à rhème pour désigner ce qui constitue l'objet d'un acte d'énonciation, le rhème (ou propos) présentant ce qui est exprimé à son sujet. Les termes viennent du grec

[333] ***Dans des états de langue antérieurs, "C'est Pierre à qui je pense" aurait été la norme. Attal (1999:170).***

en vertu de la loi du droit d'auteur.

□□□□ ce qu'on pose ou dépose [334] et qui dénote en grammaire le sujet, et □□□□ le verbe, qui induisent immédiatement la confusion avec un autre couple d'opposition : sujet prédicat. L'ambiguïté peut être évitée en utilisant topique/commentaire, information ancienne/nouvelle. Cette opposition ne doit pas non plus être confondue avec l'opposition *modus*/*dictum* qui chez Bally, relève d'une opposition entre ce qui dans les énoncés permet la référence (*dictum* du latin, *le dit*) et les manières de présenter le *dictum* (*modus* du latin, *manière*). Le *dictum* est la mise en rapport d'un sujet avec un prédicat, le contenu représentatif de l'énoncé, tandis que le *modus* est l'attitude prise par le sujet parlant. Il est souvent difficile de décider pour une phrase, ce qui relève du *modus* et ce qui relève du *dictum*. La relation entre le posé et le dit, entre un thème et un propos, appelée "lien informatif" [335] , serait le garant de la complétude de la phrase. La modalité peut se concevoir comme la manière dont le locuteur envisage la relation thème/propos.

L'attitude du sujet parlant peut-être considérée comme la manière de présenter l'attribution d'un prédicat à un objet. On pourrait classer ces modes d'attribution à la manière des logiciens, selon que l'attribution est présentée comme un fait (jugement catégorique), comme une possibilité (jugement hypothétique), comme une nécessité (jugement apodictique). La distinction *modus/dictum* conduit à donner des statuts différents aux éléments phrastiques, certains appartenant au *modus*, d'autres au *dictum*. On pourra alors se demander si la modalité porte sur l'ensemble de l'énoncé, sur l'énonciation ou sur une partie de l'énoncé. Il faudra aussi chercher ce qui détermine la portée d'un opérateur modal et s'il y a une relation entre portée sémantique et portée syntaxique.

La mise en thème dans les langues qui le permettent est préférentiellement marquée dans l'ordre des mots (topicalisation [336]). La littérature relève aussi le passif [337] comme stratégie de mise en thème. Cependant en français la structure thématique la plus fréquente à l'oral, le jeu sur l'ordre étant contraint par la langue et le passif peu employé [338] , est le mécanisme de la dislocation, illustré par l'exemple (165) ci-dessous.

(165) Mon père$_i$ à son frère$_j$ il$_i$ la$_k$ lui$_j$ a donnée la moto$_k$ [339] .

Selon Keenan (1977) ce type de construction serait rare dans les productions planifiées et donc caractéristique de l'oral non planifié (spontané).

334 Picoche & Firmin (1992).

335 Dupont (1983).

336 Parmi les "fronting constructions".

337 La fonction de topicalisation du second argument attribuée au passif par la littérature est remise en cause par des exemples mis en avant par Gaatone (1998:211) du type : "Comme président, je serais un homme qui unira toute la population, car j'aurais été élu par tous".

338 Les tournures passives sont également rares dans nos corpus.

339 Exemple de Norbert Dupont, Syntaxe, Cours de Licence de l'université Lumière Lyon2.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Certains dispositifs syntaxiques sont identifiés également comme des procédés de topicalisation de référents qui n'ont pas déjà été introduit dans le discours. Ainsi Lambrecht (2000) caractérise-t-il la construction *Il y a N qui…* ou *j'ai X qui…* On ne peut en effet pas utiliser ces constructions pour parler d'un référent qui vient d'être mentionné.

(166) Et ta nouvelle voiture ? *Il y a ma nouvelle voiture qui est en panne. *J'ai ma nouvelle voiture qui est en panne [340] ***.***

Selon Creissels (2000b:18;4) "Dans l'alignement prototypique des rôles, le sujet est le constituant nominal de l'unité phrastique qui a le degré le plus élevé de topicalité. Par conséquent, dans une langue où se développe une tendance à marquer la topicalisation d'un constituant presque dans chaque énoncé (comme en français parlé ou en japonais), il est prévisible que les sujets vont tendre à être topicalisés en quelque sorte par défaut chaque fois que le contexte discursif n'impose pas la topicalisation d'un autre constituant." Alors, la fréquence de la construction lui fait perdre son statut de construction marquée et, cela pourrait avoir pour conséquence selon Creissels (2000b:18;4) la "transformation d'un indice de sujet initialement utilisé pour représenter un constituant détaché dans une construction disloquée en un élément obligatoire du mot verbal qui marque simplement l'accord entre sujet et verbe"

Les syntagmes nominaux initiaux non sujets sont dits topiques externes selon Foley & Van Valin (1985:300). Ces auteurs différencient les constructions topicalisation et dislocation à gauche par le fait que la seconde (167) présente un élément pronominal qui réfère au topique externe tandis que la première (167, 168)présente un trou syntaxique.

(167) Cette fille je m'entends pas avec _. (168) Extrait du Corpus F1 *PCJ: l' humour au 2eme degré y'en a qui connaissent pas ptdr

Selon Foley & Van Valin (1985:300)

"Les topiques ne sont pas des constituants de la proposition mais lui sont plutôt externes, en juxtaposition à la clause entière […] ils ne portent pas nécessairement de relation sémantique avec le prédicat ou ses arguments". [341]

Ce phénomène s'illustre par l'énoncé suivant.

(169) Ma femme, son père il est banquier.

On notera cependant des objections à la caractérisation de la construction dilocation à gauche comme un simple procédé de topicalisation. Prince (1992, 1998) [342] propose pour l'anglais que la construction est corrélée à trois fonctions : simplifier le traitement du discours, déclencher un jeu d'inférences partiellement ordonnées, ou amnistier les violations d'îles [343] .

Notre corpus contient, nous l'avons évoqué au chapitre 7, un grand nombre de noms

[340] ***Ici l'astérisque ne vaut que pour l'impossibilité d'enchaînement.***

[341] ***"Topics are not constituent of the clause but rather are external to it, in juxtaposition to the clause as a whole […] They do not necessarily bear any semantic relationship to the predicate or its arguments".***

[342] Voir aussi Ward & Prince (1991).

[343] *island condition,* Ross (1967) cité par Prince (1991).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

propres de personnes qui pour la plupart entrent dans des fonctions de vocatifs. Les vocatifs permettent de pallier le caractère décontextualisé du dispositif de communication. Les locuteurs n'ont de traces perceptibles de l'ensemble des locuteurs présents que la liste des participants (liste de surnoms) et n'ont trace de la participation de chacun que depuis le moment où ils sont entrés dans le canal. Enfin, les locuteurs se manifestent aux autres locuteurs seulement par leur dire, auquel est attaché leur surnom. Reconnaître l'auteur d'un énoncé, ainsi que nous l'avons vu au chapitre 6 n'est pas un problème. En revanche, la sélection de l'allocutaire est problématique dans le cas où plusieurs locuteurs sont engagés dans la conversation et dans le cas où plusieurs conversations se mènent en parallèles, ainsi que l'illustre l'exemple suivant (déjà présenté au chapitre 6) dans lequel deux conversations se mènent en parallèle, l'une à propos d'une effraction sur véhicule et l'autre à propos de cours. Un locuteur n'a pas spécifié le destinataire de son message immédiatement (mis en relief en gras dans l'exemple) mais dans le message suivant ce qui donne lieu à confusion [344] sur le destinataire :

Extrait du Corpus F7 *PCJ: j'peux meme pas, j'ai cours, deja que je vais partir en plein milieu... %add: PFN *PDS: fallaiçs mettre un system linux embarqué pour l'alarme %add: PCK *PCK: heureusement que j'avais une alarme %add: PDS *PCK: qui les a dérangé... ils ont rien pris %add: PDS *PDS: ah oki %add: PCK *PDS: tu les as butté %add: PCK *PIB: chanceux va %add: PCJ *PIB: PCJ %add: PCJ *PDS: pas changed18nceux... intelligent %add: PIB *PCK: mais j'ai monté mon ampli+hp a l'appart %add: PDS *PCJ: PIB tu trouves vraiment que j'ai e la chance? %add: PIB

C'est ce qui explique selon nous en grande partie l'emploi de noms propres de personnes massivement en fonction de vocatifs. La syntaxe et la sémantique des vocatifs ont été très largement négligées à la fois par les grammaires traditionnelles et par les linguistiques contemporaines. Il est entendu ici que le terme de vocatif ne se rapporte pas à un cas morphologique formel tel qu'on le définit pour le grec ancien mais à un type particulier de constituant d'unité phrastique pour lequel Bally [345] remarquait qu'il apparaît "en marge de la phrase ou de la proposition grammaticale". Lambrecht (1998) fait l'hypothèse que ce manque d'intérêt provient d'une part du statut de non argument dont ils sont assortis dans l'énoncé et d'autre part de leur caractère déictique. On ne réfère pas ici aux épithètes et interjections dont Milner (1978) et Ruwet (1982) ont dégagé la syntaxe, mais aux constituants qui permettent, en première analyse, d'attribuer le statut de destinataire direct d'un message à un interlocuteur.

Par certaines propriétés, les vocatifs ressemblent aux constructions topiques ainsi que le remarquent Lambrecht (1994, 1996, 1998) et Gregory & Michaelis (2001). Cependant certains auteurs les excluent explicitement de la construction dislocation. En effet, Prince (1998:282) indique que la "dislocation à gauche […] dénote une forme syntaxique dans laquelle certains syntagmes nominaux (non vocatifs) apparaissent en position initiale pré-clausale, coréférentiels avec un pronom personnel apparaissant quelque part dans la clause" [346] . On peut ainsi distinguer dislocation gauche, dislocation

[344] Nous n'affirmons pas là que l'interprétation a été erronée, l'ambiguïté a également pu être saisie volontairement.

[345] Bally (1985) cité par Lambrecht (1998:37).

en vertu de la loi du droit d'auteur.

droite, et détachement à gauche ou à droite.

Conformément à l'analyse proposée par Lambrecht (1998), nous parlerons de constructions TOP et A-TOP [347] , de constructions VOC et A-VOC respectivement pour les constructions topique gauche, topique droit, vocatif gauche et vocatif droit. Selon Lambrecht (1998:35) ces deux constructions servent à "établir une relation de pertinence entre un élément de discours et une proposition."

Lambrecht (1998:35) affirme que la forme de ces constructions répond à un principe cognitif qu'il appelle "principe de séparation de la référence et de la relation". Ainsi, la langue parlée présenterait une tendance à présenter "l'information nécessaire à l'établissement d'un référent topical dans un discours indépendamment de l'information pertinente concernant ce référent".

Dans le corpus, 3034 noms propres sont utilisés comme vocatifs. La position à droite ou à gauche est à peu près également répartie sur le corpus (50,7% apparaissent à la droite de l'énoncé). 89,39% des vocatifs apparaissant à droite ne présentent aucun lien indiciel avec l'énoncé, tandis que la proportion des vocatifs gauches pour ce critère est de 68,52%.

Les vocatifs qui apparaissent sous d'autres formes que le nom propre ne sont que très peu représentés dans nos corpus. Il peut aussi s'agir d'un nom commun qui comme le nom propre est éventuellement assorti d'un adjectif et parfois d'un déterminant. En position de vocatif on trouve également des formes de pronom disjoint (*toi, vous*), des titres éventuellement associés à des syntagmes nominaux. Dans nos corpus, ces derniers ne sont pas représentés et les formes autres que les noms propres sont bien moins fréquentes.

Nous avons marqué de façons différentes dans nos corpus les constructions impliquant des vocatifs (170), et celles impliquant des topiques (171).

(170) Extrait du Corpus F2 *PAD: toutafait d'accord PBT :o) (171) Extrait du Corpus F7 *PDS: la saxo vts elle tue

Nous avons également différencié les cas de vocatifs présentant un lien avec un élément de la structure phrastique à laquelle ils sont associés (172) et ceux n'en présentant pas (173).

(172) Extrait du Corpus F3 *PAJ: PEQ tu me fais iech (173) Extrait du Corpus P3 *PAM: PAF jai pas un coprs de deesse.. tu doit te tromper :P

On distinguera deux types de structures comportant une dislocation : celles qui n'en comportent qu'une seule, et celles qui présentent plusieurs éléments disloqués (174).

(174) Extrait du Corpus F3 *PAJ: PDL t'es ki toi?

Nous avons également codé la position des éléments en question par rapport à l'unité phrastique à laquelle ils sont associés : à droite, ou à gauche. D'autres positions peuvent apparaître pour les vocatifs (entre le verbe et l'objet), cependant, nos corpus ne les

346 "left dislocation […] denotes a syntactic form in which some (nonvocative) NP appears in initial pre-clausal position, coreferential with a personal pronoun occuring somewhere in the clause".

347 Anti-topique.

en vertu de la loi du droit d'auteur.

manifestent pas.

La position syntaxique qui connaît le plus de dislocations est celle de sujet, dans tous les types, à plus de 80%. Ce résultat confirme bien l'hypothèse avancée par Creissels évoquée plus haut, et présente donc nos corpus comme utilisant des structures d'une manière proche de celle de l'oral.

Si on compare la proportion de dislocation à gauche par rapport à la dislocation à droite pour chacun des deux types (avec ou sans vocatif), on remarque qu'il se présente une préférence pour la dislocation à gauche, mais dans des proportions différentes. Les disloquées comportant des vocatifs ont une tendance bien plus forte pour la position à gauche.

Dans le cas le plus fréquent, un seul élément est disloqué. Toutefois, 6,43% présentent des combinaisons. Il se dégage deux types préférentiels qui sont l'apparition d'un vocatif nom propre à gauche, doublé d'un sujet à droite et l'apparition d'un vocatif pronominal sujet à gauche, doublé d'un vocatif nom propre sujet à gauche. Les contraintes positionnelles ne peuvent être plus finement étudiées.

En ce qui concerne la position droite ou gauche dans nos corpus, elle semble liée au fait que le référent du terme disloqué est plus ou moins facilement identifiable par le destinataire.

8.4. Autres phénomènes syntaxiques et morphosyntaxiques

8.4.1. Absence de marqueur d'enchâssement

Certaines structures qui dans la variété standard écrite apparaissent avec un marqueur morphologique, en font l'économie dans nos corpus. Les exemples (175) et (176) illustrent ce fait.

(175) Extrait du Corpus P8 *PBA: Plutot crever de perdre mes cheveux =D [348] ***(176) Extrait du Corpus F7 *PDA: PBF j ai une date avec ta belle ;) je te promets je fais attention a ce qu elle mange ;)***

Il faut reconnaître que l'exemple (175) est d'un type peu fréquent et est sans doute le fait d'une erreur. En revanche, l'exemple (176) est d'un type fréquent. Dans ce cas, l'apparition d'un deuxième indice de sujet ne rend pas strictement nécessaire le marquage de la complétion. Un phénomène similaire s'identifie certaines structures injonctives négatives telles qu'elles paraissent à l'oral. *t'inquiète* paraît pour *ne t'inquiète pas,* or le fait que l'énoncé commence par le clitique induit qu'il s'agit d'une négation puisque la structure positive place le verbe en premier *inquiète toi*.

[348] ***avec la signification je préfère X plutôt que Y***

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

D'assez nombreux autres exemples de ce type paraissent dans nos corpus :

Extrait du Corpus P3 *PAM: PAS ouais ! :) mais la il veut jle laisse dormir :/ Extrait du Corpus P6 *PA2: sa fait pres de 2 ans je suis aop ici moi non mais bordel

Toutefois ces phénomènes sont assez rares si on les rapporte à l'ensemble du corpus. On peut constater que contrairement aux fautes d'orthographe ils ne soulèvent aucun commentaire métalinguistique. Ils sont parfaitement ignorés.

Dans certains cas, on est en droit de se demander s'il s'agit de véritables omissions de simplification du matériel signifiant ou d'une innovation. L'interprétation de ces faits qui consisterait à évoquer un phénomène de parataxe est fausse.

(177) Il veut (178) Je te promets

(177) et (178) sont des énoncés qui ne peuvent être reconnus comme complets sans invoquer une réduction discursive.

Avec certains verbes (*penser, croire, espérer*), un phénomène similaire se produit avec détachement de ce qui serait la portion de phrase matrice dans une structure standard.

Extrait du Corpus F2 *PAL: elle a memote les sops je crois Extrait du Corpus P2 [349] *PAP: PAO lol tu parle pas dememe a ta blondasse jesper e :P

8.4.2. Spécificité du marquage du discours rapporté

Le discours rapporté dans nos corpus se présente bien entendu de la manière traditionnelle, ainsi que le montre les extraits ci-dessous.

Extrait du Corpus F3 *PCA: y dit ki parle pas français Extrait du Corpus F4 *PAT mon fils en parlait encore ce matin lol (3) *PAT "Maman..... vous aviez l'air de deux capotées hier !!!" Extrait du Corpus F4 *PAT je lui ai dit...."euh non non.... c'est naturel" lololol

Cependant les mécanismes de discours rapporté direct et indirect standards, ainsi que nous l'avons signalé au chapitre 5 ne se présentent presque que pour des propos qui se sont tenus hors des IRC ou lors de sessions antérieures. Lorsqu'un locuteur énonce les paroles d'un autre locuteur présent dans la conversation, il utilise systématiquement une construction qui prend la forme suivante :

Y: X "parole de X" parole de Y

En voici quelques exemples :

Extrait du Corpus F4 *PCD (P99): di leur de me enlever lakick just kelke minute pour ke je vienne mettre leur parler! Extrait du Corpus F8 *PDV <PDT> moi PDV... je ne crois pas.... je sais ;) d'où la question

Dans l'extrait du corpus F4, l'intervention entière est une citation, tandis que dans l'extrait du corpus F8, il faut lire : PDT a dit "moi PDV... je ne crois pas.... je sais ;)" et PDV ajoute "d'où la question".

Bien évidemment, des variantes se présentent, ainsi qu'on peut le voir dans les extraits du corpus F8.

[349] ***On notera l'emploi de de même caractéristiquement plus fréquent dans la variété québécoise.***

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Extrait du Corpus F8 *PAV: <PEK> donne moi un voice <------ encore un qui croit que je suis à son service Extrait du Corpus F8 *PDV <PDY> PDV ben je te connais ici, et je sais que c'est un truc que tu fais souvent <--- je ne sais pas si tu cherches une dispute ou une conversation, mais je ne fais pas ça ! et je t'avoues un truc, tu m'enerves

Souvent, rien ne signale X comme auteur de "parole de X" ni la démarcation entre "parole de X" et "parole de Y".

8.5. Conclusion

Ce chapitre nous a permis de qualifier la syntaxe dans les IRC au niveau des structures phrastiques. Les structures complexes sont reconnues comme caractéristiques du type de texte. Nous avons observé, évidemment, dans nos corpus que les structures complexes sont en nombre restreint par rapport à ce qu'on pourrait attendre d'un corpus écrit prototypique. En ce qui concerne ces structures, d'un point de vue qualitatif, nous avons pu remarquer également que l'inventaire des morphèmes subordonnant qu'il se conforme à ce qu'on rencontre pour l'oral en face à face (cf. inventaire des pronoms relatifs, des conjonctions et locutions conjonctives). De même, les places syntaxiques des structures enchâssées respectent les tendances rencontrées pour l'oral (peu de complétives nominales ; peu de complétives sujet ; proportion importante d'adverbiales ; etc.). La tendance pour les positions syntaxiques des structures enchâssées est également similaire à celle de l'oral (peu de structures enchâssées provoquent une interruption de la phrase matrice ; préférences d'enchâssement à droite ou à gauche en fonction des types sémantiques).

Au niveau des types de phrases, nous avons remarqué qu'ils sont tous présents, dans des proportions différentes mais proches de celles rencontrées pour l'oral en face à face, les particularités se portant essentiellement sur les interrogatives et les déclaratives. La question qui se pose à cet égard est de savoir quelles sont les stratégies mises en place pour pallier la contrainte du canal/média. Les locuteurs emploient de nombreuses structures typiques de l'oral et habituellement reconnues pour être soutenues par la prosodie (notamment les structures sans marqueur d'enchâssement).

L'examen des structures syntaxiques phrastiques montre ainsi que la syntaxe dans les IRC est fortement similaire à celle de l'oral, malgré la différence de chenal et donc l'impossibilité de recourir à des moyens prosodiques. Évidemment, la fréquence des structures sans verbe renforce cette conclusion.Les structures fréquentes sont en effet soit également fréquentes à l'oral, soit typiques de l'oral. Nous pensons que l'interprétation des énoncés dans les IRC requiert plus de recours aux connaissances extralinguistiques (contexte, situation, inférences sur les intentions d'autrui…) que les autres interactions verbales parlées. On note sporadiquement des structures qu'on ne s'attendrait pas à trouver dans des textes spontanés, mais nous les expliquons par le fait que les IRC permettent d'échapper à la contrainte temporelle et donc quelquefois, de planifier le discours.

en vertu de la loi du droit d'auteur.

en vertu de la loi du droit d'auteur.

Chapitre 9 Temporalité

Les chapitres précédents ont tenté de mettre en perspective les spécificités des productions linguistiques dans les IRC, du point de vue du code (Chapitre5), du point de vue de l'échange (Chapitre 6), du point de vue du lexique (Chapitre 7) puis du point de vue de la syntaxe des énoncés (Chapitre8).

Nous avons vu que ces domaines sont différemment touchés. Le présent chapitre tentera de montrer comment se met en place la référence temporelle dans les corpus. En effet, parmi les éléments linguistiques qui se distribuent entre les différents types de textes se présente l'utilisation des systèmes de TAM. Cette dernière serait liée à la deixis.

Notre objectif est donc ici d'examiner si ce versant des messages linguistiques est atteint par la différence de dispositif.

Nous consacrerons quelques pages à une réflexion sur le concept de temps, puis nous examinerons comment la temporalité se manifeste dans nos corpus. Enfin, nous présenterons quelques types particuliers d'unités conversationnelles de nos corpus.

9.1. Le concept de temps

La démarche que nous allons adopter dans cette section, avant de nous attaquer à la matérialité linguistique proprement dite, ressemble à un exercice de réflexion sur les

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

notions qui seront essentielles à nos analyses. En effet, avant d'aborder la question de la temporalité linguistique dans les corpus, il a paru qu'un questionnement raisonné sur le concept de temps était indispensable.

La temporalité dans le langage peut en effet être envisagée de nombreux points de vue. La littérature linguistique en témoigne par son extrême abondance, et sa diversité. Si *temporalité* est entendu comme expression du temps, il nous faut préalablement définir ce qui relève du temps. Ce qui est en question ici, ce n'est pas simplement la polysémie si fréquemment mentionnée de ce terme en français. Nous nous interrogeons sur la nature du temps, ce que c'est que le temps afin de mieux cerner ce qui se joue pour le temps dans le langage et les langues. Bien sûr, il ne s'agira pas d'une exploration rigoureuse des théories physiques ou philosophiques du temps, néanmoins nous aborderons les problématiques essentielles concernant le temps.

Cette démarche pose une question cruciale quant aux rapports entre le langage et la pensée. Faut-il commencer l'épreuve par une réflexion quasi-philosophique sur le temps ou doit-on partir des phénomènes linguistiques ? Nous pensons, au premier abord, que ces deux approches sont complémentaires. Il nous semble illusoire de penser l'une sans penser l'autre. Sur quelle base en effet sélectionner les phénomènes linguistiques sans avoir déterminé ce qui est en question, et comment, à l'inverse, déterminer ce qui dans notre pensée est dicté par la langue, ou par le langage ? Ce problème est bien sûr ancien, et donne lieu à de vifs débats. La pensée est conçue selon deux perspectives antagonistes, entre lesquelles on trouve, bien sûr, des positions plus nuancées : pensée comme masse amorphe avant qu'elle ne soit structurée par le langage ou structurée de façon innée pour permettre le langage.

De nombreux chercheurs, s'inspirant du modèle computationnel dans lequel l'ordinateur fait figure de métaphore du cerveau, ont postulé l'existence d'un "langage de la pensée", un *mentalese* selon Fodor (1983). Cette conception postule de façon analogue à ce qui se passe pour l'ordinateur un langage machine : le "langage mental" support de la pensée. Lakoff (1987:337) résume, de plus, que "Selon cette perspective, l'esprit est un ordinateur avec un matériel biologique et fonctionne en utilisant des programmes essentiellement similaires à ceux utilisés dans les ordinateurs aujourd'hui" [350] .

La pensée est alors une question de manipulation de symboles et toutes les activités cognitives sont conçues, dans ce cadre, comme algorithmiques. La raison humaine est alors abstraite et indépendante de l'expérience. Les concepts sont des représentations internes de réalités externes. Les symboles manipulés reçoivent leur signification seulement de leur capacité à correspondre à des aspects de la réalité.

Beaucoup de débats sont générés par cette question de l'interaction entre langage et cognition. À la suite de Chomsky, nombre de linguistes postulent l'autonomie du langage, en se basant par exemple sur les dissociations découvertes entre les capacités grammaticales et diverses capacités cognitives non-linguistiques. D'autres linguistes

[350] "According to this view, the mind is a computer with biological hardware and runs using programs essentially like those used in computers today".

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

postulent, principalement à la suite de Piaget, une étroite interdépendance entre langage et cognition, avec primauté tantôt de la cognition, tantôt du langage. Les défenseurs de l'hypothèse de la relativité linguistique, c'est-à-dire de l'hypothèse selon laquelle les différences entre les langues influenceraient la pensée habituelle s'opposent aux universalistes.

Dans les conceptions les plus récemment élaborées, la linguistique cognitive nie l'existence d'un mode de représentation autonome issu de la faculté de langage et semble confondre représentations cognitives et représentations linguistiques. La difficulté rencontrée est alors la même que celle qu'ont éprouvée les grammairiens tels que Ferdinand Brunot (1922) dont les investigations partaient de la pensée pour aller vers la langue. Il est véritablement difficile de se dégager de la circularité mentionnée au début de ce paragraphe. Quel(s) rapport(s) les opérations linguistiques entretiennent-elles avec les autres processus cognitifs ? Si les langues construisent une diversité de représentations du monde, est-ce à dire que chaque langue enferme ses locuteurs dans un système de pensée ? Derrière la diversité des langues se présente l'universalité de la faculté de langage, qui pose la question de la nature des invariants. Il reste que des conceptions plus ou moins larges de ce que sont le langage, la pensée, la cognition, l'intelligence, les représentations, les concepts et de ce que sont leurs interrelations etc., ne font que brouiller davantage le tableau général.

Le développement cognitif d'un individu est très certainement contraint par son environnement et par les catégories conceptuelles et culturelles véhiculées dans son environnement linguistique. Toutefois s'il existe des énoncés intraduisibles ils ne sont jamais incompréhensibles. Les langues peuvent forcer l'expression de certains concepts, elles peuvent contraindre ce que le locuteur doit transmettre et ainsi éventuellement, être dans certaines circonstances plus ou moins efficaces les unes par rapport aux autres.

Si l'on considère qu'il existe une même structure conceptuelle pour toutes les langues, on reste en interrogation devant le fait que certaines choses sont prises en charge par des éléments de classes fermées et d'autres par des éléments de classe ouverte. Talmy (2000a) propose que les éléments appartenant à la grammaire concourraient à la structure de la représentation cognitive tandis que les éléments de classes ouvertes contribueraient surtout à son contenu. Si tel est effectivement le cas, l'hypothèse serait que des différences dans ce qui est grammaticalisé dans une langue et pas dans une autre correspondraient à des différences de structuration conceptuelle. D'autre part, les procédés de grammaticalisation effectueraient des changements dans les structures conceptuelles des locuteurs d'une langue à travers le temps.

Est-ce à dire que la structure conceptuelle est superordonnée aux représentations cognitives qui, elles, seraient dépendantes des langues ? Et si tel est le cas, quel genre d'"évidence" la linguistique peut-elle avancer ? Pour l'heure, prendre position dans ce débat semble davantage résulter d'un acte de foi que d'une démarche scientifique.

Nous parlions plus haut de la polysémie que manifeste en français le terme *temps*. Cette question a été examinée à maintes reprises, et il se trouve que les différentes acceptions du terme *temps*, ne se laissent pas analyser d'une façon claire et évidente en dehors de l'acception météorologique, qui se dégage assez nettement. Nous

en vertu de la loi du droit d'auteur.

questionnerons donc en premier lieu le concept de temps.

Le temps est une réalité pour laquelle nous n'avons pas de perception immédiate. Nos sens ne permettent pas de l'atteindre, de l'appréhender directement. Aussi sa perception est-elle probablement différente entre les espèces, et pour les humains, entre les cultures. Le temps est une expérience qui ne cesse, malgré la familiarité que nous avons avec elle, de nous rappeler que nous n'en savons rien, se présentant de manière si évidente et naturelle qu'elle en devient presque indicible.

La physique prétend qu'avant le big-bang, le temps n'était pas et que c'est le refroidissement qui a permis la naissance du temps [351] . Une telle affirmation peut sembler quelque peu hermétique, voire paradoxale. L'argument est que l'instant zéro de l'univers se manifestant comme une purée de quarks homogène où tout se ressemble, où l'espace est indifférencié, ne connaît pas le temps. Le refroidissement serait donc à l'origine du temps puisqu'à l'origine de la différenciation. La recherche du commencement du temps suppose un "avant le temps sans temps" ce qui est difficilement concevable. Se demander ce qu'il y a avant le temps n'a pas de sens et pourtant notre intuition nous pousse à rechercher cet avant qui ne peut exister si le temps n'est pas.

La relation temps/température, mise en avant ici, ne serait que la médiation avec la relation temps/espace. Le temps n'existe pas dans l'espace indifférencié, le temps commence avec le changement [352] de l'espace lui-même initié par le changement de température. C'est dans l'opposition entre l'immuable et les métamorphoses que se dessine le concept de temps. Ce qui crée le temps c'est le changement. Dès que survient le changement, le temps existe.

Le temps est, du point de vue humain, également en rapport avec la mémoire. L'identification de situations comme répétées ou différentes n'est pas possible sans mémoire. Le temps ne se perçoit pas en lui-même, c'est l'observation des ressemblances/différences de l'environnement perçu qui permet de construire le concept de temps.

Les rapports que le temps entretient avec la température et l'espace sont ainsi primitifs, et de ces rapports, on peut envisager des rapports dérivés. La relation temps/espace engendre une relation d'ordre. Les changements qui surviennent sont ordonnés, ils se précèdent, se succèdent ou sont simultanés.

Le temps se mesure. La mesure du temps repose sur l'identification d'un phénomène dont la durée est clairement définie et qui soit reproductible à l'identique de façon infinie. Il s'agit de la recherche d'un étalon. Pour mesurer une distance, ou un poids, le choix de l'étalon est conventionnel. Pour le temps, ou la température, le choix de l'étalon est plus complexe. L'unité de temps est, au commencement de la physique, arbitraire, tout comme l'unité de longueur. En physique, l'unité de longueur dépend de l'unité de temps, elle est définie en fonction de l'unité de temps, de la seconde. La mesure du temps se base sur des phénomènes périodiques : course du soleil, mouvement d'un pendule… Elle s'affine

[351] Klein & Spiro (eds) (1996).

[352] Cette position est exploitée, entre autres, dans le modèle de Kamp et Reyle (1993).

en vertu de la loi du droit d'auteur.

progressivement en fonction des exigences de la société. Au XV^{e} siècle avant notre ère, en Égypte, on repérait l'heure du soleil selon l'ombre des pyramides. Le siècle suivant, l'ombre d'un bâton planté verticalement en terre, ancêtre des cadrans solaires, permet l'observation des heures claires et sans nuages. L'invention de la clepsydre permet ensuite de se libérer des contraintes météorologiques et solaires. Au IX^{e} siècle le temps se consume, en Angleterre avec la chandelle graduée et en orient avec des bâtons d'encens; puis s'égrène avec le sablier vers l'an mille. À la fin du XIII^{e} siècle, naît la première horloge à poids, puis la première horloge à pendule apparaît en 1657. Le principe de la montre à quartz est connu depuis 1920. À titre anecdotique, la seconde, depuis son invention, correspond à la 1/86 400^{e} partie du jour solaire. Cette base de calcul étant insuffisante, la seconde est redéfinie en 1955 comme la 1/36 556 925 947^{e} partie de l'année 1900. Il faut se rendre à l'évidence que tous les phénomènes utilisés présentent des irrégularités, pour remédier à ces imprécisions les scientifiques de notre siècle utilisent les propriétés des atomes et plus particulièrement celles du césium. La découverte de l'horloge atomique réduit l'imprécision, la seconde devient "*la durée de 9 192 631 770 périodes de la radiation correspondant à la transition entre les deux niveaux hyperfins de l'état fondamental de l'atome de césium 133*." [353] Ces précisions ont évidemment peu d'incidence sur notre façon courante de penser le temps, mais elles peuvent engendrer des pratiques nouvelles. De plus, la physique avance que contrairement à la perception qu'on en a, le temps est granulaire, c'est-à-dire qu'il existe une limite au-delà de laquelle il n'est plus possible de diviser le temps [354] . Einstein a également postulé que "l'écoulement du temps n'est pas le même pour deux témoins dès qu'ils bougent l'un par rapport à l'autre" [355] , ce qui remet en cause l'universalité du temps et oblige à penser ensemble espace et temps.

La relation temps/espace est encore en relation avec la distance. Les événements s'inscrivent dans un repère spatio-temporel. Mettre en rapport le temps et la distance aboutit au concept de vitesse. La propagation de la lumière est à l'échelle de la perception humaine instantanée [356] tandis que la propagation de l'onde sonore est d'environ 340 mètres par seconde. On peut donc à partir de ces vitesses calculer la distance qui nous sépare d'un point de foudre en comptant le nombre de secondes qui séparent la perception de l'éclair de la perception de la déflagration.

La question de la temporalité met donc en jeu différentes perceptions : celles du changement et de l'ordre dans lequel les changements surviennent, de la durée ou mesure du temps, et celle de la vitesse.

Les représentations du temps sont de deux ordres : cycliques ou linéaires. Le temps cyclique permet d'accéder au temps linéaire. Les premières horloges que furent le soleil

[353] Définition de la Conférence Générale des Poids et Mesures, le 13 octobre 1967.

[354] Jacquard (2001).

[355] Jacquard (2001:59).

[356] Selon les physiciens, la vitesse de la lumière ne saurait être dépassée.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

et la lune ont conduit à l'établissement des calendriers [357] et de l'Histoire. Ce sont donc des phénomènes qui se reproduisent dans le même ordre régulièrement qui permettent le repérage.

Il a été affirmé que des sociétés et des cultures différentes présentent des conceptualisations différentes du temps. Il apparaîtrait ainsi comme cyclique [358] , ramifié [359] ou linéaire. Des auteurs affirment que ces conceptualisations/représentations ne se manifestent pas au niveau de la grammaire d'une langue et qu'on peut utiliser la représentation traditionnelle du temps, c'est-à-dire une ligne (orientée) sur laquelle des points de référence peuvent être portés.

L'hypothèse de la relativité linguistique, dont les plus illustres représentants sont Sapir et Whorf, a été mise à l'épreuve en partie avec la question de la temporalité dans le langage. Ainsi que le rappelle Lakoff (1987), il existe différents types et différents degrés de relativisme, selon la façon de concevoir les différences de systèmes conceptuels. En particulier, concernant la question de la profondeur des différences de systèmes conceptuels, certains considèrent que les éléments qui composent le système n'ont pas tous le même statut. Il y aurait des concepts fondamentaux et d'autres plus superficiels, les plus fondamentaux étant les concepts de temps et d'espace. Cette distinction de statuts des parties d'un système conceptuel renvoie de façon assez proche à celle entre l'universel et le spécifique. Dire que les concepts de temps et d'espace sont fondamentaux dans un système conceptuel ne laisse que peu de probabilité que ces derniers puissent manquer, être absents. Les concepts mineurs, superficiels en revanche sont spécifiques et moins probablement présents dans les systèmes conceptuels. Toutefois cela ne dit rien de la comparabilité des concepts. De plus, les concepts fondamentaux, formant la base du système, sont plus susceptibles d'être réutilisés par d'autres parties du système conceptuel que les concepts spécifiques ; et l'on s'attend à ce que les concepts fondamentaux soient aussi plus grammaticalisés. Enfin, plus les concepts fondamentaux diffèrent, plus les systèmes diffèrent, puisque les concepts fondamentaux sont à la base de plus de concepts dérivés. Ainsi, Whorf (1970) manifeste une position radicale, en affirmant que les concepts de temps et de changement reçoivent, dans les langues amérindiennes et particulièrement en hopi, des caractérisations très différentes de celles rencontrées dans l'*Average European language* [360] . Cette conclusion est tirée du fait que ces langues ne présentent pas la catégorie

[357] Ewing Duncan, D. (1999). Cet ouvrage présente entre autres les manipulations que subissent les calendriers au cours de l'Histoire Occidentale et leurs motivations religieuses, politiques et scientifiques.

[358] Choi (2000).

[359] Diodore Cronos (fin IV av. J.C.) et ses disciples. Les logiciens de l'école de Mégare ont constitué un sytème axiomatique de la logique des propositions à partir des lois logiques établies par Aristote dans les *Premiers Analytiques*. La logique leur doit d'avoir dégagé des schèmes d'inférence (en particulier, les célèbres indémontrables : *modus ponendo ponens* et *modus tollendo tollens*) et étudié le fonctionnement des opérateurs (implication et disjonction notamment). On doit à Diodore Cronos la formulation des opérateurs modaux en termes temporels, et la conception du futur ramifié.

[360] Dénomination tirée de Whorf (1970).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

grammaticale du temps [361] . Cependant, ainsi que le remarque Comrie (1985:4) "il serait aussi logique d'affirmer que les locuteurs de langues qui ne connaissent pas la catégorie grammaticale du genre ont un concept du sexe radicalement différent des locuteurs de langues qui présentent ces catégories" [362] , et c'est bien là toute la question.

Nous avons donc pu dégager les différents aspects que recouvre le concept de temps. Il convient à présent d'examiner les moyens linguistiques d'expression de la temporalité dans nos corpus.

9.2. Les moyens d'expression de la temporalité dans le corpus

Les moyens d'expression de la temporalité dans une langue ou dans les langues sont très variés. Au niveau de la matérialité linguistique, en français, toutes sortes de catégories permettent d'exprimer la temporalité : noms, verbes, adjectifs, prépositions, adverbes etc. La syntaxe contient elle aussi des moyens de spécification du temps, et à un niveau d'analyse plus étendu encore, on peut considérer que des configurations textuelles portent des spécifications temporelles. Ce qu'il est important ici de considérer, c'est la part de ce qui est lexical et de ce qui est grammatical puis discursif dans l'expression du temps. On fait l'hypothèse qu'il y aura plus de divergence dans ce qu'il est possible d'exprimer par des moyens lexicaux et discursifs à travers différentes langues que dans ce qu'il est possible d'exprimer par des moyens grammaticaux (structurels) [363] , les moyens grammaticaux étant dédiés à l'expression de la structure temporelle (ou qualité) et non pas à la quantité.

9.2.1. Morphologie verbale

Le moyen de spécification du temps sur lequel s'est le plus précocement focalisée l'étude de la langue est le verbe. Ne trouve-t-on pas en effet chez Aristote que le verbe "est ce qui ajoute à sa propre signification celle du temps"? [364] Le verbe est conçu comme le lieu privilégié de l'expression du temps. Mais cette observation ne convient bien que pour les langues qui connaissent le phénomène de la conjugaison [365] , c'est-à-dire, si l'on se réfère aux linguistes contemporains intéressés par l'origine des langues et leur évolution, le

[361] Une réfutation de ces hypothèses se trouve dans Malotki (1983).

[362] "It would be equally logical to assert that speakers of languages lacking grammatical gender categories have a radically different concept of sex from speakers of languages with such grammatical categories".

[363] Il n'y a pas de distinction entre les concepts pour *aujourd'hui* et *maintenant* en yidiny – d'après Dixon (1977) cité par Comrie (1985:8).

[364] Aristote (1984) *De l'interprétation (Organon II),* tr. J. Tricot, Paris, Vrin.

en vertu de la loi du droit d'auteur.

phénomène né de la jonction de deux "verbes" différents pour les langues germaniques et le hittite et de trois verbes différents pour le grec et le sanskrit [366] , qui pour une même base lexicale présente une série de formes selon des variations en personne, en temps et en mode.

"Conjugaison" est attesté selon le *Petit Robert* en 1236, et provient de la terminologie grammaticale latine. Selon le dictionnaire étymologique de Picoche [367] , ce terme est issu d'une forme racine indo-européenne (*yug-) signifiant "atteler", qui donne *jugare, conjugare* "unir" et *conjugatio , -onis* "union, conjugaison" en latin. Ce terme renvoie donc à la manière dont les radicaux des verbes sont joints, attelés aux terminaisons. Pour le latin [368] par exemple, quatre conjugaisons sont distinguées selon la finale du radical présent :

amo, amas, amare deleo, deles, delere lego, legis, legere audio, audis, audire

Les définitions du verbe mêlent critères morphologiques, sémantiques et syntaxiques. La caractérisation syntaxique du verbe, dans les termes de Givòn (1984:73), est que "les verbes tendent à former le noyau obligatoire des phrases. C'est-à-dire qu'ils sont le plus souvent le prédicat de la phrase" [369] .

Nombre des caractérisations que reçoit la notion de verbe se heurtent à la comparaison inter-linguistique. La notion de verbe ne semble recevoir de définition universelle que dans la mesure où elle prend pour paramètre la notion d'intégration d'un ensemble de constituants nominaux en unité phrastique. Ainsi, Creissels (1995:11) rappelle que :

"une réflexion générale sur la grammaire n'a donc de sens que si, au-delà de la spécificité indéniable de chaque système morphosyntaxique, il est possible de définir des notions syntaxiques générales telles que les notions fondamentales dans la description de chaque langue particulière puissent se laisser reconnaître comme la forme particulière que prennent dans les différentes langues ces notions syntaxiques générales".

La caractérisation sémantique (notionnelle) [370] du verbe repose sur l'idée de répartir les classes lexicales majeures dans les langues (nom, verbe, adjectif, adverbe) sur une échelle de "stabilité temporelle". Ainsi, les langues tendraient à lexicaliser les expériences qui restent relativement stables dans le temps (à l'échelle de l'observation humaine) comme des noms, tandis que les événements et actions, qui sont des changements

365 Parmi les langues qui ne connaissent pas la conjugaison, la plus illustre est sans doute le chinois.

366 Martinet (1986).

367 Picoche & Firmin (1992).

368 Lebelle (1938:18). Pour le français, la grammaire traditionnelle distingue trois conjugaisons en fonction de la marque de l'infinitif et du participe présent. Dubois (1967) distingue les conjugaisons en fonction du nombre de bases présenté par les formes verbales.

369 "verbs tend to form the obligatory nucleus of sentences. That is, they are most commonly the predicate of the sentence".

370 Givòn (1984:64).

en vertu de la loi du droit d'auteur.

rapides donc moins stables dans le temps, seraient plus probablement lexicalisés comme des verbes. Noms et verbes sont alors disposés chacun à une extrémité de l'échelle. Les verbes tendraient donc à coder des expériences moins stables dans le temps. Ici se pose une première distinction entre chronologie et stabilité temporelle (ou changement).

S'il est courant de rencontrer des langues dans lesquelles le verbe connaît l'affixation d'indices pronominaux et/ou autres variations, il est des langues sans système d'affixes verbaux, ou qui ne connaissent qu'un système d'indices pronominaux. Toutefois, à travers les langues, il est souvent observé que le verbe fait preuve d'une complexité morphologique très supérieure à celle que connaît le nom, et qu'il est rare de ne pas être en mesure de reconnaître un verbe en fonction d'un système d'affixation particulier. Au niveau de la caractérisation morphologique, il s'agit d'identifier les sous-systèmes qui se groupent autour du verbe : temps aspect modalité, négation, accord pronominal, morphèmes transitivants ou détransitivants (voix). Toutes les langues n'ont pas le même nombre de tiroirs verbaux. Mais parmi les systèmes les plus fournis, c'est-à-dire ceux pour lesquels il existe une distinction relative au degré d'éloignement, le nombre de tiroirs n'excède pas dix.

Au sujet de la conjugaison, ou des temps grammaticaux, Comrie (1985:3) rappelle que :

"Clairement, si notre intention est de fournir une explication du temps verbal qui soit valide pour toute langue, alors cette explication ne doit pas être basée sur des concepts du temps culturellement spécifiques mais devrait plutôt être une théorie générale appropriée à toutes les cultures et ainsi à toutes les langues" [371] **.**

C'est pourquoi nous avons mené une discussion sur le concept de temps, pour laquelle nous avons vu qu'il semble toutefois difficile de se dégager complètement des problématiques culturelles, mais qui nous a permis d'identifier les diverses dimensions du concept.

> Découvrir des morphèmes flexionnels verbaux repose sur une analyse morphologique, qui procède donc par commutation. La technique de la commutation, se pratique en deux temps. Pour une forme telle que *je chante* /ʒœʃãt/, on fera varier avec *tu chantes* /tyʃãt/, puis avec *je marche* /ʒœmarʃ/. Mais ce n'est qu'après avoir épuisé le paradigme entier des formes verbales qu'on pourra clairement identifier des morphèmes pour le système de l'oral et celui de l'écrit[376].

Un morphème se définissant comme la plus petite unité à deux faces, le principe de d'identification se base sur la forme et le sens. Pour chaque catégorie grammaticale, on devrait donc pouvoir définir un jeu de conditions nécessaires et suffisantes ou un prototype représentant la signification de la catégorie en question et justifiant de son emploi, de ses emplois particuliers. Il se trouve que la signification des morphèmes de

[371] ***"Clearly, if our intention is to provide an account of tense valid for any language, then this account must not be based on culture-specific concepts of time, but should rather be a general theory appropriate to all cultures, and thus to all languages".***

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

temps présente ce que l'on pourrait en premier lieu qualifier de polysémie de la catégorie, voire d'homonymie, dans le cas où les liens entre les significations postulées ne se laissent pas expliquer de façon immédiate et transparente.

La morphologie verbale du français ne spécifie pas seulement la temporalité. En effet, trois catégories linguistiques semblent si étroitement liées qu'elles sont presque indissociables : le temps, le mode et l'aspect – toutes trois faisant partie de la caractérisation des formes verbales en français.

Dès lors, se posent des problèmes de terminologie. Si en français la polysémie du terme "temps" permet de repousser la question du point de vue, ou du niveau d'analyse ; l'emploi de *mode* et *aspect* requiert que des précisions soient apportées. Nombre de linguistes se sont penchés sur la question, se demandant s'il y avait lieu de les envisager séparément ou ensemble [372] . Pour notre propos, il nous semble indispensable de les distinguer dans un premier temps, quand bien même elles présenteraient des zones de recouvrement, afin d'être en mesure de déterminer ce qui ressort de l'une ou de l'autre.

Il faut bien entendre pour commencer que le temps linguistique se distingue du temps notionnel ou expérientiel, et qu'il n'entretient pas des rapports simples avec ce dernier. On invoque souvent pour preuve de cette distinction les termes anglais ou allemands, respectivement *time* et *tense*, *Zeit* et *Tempus* mais il s'agit là d'une dimension supplémentaire, qui parmi les moyens d'expression de la temporalité distingue les phénomènes présentés par la morphologie verbale (temps verbal) des autres phénomènes.

Dans de nombreux travaux, la notion de temps se forme selon une conception linéaire, continue et unidimensionnelle. Le temps morphologique s'analyse couramment à partir de la représentation de la temporalité par un axe horizontal tripartite orienté à droite, c'est-à-dire mettant essentiellement en jeu des questions de chronologie, d'ordre dans lequel se déroulent les événements. Cette tripartition représente le découpage du temps en époques : passé, présent, futur. Il s'agit d'une représentation linéaire, où passé et futur sont conçus comme des infinis orientés dans la même direction (vers le futur), le présent étant figuré par un point [373] .

La temporalité est construite à partir du "maintenant" du locuteur. Le moment du dit est repéré par rapport au moment du dire. Ces approches effectivement ne permettent guère que de situer le moment de l'événement énoncé par rapport au moment de son énonciation, et donc de fournir une conception tripartite du temps : ce qui est énoncé comme se présentant avant, pendant, ou après le moment de l'énonciation. La catégorie des temps est alors déictique puisque le temps des événements est situé par rapport au moment de l'énonciation.

Les temps liés au moment de l'énonciation, au moment de parole sont des temps dits absolus [374] . On parlera de temps relatifs, lorsque ceux-ci permettent le repérage d'une situation (événement) par rapport à une autre situation (événement) elle-même le plus

[372] Gosselin (1998).

[373] Notons que Guillaume (1929) présente une autre analyse, que nous évoquerons plus bas.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

souvent repérée déictiquement. La distinction se pose également en termes de temps anaphoriques ou non anaphoriques. Les temps dits anaphoriques ont besoin d'un point d'ancrage pour obtenir la référence temporelle, tandis que les temps non anaphoriques se suffisent en eux-mêmes en relation au moment de l'énonciation pour établir cette référence.

(1) Alicia a perdu son bracelet. (2) Sylvain se promène dans les bois. (3) Depuis que Pierre est parti, Marie ne mange plus.

Les approches qui situent le moment de l'événement décrit par rapport au moment de parole dérivent, de près ou de loin, de la conception reichenbachienne du système des temps [375] . Ces conceptions donnent un rôle central à l'énonciation dans le traitement du temps dans la langue, elles formulent une sémantique minimale des temps et permettent d'indiquer les relations temporelles entre les phrases dans le discours. C'est la configuration de trois "moments" sur la ligne du "temps" qui détermine les différents items de la représentation des temps : soit S le moment de parole, E le moment de l'événement et R le moment de référence.

Au niveau de la morphologie verbale, l'expression du temps serait le résultat d'une évolution qui serait caractéristique des langues indo-européennes alors qu'on trouverait dans les autres familles de langues une préférence pour l'expression de l'aspect, la morphologie du verbe restant indifférente au temps chronologique. Pour le français, qui hérite son système verbal du latin, il est traditionnel de reconnaître que les formes verbales répondent à une division du temps en trois époques : passé, présent, futur et permettent d'exprimer en outre le "futur dans le passé" et le "passé dans le futur", ce qui revient à accepter l'idée que le système verbal du français exprime essentiellement le temps chronologique. Il semble que cette évolution parte d'une innovation du latin alors que l'indo-européen d'après la grammaire comparée exprimait surtout l'aspect. Pour Vendryes (1923:131), "l'indo-européen était beaucoup moins préoccupé de marquer le temps que la durée. Ce qui l'intéressait dans une action, ce n'était pas d'indiquer à quel moment (passé, présent ou futur) s'accomplissait l'action, mais de marquer si on l'envisageait dans sa continuité ou à un point seulement de son développement, si c'était le point initial ou le point final, si l'action n'avait lieu qu'une fois ou se répétait, si elle avait un terme et un résultat."

L'aspect entretient des rapports étroits avec la temporalité chronologique. Il peut se marquer dans la morphologie verbale. Le temps verbal, d'après Comrie (1985:9) est :

"[…] l'expression grammaticalisée de la localisation dans le temps" [376]

tandis que l'aspect est :

"[…] la grammaticalisation de l'expression de la constitution temporelle interne" [377]

Il s'agit donc aussi, au point de vue sémantique, de temporalité, mais de celle du procès

374 Comrie (1985:36-82).

375 Reichenbach (1947).

376 ***"grammaticalised expression of location in time".***

en vertu de la loi du droit d'auteur.

en lui-même. Les définitions de l'aspect sont de loin celles qui manifestent le moins de consensus au sein des théories linguistiques, et il est courant de trouver dans la littérature des positions telles que celle de Andersson (1972:39) [378] au sujet de l'allemand :

"En ce qui concerne le débat au sujet de la notion d'aspect allemande, je voudrais à la question si il existe un aspect en allemand, pour ma part répondre de la manière suivante : cela dépend de ce qu'on entend par aspect et par existe."

Il s'agit en effet de définir précisément la notion d'aspect et d'identifier clairement ses moyens d'expression et les niveaux de descriptions pris en compte. Il est aujourd'hui de coutume de rappeler que cette notion a pour origine dans les théories linguistiques la linguistique slave [379] et que le phénomène originellement étudié dans les langues slaves connaît des définitions plus spécifiques et ne correspond plus à la diversité des conceptions de l'aspect. Le terme d'aspect est employé aussi bien pour parler de catégories morphologiques, que pour désigner des catégories lexico-sémantiques, et on emploie souvent dans ce dernier cas le terme d'*Aktionsart*.

Guillaume (1929) distingue ainsi le temps expliqué du temps impliqué, à mettre en parallèle avec l'opposition bergsonienne entre temps vécu, subjectif et temps pensé, objectif. C'est là l'opposition qui fonde la notion d'aspect comme temps interne au procès, temporalité intrinsèque. Le temps grammatical verbal situe les procès par rapport à un repère, ou situe les procès les uns par rapport aux autres. Le procès peut être envisagé quant à sa place sur l'axe du temps (chronologie) ou quant à son déroulement interne et aux bornes qui le limitent. Il s'agit là de deux dimensions temporelles. L'aspect exprime les manières différentes de voir la constitution temporelle interne d'une situation. Ainsi le sémantisme du verbe implique un certain rapport au temps. La signification du verbe lui confère un certain "mode d'action". Il existe alors une dimension aspectuelle inhérente au verbe, lexicale. C'est par cette dernière qu'on opposera par exemple l'inchoativité de *chercher* à la résultativité de *trouver*.

(4) a. il a cherché les clefs pendant des heures b. *il a trouvé les clefs pendant des heures [380]

L'inchoativité accepte une spécification durative, tandis que la résultativité n'accepte pas de spécification durative.

L'aspect peut se manifester au niveau morphologique, grammatical, et c'est dans ce cas que le locuteur choisit de présenter le procès comme un résultat, une durée...

Au sujet de l'aspect comme catégorie morphologique, les définitions négatives du type "l'aspect est tout ce qui dans le verbe ne correspond pas aux catégories de

377 ***"[…] the grammaticalisation of expression of internal temporal constituency".***

378 "Angesichts der Debatte über deutsche Aspektbegriffe möchte ich die Frage ob im Deutschen ein Aspekt existiert, meinerseits auf folgende Weise beantworten : Es hängt davon ab, was man unter Aspekt versteht, und unter existiert " Andersson (1972:39) cité par Wilmet (1980) – Notre traduction.

379 Cohen (1989).

380 ***Une lecture itérative permettrait de lever l'astérisque.***

en vertu de la loi du droit d'auteur.

personne, nombre, voix, temps et mode" ne sont pas satisfaisantes dans la perspective typologique dans la mesure où l'ensemble des morphèmes verbaux n'est pas fixé.

Au niveau de la terminologie, il est donc deux notions à préciser : celles d'aspect et d'*Aktionsart*. On peut relever dans la littérature différentes acceptions de ces termes, nous nous efforcerons de donner quelques pistes de différenciation de ces termes.

La distinction aspect/*Aktionsart* se joue soit sur un plan purement sémantique, soit en fonction du niveau de description auquel aspect et *Aktionsart* sont rattachés. Au niveau sémantique (notionnel), la littérature propose que l'*Aktionsart* concerne les types de procès, tandis que l'aspect réfèrerait à différentes perspectives possibles sur un seul et même procès. L'*Aktionsart* appartient alors au lexique, tandis que l'aspect appartient à la grammaire.

Les dimensions aspectuelles (au niveau morphologique ou lexico-sémantique) relevées par les auteurs sont nombreuses, et intègrent parfois des caractères qui ne sont pas strictement liés à la temporalité (locatifs, intrumentaux… [381]). Parmi ces dernières, on peut répertorier les plus usitées, en notant toutefois que les auteurs n'ont pas toujours une définition consensuelle pour chacune d'entre elles. La notion de durée est spécifiée pour l'aspect en termes de duratif *vs* non duratif (ou 'momentané', 'pontuel'). Le nombre verbal (itération) est identifié comme une opposition entre semelfactif *vs* répété (itératif ou fréquentatif), dimension étroitement liée à la celle de la durée. Une autre dimension est la détermination ou orientation du procès (télique *vs* non télique) : l'action verbale est orientée ou non vers un point à atteindre, un but (par ex. *marcher / aller*), elle contient l'idée de franchissement d'un seuil. La dimension de l'achèvement se subdivise en achèvement objectif et achèvement subjectif. L'achèvement objectif ressort de l'opposition terminatif *vs* non terminatif. L'achèvement subjectif élicite l'opposition perfectif *vs* imperfectif. On considère aussi la dimension du résultat (l'action est présentée avec ou sans son résultat), qui est, elle aussi, divisée en deux : objectif / subjectif (concernant l'agent de l'action). On trouve encore des spécifications telles que la dimension de vision de l'action comme partialisée ou non partialisée, l'expression de la phase ou du degré de déroulement de l'action. Enfin, il est examiné la collocation ou incidence (action verbale située par rapport à d'autres actions non nommées mais impliquées).

Temps et mode entretiennent également des rapports, bien que ceux-ci soient notionnellement moins évidents à déterminer. La présentation des formes verbales par les grammaires traditionnelles suggère ces rapports, puisqu'elle propose une division des modes en temps. Cependant, les critères qui permettent de définir les modes sont plus qu'hétérogènes et la division des modes en temps paraît quelque peu artificielle. Le mode le plus achevé, pour le latin, le grec classique et le français entre autres, est celui de l'indicatif, et il semble que les divisions des autres modes aient été plus ou moins calquées sur ce dernier.

Nous avons distingué temps du procès et temps intérieur au procès. Si le mode surplombe ces derniers, on peut se demander quel est son exercice. Est-il manière d'exister du procès, manière de présenter le procès ? Le mode peut être envisagé au plan

[381] Ceci est contribué par le fait que l'aspect entretient des relations avec le cadre participatif.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

formel (les modes sont alors distingués selon la présence ou absence des marques de personne, de genre...) ou au plan sémantique. Il n'est *a priori* pas possible de considérer les modes établis par la grammaire traditionnelle sous un même angle. Pour exemple, on n'est pas en mesure de donner une description sémantique pour chaque mode. Il est relativement complexe d'y parvenir pour les modes personnels, et la description est bloquée avec les modes impersonnels.

La modalité ne se cherche pas en fait uniquement dans les formes verbales, et il faut distinguer mode et modalité, ainsi qu'on distingue temps verbal et temporalité. Traditionnellement, on recherche la modalité dans le mode, ce qui renforce son statut de catégorie exprimée par la morphologie verbale. Il semble cependant que dans de nombreuses langues, la modalité ne soit pas principalement exprimée dans la morphologie verbale [382] , que l'on songe aux verbes modaux anglais, aux particules modales en Grec [383] ... La modalité se définit comme l'expression de l'attitude du locuteur à l'égard de ce qu'il dit [384] , à l'égard du contenu propositionnel de son énoncé, pour être plus précis. Il n'y a pas de raison qu'elle ne soit portée que par le verbe, en dehors du fait qu'il est souvent l'élément le plus central d'un énoncé, réalisant la prédication.

On constate que des marques temporelles, aspectuelles sont susceptibles d'exprimer des valeurs modales :

(5) a. Il aura pris la voiture de sa mère. (5) b. Il a sûrement pris la voiture de sa mère. (6) a. Marie, elle te fait une robe sans patron. (6) b. Marie peut te faire une robe sans patron.

On peut également constater de la modalité dans la diathèse :

(7) a. Les sauterelles, ça se mange. (7) b. On peut manger des sauterelles.

Il faut donc noter qu'il n'est pas de relation biunivoque entre temps et temporalité, mode et modalité, aspect et "aspectualité". Ainsi des marques temporelles ou aspectuelles peuvent recevoir une valeur modale, et la temporalité, la modalité, et l'aspectualité sont susceptibles d'être portées par des marques autres que verbales, l'aspectualité restant plus étroitement liée au verbe par le fait qu'elle spécifie le procès.

L'analyse des marques morphologiques à laquelle nous adhérerons, est celle proposée par Touratier (1996). Selon lui,

"les temps simples, les temps composés et les temps surcomposés de la conjugaison française constituent trois séries de temps entièrement parallèles. Les temps composés sont en effet formés à l'actif, à partir des temps simples de l'auxiliaire avoir ou être, suivis du participe dit passé du verbe à conjuguer. Et les temps surcomposés sont formés à partir des temps composés du seul auxiliaire avoir, suivis eux aussi du participe passé du verbe à conjuguer, c'est-à-dire par conséquent qu'ils ajoutent le participe passé du verbe à conjuguer non plus au temps simples d'un auxiliaire, mais à la combinaison que constituent les temps

[382] Palmer (1986).

[383] Basset (1989).

[384] Bally (1965).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

simples d'un premier auxiliaire et le participe passé d'un second auxiliaire." [385] .

Il se pose d'emblée une question au sujet du passif. En effet, si on oppose temps composés à temps simples, alors le problème se pose pour le passif, qui est constitué de formes composées et surcomposées. La conjugaison des formes de passif n'est que très rarement présentée dans les ouvrages de référence (grammaires, dictionnaires…). La raison en est sans doute, que les formes de passif n'apparaissent que pour les verbes transitifs et répond plutôt à une question de syntaxe. Le passif de première et deuxième personnes n'apparaît que pour les verbes qui acceptent à l'actif un objet animé/humain. Ainsi que le remarque Gaatone (1998:66) :

"la question qui se pose est donc la suivante : est-il possible de prédire la passibilité d'un verbe en français ? Comme pour beaucoup d'autres phénomènes syntaxiques, la réponse ne va pas de soi. L'étude minutieuse des données révèle un comportement hautement idiosyncrasique des verbes par rapport au passif, à tel point qu'on peut être tenté d'y voir un problème à reléguer au lexique".

D'autre part, sont ordinairement considérées deux formes de passif : passif d'action et passif d'état [386] . Touratier (1996:192) propose de ne considérer que la première comme véritable forme passive du verbe [387] . Nous considérerons également avec Touratier (1996) que le morphème du passif est homonyme de certaines formes du morphème d'accompli.

L'analyse de Touratier (1996) amène à considérer des catégories sans marques formelles (indicatif, présent, actif, formes simples) et des morphèmes de temps, d'aspect, de mode et de voix, ce qui lui permet de construire la représentation suivante (Figure 9-1).

385 ***Touratier (1996:33).***

386 Blanche-Benveniste (1984), Le Goffic (1993:200-201).

387 Les formes de passif d'état s'analysent comme un participe passif combiné avec *être* et occupant une fonction d'attribut. Le participe passif serait une variante de la forme de la série composée.

en vertu de la loi du droit d'auteur.

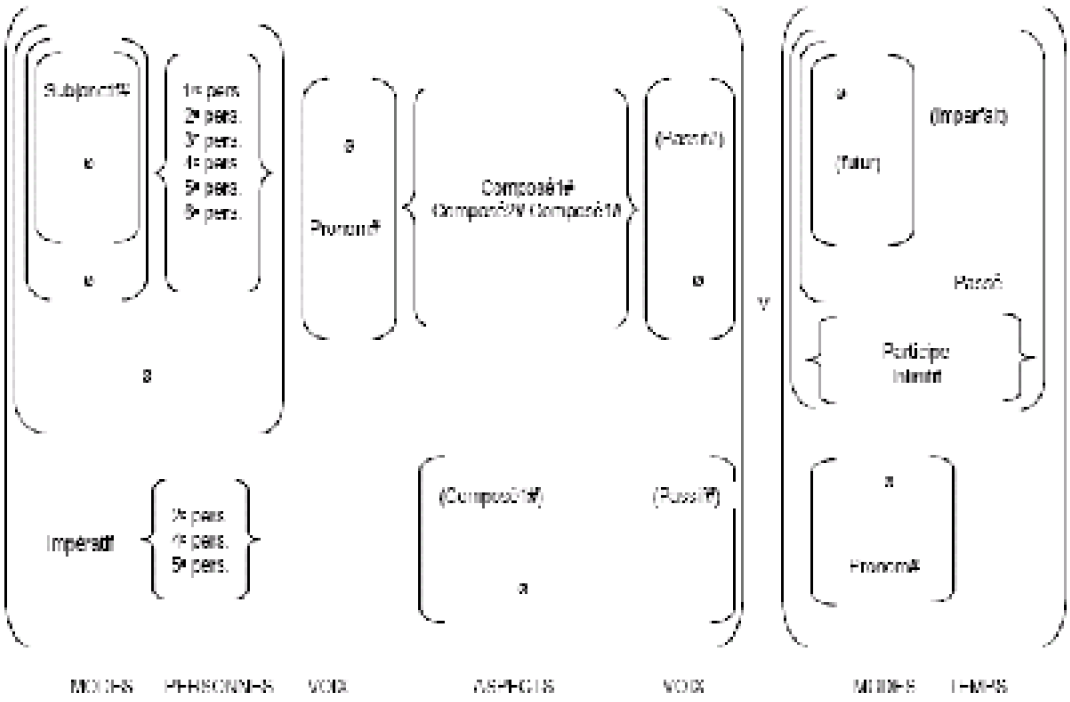

Figure (9-1) – "Structure morphologique ordonnée" du verbe français – Touratier (1996:62 Fig. 9) [388]

Touratier (1996) résout le problème des catégories sans marques formelles (mode indicatif, présent, tiroirs simples, voix active) en postulant pour le présent que "Dépourvu de signifié propre, le présent est en effet apte à apparaître dans les types d'énoncés les plus divers, voire les plus curieux au point de vue de la temporalité"(Touratier 1996:96). L'actif, l'indicatif et les formes simples sont caractérisés de la même manière :

Touratier (1996:98) propose en effet que "l'absence de marque formelle propre qui caractérise l'indicatif par rapport au subjonctif n'est nullement le signifiant zéro d'un morphème modal, dont le signifié ne semble pas définissable, et que le terme d'indicatif désigne simplement l'ensemble des formes verbales qui ne contiennent aucun morphème

[388] *Apparaissent entre accolades les catégories qui constituent une classe dont un des éléments doit être présent. Entre parenthèses apparaissent les catégories qui peuvent ne pas apparaître. Les catégories compatibles apparaissent sur une même ligne. sont rassemblée entre crochets.*

en vertu de la loi du droit d'auteur.

de mode."

De même, "les formes dites simples du verbe n'associent aucune signification à leur absence de marque formelle […] les prétendues formes simples ne sont que les formes verbales dépourvues de morphème aspectuel". (Touratier 1996:99)

Enfin, selon Touratier (1996:100) "Il est clair qu'il n'y a rien de commun du point de vue de la sémantique entre tous les verbes à la voix active. Il est donc préférable de rattacher à l'actif l'ensemble des formes verbales qui n'ont aucune unité morphologique là où les formes dites passives présentent une unité morphologique /*et*…PP [389] , et les formes dites pronominales une seconde unité morphologique de personne dite réfléchie. L'actif, n'ayant ni signifiant ni signifié, correspond donc à une absence de morphème de voix."

Les morphèmes de temps sont alors le passé simple, dont le signifié est "passé factuel", l'imparfait auquel est associé le signifié de "non actuel".

Un seul morphème d'aspect est envisagé, associant un signifiant discontinu à un signifié d'"accompli" ou "achevé".

Les morphèmes de mode sont le subjonctif, le futur, l'infinitif, le participe, et le gérondif. Le cas du subjonctif est particulier, puisqu'il s'en dégage trois morphèmes : morphème de volonté [390] , de possibilité, et de servitude grammaticale. Le morphème du futur reçoit le signifié de "projeté". Infinitif participe et gérondif sont les signifiants de morphèmes ayant un signifié purement grammatical et ne sont combinables qu'avec le morphème d'aspect et/ou les morphèmes de voix.

Les morphèmes de voix identifiés par Touratier (1996) sont celui du "passif" et le "pronominal" subdivisé en "réfléchi/réciproque" et "passif".

Nous avons établi l'inventaire des tiroirs verbaux dans le corpus. Les formes simples sont proportionnellement bien plus usitées que les formes composées. Toutefois, c'est le présent qui contribue majoritairement ce résultat. En effet, ce tiroir représente plus de 70% des formes verbales finies. Cependant cela ne signifie pas que les locuteurs ne réfèrent qu'à des situations afférentes au moment de parole. Le présent dans les corpus représente une constellation de valeurs, il est loin d'être toujours déictique.

Tableau (9-1) – Répartition des tiroirs verbaux dans le corpus [391]

[389] *être* fléchi associé à un verbe au participe passé.

[390] L'impératif est également le signifiant du morphème de volonté selon Touratier (1996:167).

[391] Les tiroirs verbaux s'entendent toutes voix confondues (active, passive, pronominale). Nous avons conservé la terminologie traditionnelle.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

tiroir	proportion
présent	74,43
impératif présent	6,80
imparfait	3,42
futur simple	2,03
conditionnel présent	1,88
subjonctif présent	1,12
passé simple	0,09
Total formes simples	89,79
passé composé	6,79
plus-que-parfait	0,39
futur antérieur	0,03
conditionnel passé	0,49
subjonctif passé	0,04
passé antérieur	0,01
futur périphrastique	2,45
Total formes composées	10,21
	100

Au niveau de ce simple inventaire, on est frappé de constater que le présent, c'est-à-dire les formes verbales qui ne comportent que la marque de la personne, présente une proportion si importante.

On pourrait être tenté d'ajouter à cet inventaire trois tiroirs qui ne sont traditionnellement pas reconnus comme tels. Il s'agit du futur périphrastique à l'imparfait (8), des formes en *venir de* sans autres marques formelles que celles de personnes (9), et des formes en *venir de* à l'imparfait (10).

(8) Extrait du Corpus F7 *PJQ: j'allais acheter une 607 HDI à 200 kF, mais une TT à 150 pas mal aussi lol %add: PEI (9) Extrait du Corpus P2 *PAP: PAY pourtant jvien douvrir la fenetre.. :/ %add: PAY (10) Extrait du Corpus F3 *PAE: PCO bah tu m'a donne l'url j'y suis alle (2) *PAE: mais j'ai rien ecoute je venais de mangfer un TUC %add: PCO

Ces "tiroirs" ressortent davantage de ce qu'on a coutume de considérer comme des semi-auxiliaires, mais dans ce cas, la position du futur périphrastique, avec lequel ils présentent de nombreuses similarités, est à revoir. D'un point de vue sémantique, ils appartiennent aux moyens de repérage temporel, tandis que les semi-auxiliaires couramment identifiés pour le français se rapportent à l'aspectualité ou à la modalité.

9.2.2. Autres marques de temporalité

Nous avons identifié deux types principaux concernant la temporalité : d'une part le repérage chronologique et d'autre part la temporalité interne, comme catégories de la morphologie verbale. D'autres moyens linguistiques permettent de les manifester.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

9.2.2.1. Temporalité

Pour ce qui concerne le repérage chronologique d'un événement, si couramment différents moyens lexicaux et grammaticaux sont identifiés en dehors de la morphologie verbale, il est assez difficile de soutenir cette position. Le repérage d'un événement par rapport au moment de l'énonciation ou par rapport à un événement énoncé ne se fait pas directement par un item lexical mais par la fonction qu'occupe un syntagme dans une unité phrastique.

On peut ainsi déterminer un ensemble de prépositions et de conjonctions, de noms autodéterminés, d'adverbes, de noms et de verbes qui peuvent, suivant leur mode d'intégration à l'unité phrastique, contribuer à la localisation d'un événement dans le temps.

Dans nos corpus, l'inventaire de ces unités se présente comme suit :

Les connecteurs interphrastiques dans notre corpus qui permettent le repérage chronologique d'un événement sont : *et, puis, après, ensuite, avant, au préalable, entre-temps*.

Les conjonctions de subordination présentes dans le corpus, manifestent un éventail de valeurs sémantiques plus large de possibilités dans notre corpus. L'inventaire est le suivant : *quand, lorsque, pendant que, dès que, après que, après* +Vinf*, avant (même) que, avant de* +Vinf*, (à) chaque fois que, tant que.* Dans notre corpus, elles permettent de spécifier l'ordre des événements, mais également des cooccurrences d'événements, et la simultanéité d'événements.

En ce qui concerne la catégorie des adverbes, nous avons mentionné au chapitre 7 que nous considérerions certains d'entre eux d'après Creissels (1995:139-143) comme des noms autodéterminés. Ainsi, selon Creissels (1995:140) *maintenant, jamais, quand ?* "peuvent se définir comme substituts de syntagmes construits au moyen du substantif *moment"* et des "adverbes" tels que "*hier, aujourd'hui, demain* peuvent se définir comme des substituts de syntagmes construits au moyen du substantif *jour*".

L'inventaire que nous pouvons dresser pour notre corpus est donc le suivant pour les adverbes et noms autodéterminés.

Certains, dans leur sémantisme contiennent l'idée de repérage chronologique. Il s'agit des unités suivantes : *maintenant, là, tout de suite, tout à l'heure, présentement, dernièrement, aujourd'hui, demain, hier, avant-hier, tôt, tard, bientôt, tantôt, soudain, (tout) d'un coup, tout à coup.* On notera que ce repérage peut être plus ou moins précis.

D'autres font apparaître la qualification de l'occurrence d'un événement. Il s'agit pour nos corpus des unités *déjà, encore, de nouveau* et *toujours, tout le temps, souvent, généralement, parfois, uniquement, seulement.*

Dans nos corpus, le repérage d'un événement peut faire appel à la mesure du temps. La durée peut être dénotée par ce type d'unités et dans nos corpus, l'inventaire se limite à *longtemps.*

La vitesse, dimension connexe à celle de la durée, connaît dans nos corpus un

en vertu de la loi du droit d'auteur.

inventaire plus important d'unités : *vite, rapidement, lentement, précipitamment.*

Les noms qui contiennent une spécification temporelle peuvent, pour nos corpus, également se répartir selon les dimensions identifiées pour le concept de temps. Les noms en relation avec la mesure du temps (durée) dans nos corpus sont les suivants : *temps, laps de temps, bail, an, année, mois, semaine, jour, journée, soirée, matinée heure, plombe, minute, seconde, moment, instant, période*.

Certains noms dénotent un repérage relatif à un événement et non pas à l'axe temporel fléché, ils permettent la saisie d'un moment du déroulement d'un événement. Il s'agit pour nos corpus de *fin, début,* et *milieu.*

Des noms permettent de référer au temps cyclique et calendaire et autorisent le repérage (temps situé : *matin, midi, soir, après-midi, nuit, époque, lundi, mardi, mercredi, vendredi, samedi, dimanche, week-end, février, avril, mai, juin, juillet, août, septembre, octobre, décembre, été, Noël, Pâques*). Enfin *habitude, fois,* et *coup* sont les seuls noms de nos corpus permettant la qualification de l'occurrence d'un événement sur l'axe temporel fléché.

Un certain nombre d'adjectifs dans nos corpus ont une spécification temporelle. Ils permettent de dénoter la qualité d'une durée : *court, long, lent, rapide, bref,* ou un ordre : *dernier, prochain, premier ;* voire un véritable repérage chronologique *futur, nouveau, ancien, vieil.*

En ce qui concerne les verbes, il est évident que la plupart d'entre eux ont un signifié qui implique de la temporalité [392] , mais quelques-uns, dans nos corpus ont un signifié en rapport étroit à la temporalité. Il s'agit des unités *durer, tarder, approcher* [393] *, arriver.*

Enfin, des semi-auxiliaires permettent de spécifier la saisie d'un procès dans son déroulement (*commencer à, recommencer à, être en train de, continuer à, terminer de, finir de*) ou de spécifier un repérage chronologique (*venir de).*

Les prépositions qui apparaissent pour introduire des adverbiaux temporels ne sont pas spécifiques. Toutefois, tous les membres de l'inventaire total des prépositions du français ne permettent pas d'introduire des adverbiaux temporels. Dans nos corpus, l'inventaire des unités qui le permettent est limité à la liste suivante : *dans, en, pendant, à, il y a, depuis, jusqu'à, après, avant, à partir de, entre.*

La négation entretient également des rapports avec la référence temporelle. Les éléments *ne…plus, ne…jamais* spécifient les occurrences d'un événement sur l'axe temporel fléché.

Des constructions spécifiques, enfin, permettent la référence temporelle. Dans nos corpus, elles permettent de dénoter la durée d'un événement (*mettre* CN durée *à* P, *ça fait* CN durée *que* P).

Le moyen de la référence temporelle chronologique préférentiel dans nos corpus est l'emploi d'unités lexicales dont le sémantisme est temporel dans des syntagmes

[392] Cf. section suivante.

[393] Dans des contextes spécifiques.

en vertu de la loi du droit d'auteur.

prépositionnels en position d'adverbiaux temporels.

9.2.2.2. Sémantisme des verbes

En ce qui concerne le verbe et la spécification de la temporalité intrinsèque, pour le français, sont signalés trois types de marquage grammatical au niveau morphologique : tiroirs verbaux composés/surcomposés/simples, affixes verbaux (préfixes, suffixes), semi-auxiliaires aspectuels. L'aspect affixal proposé par Wilmet (1998), est remis en cause pour le français dans la mesure où les processus ne sont pas réguliers (ils ne s'appliquent pas sur tous les verbes ni sur des classes de verbes). On ne peut pas non plus affirmer que les morphèmes répertoriés constituent toujours des marques d'aspect.

Les préfixes identifiés comme aspectuels par Wilmet (1998) sont *re-, de-, a-, en-, par-* et ils sont caractérisés comme aspectuels de la façon suivante :

re-	duplicatif	refaire, rejouer, repartir
dé	inversif	défaire, découdre
a- et en-	perfectivisant	apporter, attirer, emporter
par-	perfectivisant	parachever, parfaire

Les suffixes formeraient un aspect multiplicatif, répétitif avec effet secondaire de minimisation : *-aill-, -ass-, -el-, -ill-* … (*bouillonner, sautiller*…)

Les semi-auxiliaires aspectuels permettent d'exprimer différents points de vues sur le déroulement du procès, considéré à différentes étapes de sa réalisation.

On considère d'autre part qu'il existe une spécification de l'aspect inhérente au verbe, c'est-à-dire que l'aspect s'inscrit dans le sémantisme du verbe lui-même. De cette observation découlent des typologies de procès, parmi lesquelles, la plus usitée sans doute est celle de Vendler (1967).

Pour le français, parmi les travaux précurseurs, on doit citer ceux de F. Brunot (1922) dont l'entreprise est particulière, à cette époque, puisqu'il se propose de partir des idées pour accéder aux moyens d'expression. C'est en réaction à la bipartition des verbes en verbes d'états opposés aux verbes d'actions et au sujet de laquelle il formule deux critiques (*i. e.* tous les verbes ne rentrent pas dans au moins une de ces catégories et un même verbe peut entrer dans l'une ou l'autre des catégories selon la forme à laquelle il est employé) qu'il élabore une théorie générale de l'aspect lexical en français. La classification de Brunot se base sur la formation morphologique des verbes (verbes de production, instrumentaux, de matière, de manière, de mise dans un état, d'entrée dans un état, de renouvellement, négatifs…). Il propose aussi quatre grands types d'actions : instantanées, limitées, illimitées, partiellement limitées et considère que c'est seulement à des énoncés pris dans leur intégralité qu'on peut associer des représentations aspectuo-temporelles, le contexte linguistique pouvant modifier le caractère d'une action.

C'est la redécouverte d'une catégorisation des procès [394] en germe dans la théorie d'Aristote, les verbes de *kinesis* (*performance*) et les verbes d'*energeia* (*activity*) qui est à

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

l'origine de ce type d'analyses dans les travaux contemporains.

Parmi les premiers travaux, la classification de Kenny (1963) repose sur des tests linguistiques explicités, et forme la classification présentée Figure (9-2) :

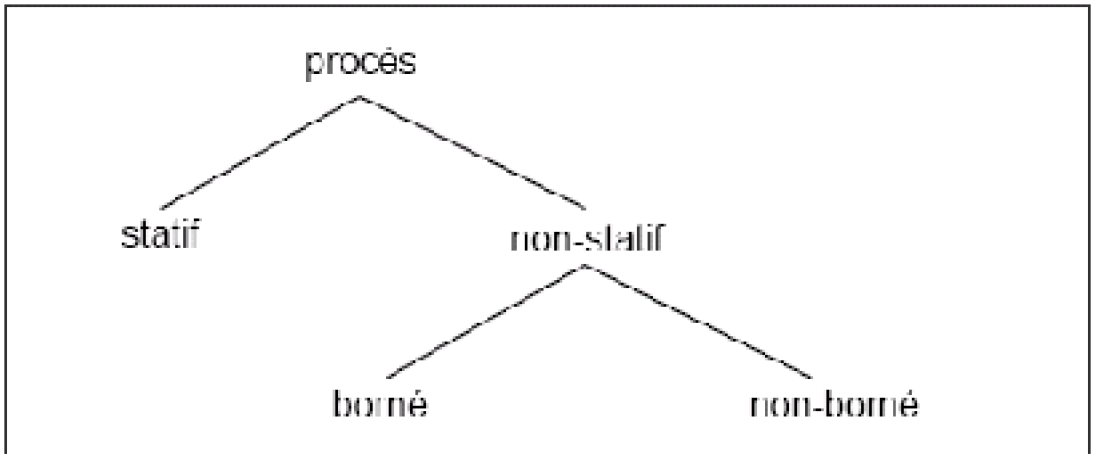

Figure (9-2) – Classification des types de procès selon Kenny (1963)

Les travaux de Kenny portent sur l'anglais. La distinction entre statif et non-statif se fait sur un test de progressivisation, et la distinction entre *performance* et *activity* se fait ensuite par test de compatibilité avec le *pluperfect* (*activity* si *Mary is listening* implique *Mary has listened ;* et *performance* si *Mary is building a house* n'implique pas *Mary has built a house*).

Selon Vendler (1967), les procès (lexicalisés par des verbes) peuvent être classés en quatre groupes : *states*, *activities*, *accomplishments*, *achievements*, les verbes des deux premiers se caractérisant comme non-téliques, et des deux derniers comme téliques. Ainsi, les quatre classes fondamentales de types de procès sont présentées dans le tableau (9-2).

Tableau (9-2) – Types de procès d'après Vendler

	dynamique	borné	Ponctuel
States	-	-	-
Activities	+	-	-
Accomplishments	+	+	-
Achievements	+	+	+

Cette analyse revient à scinder la classe des verbes de *performance* de Kenny en deux selon le caractère plus ou moins ponctuel d'une action.

La combinatoire complète, sur laquelle il faudrait s'interroger afin de se demander si réellement 3 traits sont nécessaires à la classification des types de procès est : ---, --+, -+-, -++, +--, +-+, ++-, +++, soit 8 types. L'impossibilité rencontrée pour certaines de ces

[394] Notons que le terme de procès est souvent utilisé à défaut de mieux comme désignant les représentations sémantiques associées aux prédicats verbaux (et aux substantifs déverbaux).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

combinaisons (par exemple, procès non dynamique, non borné, et ponctuel) remet en cause le choix des traits.

Dowty (1977, 1979, 1986), et Bennett et Partee (1978) dégagent 3 classes aspectuelles : les états, les *activities*, et les *accomplishments/achievements.* La différence majeure entre les analyses de Dowty et de Vendler, est comparablement à la distinction entre les analyses de Kenny et Vendler, qu'il n'y a pas de distinction entre *accomplishment* et *achievement*. D'autre part, chez Vendler, l'accent est mis sur les verbes seulement, tandis que Dowty propose que le fait d'être une *activity*, un *accomplishment*… est une propriété de la proposition entière qui peut être déterminée par les syntagmes nominaux sujet ou objet, par la présence ou l'absence de syntagmes directionnels, ou le choix d'adpositions (pour l'anglais, l'opposition entre *to* et *toward*)…

Toujours dans le domaine des classification de procès, on trouve, à partir des travaux de Culioli [395] sur les substantifs, une analyse effectuée par des représentants de son école, des types de procès en 3 classes : discret, dense, compact. Les procès discrets reçoivent une délimitation qualitative et quantitative, i.e. il existe un état résultant du procès, et le procès est dénombrable. Les procès denses sont discrétisables, et c'est l'objet qui discrétise. Enfin, les procès compacts ne sont pas discrétisables. Cette analyse semble revenir à la tripartition originelle entre *states*, *activities*, *performance.*

Talmy (2000:68) propose également les critères de classification sémantiques qui se résument par le schéma suivant.

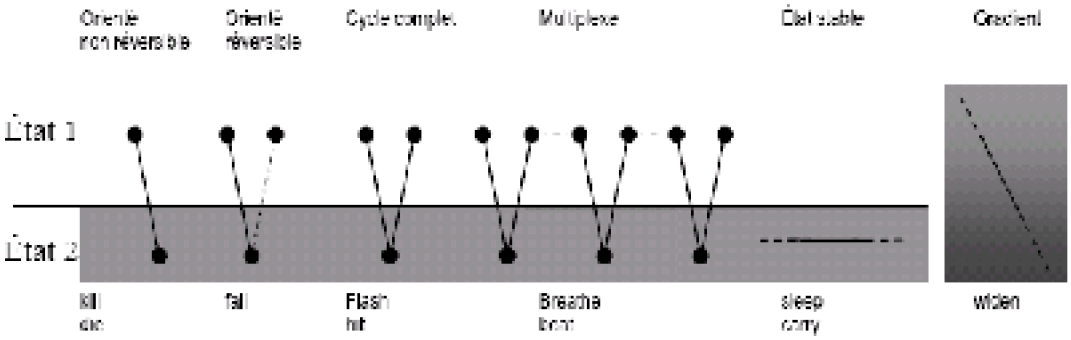

Figure (9-3) – Sémantique de l'aspect selon Talmy (Talmy 2000b:68)

Ces différentes observations montrent la diversité des conceptions de l'aspect au point de vue sémantique. Au sujet de ces classifications, une des difficultés principales est de définir des critères de validation pour l'attribution d'une catégorie (en l'occurrence un type de procès) à un verbe. Les linguistes utilisent pour ce faire l'intuition immédiate, ou des tests linguistiques, ce qui ne va pas sans poser de difficultés. Comment distinguer ce qui relève du sémantisme du verbe de ce qui relève de caractéristiques de la construction utilisée pour le test ? Et comment échapper à la circularité de l'entreprise ?

Gosselin (1996:41-47) présente les tests linguistiques les plus couramment utilisés pour essayer de classer les procès en trois ou quatre catégories :

[395] Culioli (1980), (1990).

en vertu de la loi du droit d'auteur.

Tableau (9-3) – Principaux tests linguistiques pour décider de l'aspect lexical d'un verbe.

Hypothèse testée	Moyen linguistique	problèmes
Dynamicité	être en train de Vinf	Ce test est incompatible avec les *achievements* bien qu'ils soient dynamiques. *? La bombe est en train d'exploser*
Caractère atélique	*+pendant* + GNdurée Si PC compatible, non-borné.	Des glissements possibles peuvent conduire à des interprétations de réitérations *? Il a atteint le sommet pendant des heures*
Caractère télique	+ *en*+durée	Des glissements peuvent affecter états et activités (procès non-bornés). Dans ce cas n'est retenue que le changement initial, la phase qui précède immédiatement ce changement initial. *Il a marché en une heure*
ponctualité	*Mettre n temps à/pour Vinf*	L'utilisation d'un circonstanciel ponctuel n'est pas un test fiable, dans la mesure où les glissements consistant à ne retenir que l'une des bornes sont réguliers. Indique la durée du procès pour les *accomplishments* Indique la durée précédant la culmination du procès pour les *achievements*

Un point essentiel ressort de ces différentes analyses : la nature compositionnelle de la spécification temporelle. Si on observe le fonctionnement des types de procès avec les différents temps et aspects verbaux, on s'aperçoit que les phénomènes de glissement de sens sont fréquents et réguliers. L'aspect dépend du type de procès, de l'objet, du sujet du verbe, du tiroir verbal employé, d'adverbiaux…

Cette section nous a permis d'une part de présenter les phénomènes linguistiques que nous considérons comme relevant de la temporalité, et d'autre part de dresser l'inventaire des différents types d'unités concernées dans notre corpus.

Nous envisagerons plus précisément, dans la section suivante, les moyens de la référence temporelle de nos corpus.

9.3. La référence temporelle dans les corpus

Nous avons présenté la proximité des catégories conceptuelles de temps, mode, aspect. Ainsi que le remarque Gosselin (1996:23) "les marques temporelles et aspectuelles se répartissent sur divers éléments de l'énoncé (le verbe, le temps verbal, les compléments du verbe, les circonstanciels, les constructions syntaxiques, etc.) qui paraissent interagir les uns avec les autres de telle sorte que la valeur de certains marqueurs semble ne pouvoir être fixée indépendamment du calcul global de la valeur du tout". Cet état de fait a été repéré pour plusieurs phénomènes dans la littérature.

en vertu de la loi du droit d'auteur.

Dans cette section, nous présentons les valeurs des temps (morphologie verbale) usuellement relevées dans la littérature en notant leur occurrence dans le corpus. Nous relevons ensuite les phénomènes combinatoires principaux qui surviennent dans le corpus, afin de déterminer le mode de la référence temporelle particulier à nos corpus.

9.3.1. Valeurs des temps

La littérature relève des valeurs pour les différents temps. La perspective structuraliste [396] a fait naître l'idée de considérer que les temps verbaux font système [397] . Celle-ci a été développée par la perspective textuelle et conduit à reconnaître des sous-systèmes en fonction des types de textes.

Nous commencerons ici par dessiner les valeurs rencontrées dans nos corpus pour les tiroirs verbaux, simples puis composés, d'après la synthèse proposée dans Touratier (1996) [398] .

9.3.1.1. Présent et catégories sans marques formelles

Le présent est le tiroir le plus largement employé dans nos corpus. S'il peut indiquer les trois époques (passé, présent, futur), il est toutefois le seul qui s'applique à l'actuel, la simultanéité de l'événement dénoté par rapport au moment de l'énonciation étant rarement stricte [399] . L'intervalle de temps occupé par le procès dénoté déborde fréquemment de l'intervalle du dire du côté du passé et/ou du futur [400] . Le présent peut aussi revêtir des valeurs de passé, récent (11) ou non (12) (présent de narration), et des valeurs de futur, proche (13) ou non (14).

(11) Je sors de réunion à l'instant (12) En 1920, Pierre achète une maison puis se marie (13) Je viens dans deux minutes (14) Dans 50 ans, la mer c'est une poubelle.

On relève également des valeurs non temporelles, telles que le présent dit d'habitude, ou de répétition (dans le passé ou dans l'avenir, mais pas dans le présent) et le présent gnomique, permanent ou pantemporel [401] .

Dans nos corpus, on trouve les valeurs de passé (Extrait du Corpus P6) [402] , de futur

396 Ducrot et Wahl (1968).

397 Benveniste (1974) repris et exploité par Weinrich (1973). On peut citer également les travaux de Martin (1971), (1987), Vet (1980), Klum (1961), (1975).

398 Cf. section précédente pour une présentation.

399 Kahn (1954:59) cité par Touratier (1996:74). Les cas de simultanéité stricte nécessaires sont ceux qui manifestent des performatifs de première personne au présent (*je promets*).

400 Guillaume (1929).

401 Kahn (1954:65) cité par Touratier (1996:77).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

(Extrait du Corpus F5), de présent contemporain du moment de l'énonciation (Extrait du Corpus F1), d'habitude (Extrait du Corpus F7), et le présent gnomique (Extrait du Corpus F2).

Extrait du Corpus P6 *PA1: jleur demande si ils font tourner l'adsl sur macosx il me reponde aplé france telecom Extrait du Corpus F5 *PEB: mais demain j'ai bac blanc de francais!!! Extrait du Corpus F1 *PBY: j'ai des petites questions sur la France Extrait du Corpus F7 *PJQ: le vendredi il parte plus tôt :) Extrait du Corpus F2 *PBT: les pb relationnelles entre un aop et un user le doiventy pas se terminer pas une shitlist

On trouve encore un présent fictionnel, ou "magique". Ce dernier se rapproche du présent coïncidant nécessairement strictement au moment de l'énonciation.

Extrait du Corpus P7 *(action): PBH fait de gros bisous a PAK :))

Nous reviendrons plus précisément sur ce cas dans la deuxième section de ce chapitre.

Le présent est de très loin le tiroir le plus utilisé dans nos corpus. Il reçoit, nous le verrons, ses différentes interprétations par défaut [403] , ou de la combinaison avec d'autres éléments de l'unité phrastique.

9.3.1.2. Morphèmes de temps : passé simple, imparfait

Les morphèmes dits de temps sont ceux du passé simple et de l'imparfait.

Pour le passé simple, la notion de passé factuel est proposée par Touratier (1996:101) pour rendre compte de ses différents emplois, c'est-à-dire que l'accent est mis sur le fait que l'événement dénoté appartient au passé et qu'il est envisagé comme un fait parfaitement délimité. Ainsi a-t-on pu dégager une valeur textuelle de ce tiroir verbal, permettant de marquer les événements appartenant à la trame événementielle dans une narration.

Une valeur gnomique est encore attribuée au passé simple dans certains énoncés (15).

(15) Qui ne sait se borner ne sut jamais voyager (Boileau, Art poétique) [404]

Cette dernière valeur ne se présente pas dans nos corpus.

Le passé simple, si son usage semble limité à des circonstances particulières n'en est pas moins vivant [405] . On en trouve quelques exemples dans nos corpus. Il s'agit d'énoncés indépendants et isolés, ainsi que l'illustre l'extrait suivant :

Extrait du Corpus P9 *PBF: Tout d'un coup une secheresse equivoque toucha le channel

[402] Présent de narration.

[403] Nous serions également tentée de dire, qu'un système d'inférence sur les intentions d'autrui et/ou de reconstruction de la pertinence d'un énoncé sont à l'œuvre dans l'interprétation de certains emplois de temps.

[404] ***Exemple exposé par Touratier (1996:107).***

[405] Blanche-Benveniste (1985).

 Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Ce type d'énoncés produit une rupture. Il s'agit d'un changement de type de discours. Ces énoncés font en effet passer le texte dans le système du récit, toutefois, il est évident que l'intention du locuteur n'est pas de poursuivre une narration. Ces énoncés surviennent lorsqu'un événement vécu comme négatif survient sur le canal (p. ex. absence de messages, conflits). La fonction de ces énoncés semble être de détourner l'attention des participants de l'événement négatif. Mais nous ne nous prononcerons pas davantage sur l'emploi de ce tiroir dans nos corpus, dans la mesure où ses occurrences sont rares.

En ce qui concerne l'imparfait, ses utilisations ont été recensées dans leurs aspects textuels et discursifs. L'imparfait, selon Touratier (1996:107-108) "doit être associé à un signifié non proprement temporel comme "non actuel" plutôt qu'au signifié "passé" […] si l'on entend rendre compte de façon un peu rigoureuse de tous ses emplois". En effet, il est sans doute le morphème qui présente le plus de diversité. Ses emplois temporels permettent de référer au passé, ils sont distingués selon qu'ils apparaissent dans le récit ou dans le discours [406] . L'imparfait est alors en opposition avec le passé simple qui est factuel, délimité. La valeur de passé non délimité est selon Touratier (1996:111) "une conséquence possible des oppositions dans le système verbal français et dans les contextes d'emploi".

Les valeurs de l'imparfait ont été très étudiées pour ce qui concerne le récit. Weinrich (1973) en fait le centre du plan du commentaire, servant la mise en place d'un cadre pour la trame événementielle, ou au cours de la présentation de cette trame.

L'imparfait qu'on trouve dans des complétives associées à un verbe au passé est dit de concordance.

(16) je sentis que je l'aimais toujours. (Racine, Andromaque I, 1.) [407]

Mais l'énoncé (17) [408] montre que le contenu propositionnel de la subordonnée est totalement intégré au procès passé introduit par le verbe principal.

(17) Vous avez dit que j'étais là (Courteline, Coco, Coco et Toto, M. Felix, p.70) [409]

De plus, la concordance n'est pas toujours respectée, le second verbe se réalisant parfois au présent.

L'imparfait est encore relevé comme événementiel : imparfait d'événement, ou de rupture. Ces derniers ont attiré l'attention des grammairiens dans la mesure où ils

406 Benveniste (1974) postule, pour le français, la coexistence de deux systèmes distincts complémentaires dans l'emploi des temps. Il s'agit de l'opposition récit/discours. Cette opposition, établie en prenant en compte la dimension énonciative, fait éclater l'apparente homogénéité des temps de l'indicatif. Le système du discours implique un embrayage sur la situation d'énonciation, tandis que le système du récit en est dépourvu. Un certain nombre d'indices linguistiques sont alors déterminés comme appartenant à l'un ou à l'autre des systèmes (ex. le discours comporte des embrayeurs, déictiques… tandis que le récit n'en manifeste pas). L'examen de textes divers (et notamment littéraires – Maingueneau (1986)) montre cependant que les emplois de ces systèmes s'imbriquent ou qu'on peut imaginer des degrés de l'un à l'autre.

407 Exemple relevé par Touratier (1996:116)

408 Initialement relevé par Damourette et Pichon

409 Exemple relevé par Touratier (1996:116)

en vertu de la loi du droit d'auteur.

présentaient une incohérence avec la valeur durative communément attribuée à l'imparfait.

(18) À 8 heures, la bombe explosait

Dans le discours, les valeurs relevées sont multiples. Les emplois temporels peuvent signifier le passé. Toutefois, Le Guern (1986) montre que dans certains cas, l'événement est présenté comme ayant été vrai dans le passé, mais que sa vérité est niée pour le moment de l'énonciation à moins d'une mention explicite indiquant le contraire.

(19) À l'époque, Pierre vivait à la campagne (20) À l'époque, Pierre vivait à la campagne, comme maintenant.

Berthonneau et Kleiber (1993:57) affirment de plus que "L'imparfait ne localise pas lui-même la situation qu'il introduit". L'imparfait est ainsi qualifié métaphoriquement de temps anaphorique. En effet, selon Berthonneau et Kleiber (1993:66) "son interprétation exige toujours la prise en compte d'une situation temporelle du passé, donc d'un antécédent, explicite ou implicite".

Sont encore relevées des valeurs temporo-aspectuelles, où l'imparfait est dit apporter une information de durée, ou encore d'habitude.

(21) Tout ce jour là, nous ne marchions pas, nous courions. [410] ***(22) Pierre fumait après chaque repas.***

Mais dans les exemples (21) et (22), ainsi que le remarque Touratier (1996:132-134) l'information est portée par un autre constituant (*tout ce jour là, après chaque repas*).

Les auteurs relèvent encore des valeurs modales pour l'imparfait, que Touratier préfère qualifier de non temporelles. Il s'agit de l'emploi de l'imparfait qui présente le procès comme ne relevant pas du monde réel. La possibilité est ainsi exprimée dans des conditionnelles.

(23) Si le train partait en retard, j'aimerais que tu m'appelles.

Les imparfaits reconnus comme présentant l'irréel du présent (24), ou l'irréel du passé (25), le sont comme présentant ces valeurs par le seul fait que le contexte montre l'impossibilité du procès dénoté à l'imparfait.

(24) Un peu plus, il manquait son train (25) Si je ne l'avais pas rappelé, il partait

Enfin, sont identifiés des imparfaits de politesse (26) et hypocoristiques (27).

(26) Je voulais des fraises (27) Il faisait des risettes à sa maman le bébé

On rencontre encore des emplois qui se distinguent de ceux évoqués jusqu'ici :

(28) Et en plus il tournait.

Cet énoncé, émis par le conducteur d'un véhicule qui venait d'en doubler un premier, suite à une maladresse commise par le premier conducteur, semble intéressant à plus d'un titre. Au moment où l'énoncé est produit le véhicule doublé est en train de tourner. L'explication qui recourt à la position de Kleiber et Berthonneau semble la plus valide. Ce que signifie le locuteur, c'est que l'autre conducteur savait qu'il allait tourner au moment où il a commis la maladresse.

De même, l'énoncé suivant, prononcé par un passant s'adressant à un autre au sujet

[410] ***Exemple relevé par Touratier (1996:132).***

en vertu de la loi du droit d'auteur.

d'un appartement avec terrasse devant lequel se prépare la construction d'un immeuble, semble mettre en défaut l'analyse de l'imparfait.

(29) Ils avaient de la lumière.

Il se trouve que les habitants de l'appartement en question ont toujours et encore pour quelques mois de la lumière puisque les fondations de l'immeuble ont à peine commencé au moment du dire. La définition donnée pour l'imparfait de discours est ici en défaut. L'événement est présenté comme ayant été vrai dans le passé, sa vérité n'est pas niée pour le moment de l'énonciation mais pour un futur prochain.

Dans nos corpus, l'imparfait a des valeurs limitées. Toutes les valeurs évoquées ici ne se présentent pas et elles n'ont pas toutes égale importance dans les corpus. Les valeurs proches de l'imparfait de discours sont les plus usitées (Extrait du Corpus F1). On trouve également quelques imparfaits de politesse (Extrait du Corpus P1). Les autres emplois de l'imparfait naissent de l'interaction avec d'autres éléments de l'énoncé (conditionnelles).

Extrait du Corpus F1 *PAD: ca me mnquait de pas me faire virer lolol Extrait du Corpus P1 *PAY: je me demandais combien est-ce que je devais gagner pour avoir la belle vie à paris

Ce qui est commun à tous les emplois de l'imparfait dans notre corpus est qu'ils présentent les événements d'une façon non bornée.

9.3.1.3. Morphèmes d'aspect : passé composé, plus-que-parfait, passé antérieur, temps surcomposés.

Le passé composé est reconnu recevoir des valeurs temporelles et non temporelles. Les emplois non temporels du passé composé relevés par Touratier (1996) sont en fait des emplois dans lesquels seule la valeur aspectuelle du tiroir serait décodée. Le procès serait alors présenté comme achevé sans être situé dans le temps.

(30) Quiconque a beaucoup vu peut avoir beaucoup retenu (Lafontaine, Fables, I, p.8) [411]

Toutefois, ce qui est achevé est presque nécessairement antérieur à un moment de référence (pour dire que quelque chose est achevé, il faut qu'un intervalle de temps si petit soit-il, se soit écoulé, permettant le constat). Ainsi le passé composé correspond-il souvent à une antériorité.

Les emplois temporels relevés réfèrent à l'époque *passé*. On relève également dans la littérature des passés composés gnomiques, mais ces derniers s'appuient sur d'autres constituants à référence temporelle.

Dans nos corpus, les passés composés reçoivent essentiellement des valeurs temporelles. Ils présentent un procès borné antérieur au moment de parole (extrait du Corpus F2).

Extrait du Corpus F2 *PAL: PAG il a abuse de son pouvoir et s'est entete dans ses conneries

[411] ***Cité par Touratier (1996:146).***

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

La littérature relève, pour le plus-que-parfait, tiroir combinant les morphèmes portant les signifiés d'accompli et de non actuel, des valeurs temporelles et non temporelles.

Un procès achevé dans le passé peut être dénoté par un plus-que-parfait, ou encore, l'antériorité d'un procès par rapport à un procès passé. Dans le récit, le plus-que-parfait ne concernerait que les procès qui n'appartiennent pas à la trame événementielle.

L'irréel du passé se présente pour le plus-que-parfait dans le contexte des subordonnées conditionnelles attachées à un verbe au conditionnel passé, ou à l'imparfait.

(31) Si il avait mangé tout le gâteau, il aurait été malade. (32) Si je n'avais rien dit, il venait.

Le plus-que-parfait n'aurait qu'une valeur temporelle d'antériorité dans les énoncés de structure similaire mais présentant un verbe principal au conditionnel présent.

(33) Si il avait mangé tout le gâteau, il serait malade.

Dans cette structure, il aurait une simple valeur d'accompli, lorsque le verbe principal engage l'avenir. Cette seule valeur est également reconnue lorsque le plus-que-parfait est construit avec *comme si*.

(34) Si demain vous aviez perdu toute votre fortune, que feriez-vous (Sten, 1952:225 [412])

Dans nos corpus, excepté les valeurs contribuées par les conditionnelles, les plus-que-parfaits ne répondent évidemment pas aux valeurs identifiées pour le récit. Ils dénotent effectivement l'antériorité d'un procès par rapport à un procès passé dans certains énoncés (Extrait du Corpus P9).

Extrait du Corpus P9 *PAB: g failli lui gerbé dessus %add: PAZ *PAZ: lol PAB %add: PAB *PAB: rigol pas heureusement k'eel mavé pressé un jus d'orange :p %add: PAZ *PAB: g pu rincer %add: PAZ

Mais cette valeur se rapproche parfois de celle avancée pour l'imparfait de discours. L'événement au plus-que-parfait est posé comme ayant été vrai dans un passé antérieur à un événement souvent implicite après lequel l'événement au plus-que-parfait n'est plus vrai (Extrait du Corpus F1).

Extrait du Corpus F1 *PZZ: PAD donne un susucr a PCJ lol gentil garcon gentil ;) %add: PCJ *PCJ: j'av lu un suçon lolololololololol %add: PAD

Le passé antérieur, lui, combine le morphème d'accompli avec le morphème de passé factuel. Il en résulte plusieurs valeurs possibles. Soit le morphème d'accompli n'est pas appliqué à la localisation temporelle, et le passé antérieur dénote un procès accompli dans le passé qui appartient à la trame événementielle dans le récit. Toutefois, il semble que le plus souvent, le morphème aspectuel soit appliqué à la localisation temporelle relative, c'est-à-dire qu'il permet de spécifier un procès comme s'étant produit antérieurement à un autre, mais appartenant à la trame événementielle.

(35) La cigale ayant chanté tout l'été se trouva fort dépourvue quand la bise fut venue.

Pour cet exemple, Touratier (1996:158) précise que "Lorsque la proposition au passé

[412] ***Cité par Touratier (1996:157).***

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

antérieur vient en second, le procès antérieur est simplement signalé comme étant une authentique donné factuelle spécifique, sans pour autant être intégré à la trame événementielle".

Dans nos corpus, l'utilisation du passé antérieur est très sporadique et correspond à un changement ludique de niveau de langue.

Extrait du Corpus F6 *PCW: PBF j'eûs aimé une réponse %add: PBF

Les temps surcomposés, selon l'analyse proposée par Touratier (1996) présentent deux marques d'accompli. Ainsi, l'une se spécifierait dans la dénotation de l'antérieur tandis que l'autre marquerait l'accompli, le tout dénotant finalement un procès achevé antérieur à un autre procès passé. En français standard, les emplois des temps surcomposés sont essentiellement relevés dans des subordonnées temporelles ou en combinaison avec des spécifications lexicales de brièveté.

(36) Dès qu'il a eu mangé, il s'est vautré devant la télé. (37) Il a eu vite fait de virer sa femme.

Pour des variétés régionales du français (notamment du sud de la France) les auteurs relèvent un usage moins limité, le premier morphème d'accompli apportant la valeur temporelle de passé et le second celle d'antériorité [413] , ce qui aurait pour conséquence de marquer un passé éloigné ou lointain.

(38) La marseillaise, je l'ai eu sue par cœur.

Mais la spécification du degré d'éloignement n'est pas interprétable dans tous les cas, l'exemple de Foulet (1925:227) faisant dire à Touratier (1996:161) qu'alors "il n'est guère possible en effet de parler de recul dans le passé" mais simplement d'un "accompli dans le passé".

(38) J'ai eu fini de bonne heure aujourd'hui. (Foulet 1925:227)

Nous n'avons relevé aucune occurrence de formes surcomposées dans nos corpus.

9.3.1.5. Morphèmes de mode : futur simple et futur périphrastique, conditionnel, subjonctif, impératif.

Touratier (1996:176) propose d'attribuer le signifié de "projeté" au morphème de futur. Plusieurs tiroirs le présentent : le futur simple, le conditionnel, le futur antérieur et le conditionnel passé. Nous présenterons également ici les valeurs du futur dit *prochain* ou *périphrastique* se réalisant par un morphème à signifiant discontinu *aller*+prés. Vinf.

Le futur simple connaît des emplois temporels et non temporels. Il indique que le locuteur projette ou prévoit la réalisation d'un procès après le moment où il parle. Il s'agit là de la valeur intuitivement la plus courante, cependant, la référence par rapport à laquelle le procès dénoté par un verbe comportant le morphème de "projeté" est dit ultérieur n'est pas toujours le moment de parole. Touratier (1996:177) précise qu'il "peut faire suite à un présent employé dans un énoncé de vérité générale […] ou à un impératif de portée générale […] [correspondant alors à] une conséquence logique. Il peut faire suite aussi à un présent dit de narration […] ou même à un passé simple dans ce qu'on appelle le futur historique". Il connaît donc des emplois absolus ou relatifs.

[413] Blanche-Benveniste (1977:144-145) cité par Touratier (1996:164).

en vertu de la loi du droit d'auteur.

Parmi les autres emplois on distingue des emplois qui se rapprochent de ceux du morphème de volonté, exprimant l'ordre, la prière. Mais on peut reconnaître avec Touratier (1996:179) que ce n'est pas le morphème de projeté qui contribue ces valeurs, mais bien le contexte plus ou moins large de l'énoncé.

Le futur simple peut être utilisé pour exprimer une vérité générale, il est alors dit pantemporel ou gnomique.

(39) Qui vivra verra

Certains futurs sont employés pour dénoter des procès actuels, voire passés.

(40) C'était quoi ce bruit? Ce sera le chat qui a encore renversé la corbeille

Ces emplois sont dénommés futurs de conjecture.

On reconnaît encore des futurs d'atténuation et de protestation

(41) Je ne vous cacherai pas que je me suis un peu ennuyée à cette soirée (42) Quoi, je souffrirai, moi, qu'un cagot de critique Vienne usurper céans un pouvoir tyranique (Molière, Le Tartuffe, I, 1.) [414]

Leur particularité est que tous les deux concernent des procès actuels.

Enfin, le futur simple est dit d'éventualité lorsqu'un locuteur présente une objection possible mais réfutable.

(43) C'est là, me direz-vous, une erreur, mais on peut envisager les choses autrement.

Dans nos corpus, les emplois du futur dit simple sont toujours temporels et projettent ou prévoient la réalisation d'un procès après le moment de parole.

Extrait du Corpus F5 *PAB: PCZ je ne sais pas si il y a une réelle efficacité..... mais chose sure..... c'est que ca rendra ton rhume plus joyeux :P %add: PCZ

Le futur périphrastique est dit prendre seulement des valeurs proprement temporelles. Il concerne fréquemment un procès projeté dans un avenir proche, mais il ne s'agit pas là d'une contrainte stricte.

(44) Pierre,il va prendre sa retraite dans trois ans.

De plus, Touratier (1996:181-182) remarque que le futur périphrastique peut s'appliquer à des procès qui ont une étendue temporelle importante.

(45) Ça va durer dix-sept ans comme ça. (Labiche, Le cachemire XBT, sc. 7.) [415] **.**

Il propose alors, reprenant la terminologie de Guillaume, de qualifier le morphème *aller*…Inf de futur efférent. Ceci englobe les emplois où le morphème se combine à celui de l'imparfait et forme une postériorité efférente.

(46) Il allait faire ses devoirs quand Marie est arrivée.

Ces énoncés véhiculent en général une présupposition sur l'événement dénoté par le verbe modifié par *aller*+impft…Vinf. Cet événement en effet est présupposé ne pas s'être déroulé comme il était attendu.

Dans nos corpus, les emplois du futur périphrastique concernent le plus fréquemment

[414] ***Exemple relevé par Touratier (1996:181)***

[415] ***Exemple relevé par Touratier (1996:182).***

en vertu de la loi du droit d'auteur.

des événements dont la réalisation dans le futur est plus que probable et dépendent souvent de la volonté du locuteur (acte de langage indirect de la famille des ordres).

Extrait du Corpus F5 *PBP: haaaaaaaaaa MAIS JE VEUX PAS TRAVERSER LA RUE MOI! %add: PDH (2) *PDH: sisi tu vas traverse la rue %add: PBP

Pour ce qui concerne les formes *aller*+impft…Vinf, on peut constater, dans nos corpus, qu'elles ne présentent pas toujours la valeur mentionnée plus haut. Ainsi, dans le discours indirect, la complétive objet (Extrait du Corpus F8) qui intègre un *aller*+impft…Vinf ne peut être analysée comme portant la présupposition.

Extrait du Corpus F8 *PDT: bon... nous disions donc que PBO allait faire la vaisselle ;)

Des contraintes de concordance imposent en effet l'utilisation de ce tiroir, et dans ce cas la présupposition ne se réalise pas sur l'événement mais sur le dire.

En ce qui concerne le conditionnel, certaines grammaires récentes en accord avec la proposition de Guillaume, le considèrent comme un temps plutôt que comme un mode. Il est d'une part constaté que l'emploi du conditionnel n'a pas toujours valeur modale et d'autre part, que d'autres temps peuvent prendre des valeurs modales, sans qu'on parle de mode à leur sujet (ex: imparfait pour la supposition, l'hypothèse). Le conditionnel pourrait donc être envisagé comme temps ayant la possibilité d'exprimer des valeurs modales. Cette considération est appuyée par le fait que les diachroniciens du français attribuent la désinence du conditionnel à la forme de *avoir* à l'imparfait, tandis que la terminaison du futur provient du verbe *avoir* au présent (un nouvel auxiliaire apparaît : avoir, qui est à l'origine d'un nouveau futur et du conditionnel ; "cantare habet > chantera" [416]). La position de Touratier, est de le compter parmi les morphèmes de mode, puisqu'il combine le morphème de "projeté" (modal) et celui de non actuel. C'est donc en se basant sur la morphologie que Touratier attribue le statut de morphème de mode.

On compte une valeur temporelle pour ce tiroir. Le conditionnel peut en effet prendre une valeur de futur dans le passé, les deux morphèmes actualisant une valeur temporelle.

(47) À ce moment là j'ai pensé qu'il partirait le lendemain.

Les autres valeurs sont modales, et présentent l'hypothèse. Touratier (1996:183) l'explique par le fait que le "projeté" "indique simplement qu'une hypothèse est envisagée" et le "non actuel" "pris lui aussi dans une acception non temporelle, précise que cette hypothèse est étrangère à la réalité objective".

Des effets de sens tels que l'indignation, l'atténuation ou la politesse lui sont également reconnus.

Dans nos corpus, le conditionnel présente seulement les deux dernières valeurs de façon fréquente.

Extrait du Corpus F7 *PFN: qqun pourrait traduire siouplait? ce seraitbien urbain %add: ALL

Les valeurs combinées de non actuel et de projeté, dans ces cas, permettent en effet le glissement fonctionnel. Elles fonctionnent comme adoucisseur de FTAs.

[416] Picoche et Marchello-Nizia (1998:250).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Enfin les deux derniers tiroirs dont nous allons recenser les valeurs sont le futur antérieur et le conditionnel passé, qui sont les formes composées du futur simple et du conditionnel et leur ajoutent donc le morphème d'accompli.

Le futur antérieur peut marquer l'antériorité d'un procès par rapport à un autre procès projeté.

(48) Je reviendrai quand tu auras appris ta leçon.

Cependant il peut marquer la probabilité s'il est en rapport avec un procès passé.

(49) C'était quoi ce bruit ? C'est le chat qui aura renversé la corbeille.

On lui relève encore des effets de sens de protestation ou d'atténuation.

(50) Tu auras tout manqué par ta bêtise. (51) Tu auras appris que rien ne sert de s'énerver.

Dans nos corpus, le futur antérieur se fait rare et sert seulement l'antériorité d'un procès par rapport à un autre procès projeté.

Extrait du Corpus F1 *PAY: cool PAZ je te dirai ou on sera alle :) %add: PAZ

Quant au conditionnel passé, il peut signifier l'antériorité d'un procès par rapport à un autre présenté au conditionnel et ayant pour valeur le futur dans le passé. Il peut encore formuler une hypothèse sur le passé. Cet emploi est fréquent dans la production journalistique.

(52) Le premier ministre aurait reçu la visite du maire de paris ce matin.

Cette hypothèse peut connaître des degrés de probabilité.

Si une subordonnée conditionnelle ou un syntagme prépositionnel exprime une condition, la valeur du conditionnel passé est alors l'irréel du passé.

(53) Si j'avais eu un manteau, je l'aurais emporté.

Les effets de sens du conditionnel passé sont également la protestation ou l'atténuation.

Il faut constater une ambiguïté quant au conditionnel "passé". Soit il présente l'attitude épistémique comme irréelle (54), soit il manifeste l'hypothèse que fait le locuteur (55).

(54) (Il paraît que) il aurait su que Pierre partait. (55) (Si j'avais téléphoné) il aurait su que Pierre partait.

Quant à la modalité rapportée et le conditionnel, on remarque, dans ces structures complétives, que l'attitude épistémique rapportée (56) (57) est mise en doute, ou soumise au statut d'hypothèse lorsque le verbe épistémique est au "conditionnel", tandis que c'est l'objet de l'attitude épistémique rapportée (58) qui est mise en doute, ou soumise au statut d'hypothèse par le locuteur lorsque le verbe de la complétive porte la marque en -rai.

(56) Il constaterait que Pierre est parti. (57) Il aurait constaté que Pierre est parti. (58) Il constate que Pierre serait parti.

Dans nos corpus, le conditionnel passé présente la valeur d'irréel du passé (extrait du Corpus F7) et l'hypothèse sur le passé se trouve dans des interrogatives seulement (extrait du Corpus P9).

Extrait du Corpus F7 *PAK: tu as raison PIW.... le vrai m'aurait déjà insulté depuis longtemps lol %add: PIW Extrait du Corpus P9 *PBG: je drague?? ah bon??? on m'aurait menti alors??? %add: PAA, PAZ,

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Au sujet du subjonctif, on relève dans la littérature deux attitudes selon qu'il est envisagé de façon unitaire ou non. Pour Guillaume (1929), le subjonctif se caractérise par une image moins élaborée que celle de l'indicatif, ne connaissant que la distinction accompli/non-accompli. Si on se rapporte à l'étymologie du terme *subjonctif* on est tenté d'y voir le mode de la subordination, il peut toutefois être employé dans des unités phrastiques simples ou matrices. Le subjonctif a alors valeur d'impératif ou d'optatif. [417]

(59) Que personne ne sorte ! (60) Que personne ne sorte avant qu'il entre !

Il faut aussi mentionner quelques structures archaïsantes telles que :

(61) "Je ne sache pas que ce travail ait paru." [418]

Les exemples en proposition subordonnée foisonnent également

(62) Je ne crois pas qu'il vienne. (63) Il doute que tu sois tout à fait guérie. (64) Je ne doute pas que tu sois venue. (65) Je suis venue vous dire au revoir avant que vous ne partiez.

En proposition subordonnées, les valeurs relevées pour le subjonctif, sont la possibilité, ou le non-engagement du locuteur sur la vérité de ce qu'il dit. Toutefois, des cas présentent le subjonctif dans des contextes où le locuteur asserte une certitude :

(66) Il n'est pas possible que tu t'en ailles. (67) Il n'est pas possible que les philosophes soient des caméléons. (68) Il est exclu qu'il parte. (69) Il est impossible qu'il prenne le train de 4 heures.

Dans le champ de la négation, de l'interrogation, de l'hypothèse, l'indicatif et le subjonctif sont en concurrence.

Touratier (1996:167) propose de considérer que pour le subjonctif on doit reconnaître différemment les cas où il y a un choix de la part du locuteur (commutation possible avec l'indicatif) de ceux où il est obligatoire d'une part, et d'autre part, qu'il existe pour le premier ensemble deux morphèmes homonymes : morphème de volonté et morphème de possibilité.

Le morphème de volonté se trouve dans les unités phrastiques simples, matrices ou dans des subordonnées. Une valeur de supposition peut être dérivée de cette première valeur.

(70) Qu'il me fasse attendre encore un quart d'heure et je m'en vais. (Musset) [419]

Le morphème de possibilité se rencontre essentiellement en unité phrastique enchâssée. Ainsi dans une relative, le subjonctif permet de caractériser différemment le nom auquel elle est rattachée. En (71), l'appartement est asserté exister tandis qu'en (72) il est supposé qu'il existe.

(71) Je cherche un appartement qui a un balcon. (72) Je cherche un appartement qui ait un balcon.

[417] Ces deux notions sont assez proches. Ce qui différencie optatif et impératif, c'est que l'acte dépend ou non de la volonté humaine.

[418] ***Exemple de F.Brunot (1905). On trouvera d'autres exemples dans Grevisse et Goosse (1993:1266§865).***

[419] ***exemple relevé par Touratier (1996:168).***

en vertu de la loi du droit d'auteur.

Cette valeur de possibilité est également recensée dans les consécutives au subjonctif.

Quand le subjonctif ne commute pas, on ne peut le reconnaître comme un morphème de mode, il est alors un simple indicateur d'enchâssement faisant partie du signifiant discontinu *que*… subj., puisque son apparition est limité à un inventaire défini de subordonnées. Toutes les complétives ne requièrent bien sûr pas le subjonctif, ce sont les verbes des unités phrastiques matrices qui commandent ce phénomène. Ainsi, les verbes de volonté (*vouloir, souhaiter, désirer, permettre, défendre, ordonner, interdire, empêcher,* …), verbes de sentiments (*regretter, s'étonner, se réjouir, être content*…), verbes de crainte ou de doute (*craindre, redouter, douter*), verbes d'opinion en contextes négatifs (*croire, penser,…*) et des impersonnels exprimant la nécessité (*il faut que*…), la possibilité (*il est possible que*…), sont reconnus appeler le subjonctif.

Certaines autres unités phrastiques subordonnées ne permettent pas la commutation du subjonctif avec l'indicatif. Il s'agit des adverbiales exprimant le but, des concessives (introduites par *bien que, quoique, si…que, quelque…que, …*), les temporelles introduites par *avant que* et *jusqu'à ce que*, les oppositives en *sans que* et les conditionnelles en *pourvu que* et *à condition que*.

Il reste à noter au sujet du subjonctif, que si les grammaires présentent imparfait et plus-que-parfait du subjonctif, ne sont en usage courant dans la langue moderne que le présent et le passé (forme composée). La valeur ajoutée par le morphème d'accompli au subjonctif passé peut être celle d'antériorité, par rapport à un procès actuel, passé ou futur, ou celle d'accompli.

(73) J'ai peur qu'il ait perdu les clefs. (74) J'avais peur qu'il ait perdu les clefs. (75) J'aurai peur qu'il ait perdu les clefs. (76) Il faut qu'il ait conduit avant demain.

Dans nos corpus, le subjonctif est essentiellement utilisé dans des constructions où il ne commute pas (Extrait du Corpus F1).

Extrait du Corpus F1 *PCJ: lol bah oui... moi j'ai envie que tu sois toute belle, même sur IRC. y'a pas de raison!

Le subjonctif passé est rare, ainsi que le subjonctif relevant du morphème de volonté.

Au terme de cet inventaire pour les formes verbales, on se rendra compte qu'il est important de reconnaître les interactions qui existent entre les divers éléments d'une unité phrastique pour constituer les valeurs temporelles. Nous consignons ci-après les phénomènes relevés dans la littérature et apportons quelques précisions quant aux interactions qu'on peut relever dans le corpus.

9.3.2. Combinaisons principales concernant la référence temporelle

Les divers éléments d'un énoncés reconnus pour intéragir dans le calcul de la valeur globale pour la référence temporelle sont l'aspect lexical et les tiroirs verbaux, l'aspect lexical et les semi-auxiliaires aspectuels, l'aspect lexical et la transitivité, l'aspect lexical et les voix, l'aspect lexical et la modalité, le tiroir verbal et la négation, les adverbiaux temporels et l'aspect lexical, etc. La liste est longue, et tous les phénomènes n'ont sans doute pas encore été identifiés.

en vertu de la loi du droit d'auteur.

Le jeu entre l'aspect lexical et les tiroirs verbaux a été étudié pour l'anglais, par exemple, par Langacker (1995; 1999:222-229). En français, le bornage du procès déterminerait le choix de l'auxiliaire (*être* ou *avoir*) aux tiroirs composés : les verbes perfectifs se conjugueraient avec *être* et les imperfectifs avec *avoir*. Cette hypothèse, avancée par Guillaume (1929:26-27) entre autres, ne semble pas convenir systématiquement. En effet, certains verbes perfectifs requièrent l'auxiliaire *avoir*. Les verbes pronominaux prennent l'auxiliaire *être*, sans qu'on puisse relier ce fait avec l'aspect lexical du verbe. Enfin, fait relevé par Guillaume (1929:27), certains verbes se conjuguent avec *être* ou *avoir* ce qui selon Guillaume ne pose pas de difficulté à sa théorie puisque "ces verbes comportent, à peu près nécessairement, l'alternance du point de vue *devenir* avec le point de vue *être*".

(77) La tache est disparue. (78) La tache a disparu.

Cependant d'autres verbes présentent un emploi avec *être* ou *avoir* selon que la construction est transitive ou intransitive :

(79a) Il est sorti de la maison. (79b) Il a sorti deux assiettes pour le déjeuner. (80a) Il est monté sur le banc. (80b) Il a monté un sac de ciment.

De plus, la plupart des verbes qui sont dits prendre l'auxiliaire *avoir* ou *être* selon qu'ils expriment l'action ou l'état résultant de cette action : *changer*, *crever*, *enlaidir* ; ne se conjuguent pas véritablement, et il s'agit d'une forme adjective du participe passé :

(81) Il est enlaidi. (82) Il a enlaidi.

Les verbes qui collent a l'hypothèse de Guillaume restent en petit nombre : *aboutir* ; *accourir* ; *apparaître* ; *disparaître* ; *grandir* ; *jaillir* ; *paraître* ; *ressusciter*…

Au sujet de ce type de combinaison, on relève également que le présent et l'imparfait [420] forcent une lecture non télique de verbes téliques. L'imparfait en effet est naturellement compatible avec les verbes atéliques, tandis qu'il est discordant avec les verbes téliques si on lui octroie une valeur aspectuelle en fonction de sa forme simple :

(83) La bombe explosait.

Dans nos corpus, les verbes à l'imparfait dénotent essentiellement des états et activités, si bien que ce type de coercion est rare. On observe toutefois que l'hypothèse n'est pas concordante avec l'imparfait de discours (Extrait du Corpus F7).

Extrait du Corpus F7 *PHQ: PFL alors je donnais une solution sans chanserv

Le passé simple d'un verbe télique présente le procès comme accompli de façon systématique, alors que le passé simple d'un verbe atélique peut présenter le procès comme inaccompli, inchoatif (84).

(84) Dès huit heures, il marcha.

Il s'agirait là d'une incompatibilité, que nous ne rencontrons pas dans nos corpus.

Les combinaisons reconnues entre aspect lexical et semi-auxiliaire aspectuel ont été décrites de la façon suivante.

Un verbe télique peut devenir non télique avec un auxiliaire aspectuel.

(85) Jean a commencé à trouver la solution.

[420] Michaelis (à paraître).

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Ce cas est très rare dans nos corpus (Extrait du Corpus P9).

Extrait du Corpus P9 *PBI: Tu es le sherrif ici? Tu arrive et tu commences a balancer tout le monde

Ce type d'interaction est banal, il découle de la fonction du semi-auxiliaire aspectuel, qui est justement, de saisir un intervalle dans le déroulement d'un procès.

Les semi-auxiliaires présents dans le corpus ne semblent pas interagir avec d'autres éléments de spécification temporelle dans les énoncés. *en train de* produit une saisie du procès dans son déroulement et est combinable avec tous les types de procès. *Commencer à* permet de saisir le procès dans sa phase initiale tandis que *finir de* le saisit dans sa phase terminale.

Extrait du Corpus F3 *(action): P97 est entrain de manger une glace %add: ALL

La question de l'interaction entre l'aspect lexical et le caractère transitif du verbe postule que le contenu aspectuel d'un verbe peut varier avec sa construction (transitive, intransitive), et d'autre part, les déterminants du substantif. C'est depuis Verkuyl (1972) qu'a été exploré le fait que les syntagmes nominaux arguments du verbe ont un effet sur les propriétés aspectuelles de la phrase, selon que ces arguments sont comptables ou non comptables.

Les grammaires constructionnelles, issues des travaux de Fillmore, et principalement représentées actuellement par les travaux de Goldberg, proposent un traitement de l'aspect intégré au niveau d'analyse de la construction.

Les analyses issues de ce courant et d'autres, mettent en avant le fait que l'événement est borné par l'objet sur lequel il s'applique. La notion à laquelle ces travaux font appel est celle de l'exhaustion (*measuring out*). Le traitement de l'aspect est ici lié aussi à la notion de télicité, c'est-à-dire le fait qu'un procès présente en lui même l'indication de sa propre fin. On peut ainsi opposer les verbes *chercher* et *trouver*, *chercher* étant non télique tandis que *trouver* est télique. Le test auquel on recourt le plus fréquemment pour distinguer les procès téliques et non téliques est celui de en + CN durée, et pendant + CN durée, test qui fonctionne dans nombre de langues.

(86a) Il a trouvé le rouleau de scotch en 5 minutes. (86b) ??Il a trouvé le rouleau de scotch pendant 5 minutes. (87a) *Il a cherché le rouleau de scotch en 5 minutes. (87b) Il a cherché le rouleau de scotch pendant 5 minutes.

On remarquera que la lecture non-télique d'un procès spécifié par un verbe reconnu comme télique semble potentiellement plus acceptable que la lecture télique d'un procès spécifié par un verbe reconnu comme non télique.

Le constat que pour un même verbe, présentant la même construction, le test fonctionne pour les 2 alternatives, amène à considérer le phénomène d'exhaustion :

(88) J'ai lu ce roman pendant deux heures (lecture non télique). (89) J'ai lu ce roman en deux heures (lecture télique).

La proposition de Tenny (1994) est que le lien entre syntaxe et sémantique s'il est en partie déterminé par la mise en correspondance de rôles sémantiques et de positions syntaxiques est grandement régulé par des propriétés aspectuelles. La structure aspectuelle est vue comme une sous partie de la structure de l'événement associée à la présence de rôles aspectuels.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

L'aspect dans le lien entre syntaxe et sémantique est la délimitation, c'est-à-dire la propriété 'un événement ayant une fin (borne) distincte, définie et inhérente.

(90) Paul mange une orange.

Ce procès consomme un certain temps, et est terminé quand il n'y a plus d'orange. Selon Tenny, seuls les arguments internes sont concernés par les rôles aspectuels, et parmi ceux-là, seul le complément d'objet direct peut délimiter l'événement. Pour ce complément, la qualité délimité/non délimité (de l'eau, un verre d'eau) influe sur l'aspect. Il existe des verbes qui ont un argument interne mais qui ne délimite pas (statifs et non statifs). Il reste un problème pour traiter les verbes qui sont en construction intransitive, pour lesquels la notion d'exhaustion ne peut donc apporter de solution. À la voix passive, l'argument qui délimite serait le sujet. Ce ne serait donc pas le rôle syntaxique qui porterait la transitivité aspectuelle, mais le rôle sémantique de Patient.

Des phénomènes relatifs à la référence temporelle ont également été relevés dans la combinaison aspect lexical/voix. Les grammaires traditionnelles placent au sommet de la hiérarchie constitutive du système de conjugaison la notion de voix parmi lesquelles elles en distinguent deux ou trois parmi actif, passif, moyen, pronominal ou réfléchi. L'actif exprime que le sujet est agissant, le passif présente l'action comme subie par le sujet. Le moyen exprimerait la part que le sujet prend à l'action, il est intermédiaire entre actif et passif. La voix pronominale pose problème, elle peut être réfléchie ou réciproque. Cette position n'est pas tenable à notre avis, à la lumière des travaux récents sur la notion de voix. Le choix de la voix par le locuteur est seulement envisagé comme gouverné par l'adéquation de la parole à la réalité. Il l'est cependant aussi par le point de vue adopté par le locuteur. Les phénomènes de voix sont caractérisés par l'association régulière d'un changement de la forme verbale à un changement de la valence du verbe.

La notion de voix est liée à la notion de transitivité et cette dernière est liée à la notion d'aspect. Selon Givòn (1984:566-599), il y aurait dans le prototype du passif une affinité avec la valeur aspectuelle de stativisation. Cependant, il ne s'agirait, selon Creissels, que d'un résidu de l'histoire d'un certain passif, issu de formes statives résultatives [421] .

Lorsqu'on observe la mise en relation des phénomènes de voix et le caractère borné/non borné, agentif/non agentif du procès on note qu'il existe deux valeurs du passif selon que le verbe est perfectif ou non perfectif. Ainsi, un Verbe perfectif au passif sans complément d'agent dénoterait un état qui succède à un procès achevé tandis qu'un verbe imperfectif ou perfectif avec complément d'agent dénoterait du temps situé.

Les valeurs modales et l'aspect lexical seraient également à la source de phénomènes de coercion. Selon Le Querler (1996) les constructions *si*+Vimpft, telles que *si je gagnais au loto…* reçoivent des valeurs modales qui sont fonction de l'aspect lexical du verbe :

Ainsi, un verbe atélique forme le potentiel ou l'irréel du présent (91) tandis qu'un verbe télique ne peut former que le potentiel (92).

(91) Si j'avais de l'argent, je partirais. (92) Si gagnais au loto, je partirais.

Les verbes d'activité et d'accomplissement, sont également dits [422] avoir une incidence

[421] Creissels (2000).

en vertu de la loi du droit d'auteur.

sur la valeur modale d'un énoncé :

(93) ce type là, il court pendant des heures. (94) ce type là, il peut courir pendant des heures. (95) ce type là, il te peint un tableau en une heure. (96) ce type là, il peut te peindre un tableau en une heure. (97) ce type là, il te trouve un appartement à Paris en 2 jours. (98) ce type là, il peut te trouver un appartement à Paris en 2 jours.

On notera toutefois que la lecture modale de ces énoncés est contribuée à la fois par l'aspect du verbe et le tiroir et est renforcée par l'adverbial temporel de durée.

9.3.3. Tiroir verbal et négation

Dans nos corpus, on relève que les négations *ne…plus,* et *ne jamais* sont en relation avec le tiroir utilisé.

ne jamais, employé avec le futur découpe un événement dans l'avenir, mais ne spécifie rien en ce qui concerne le moment de parole ou le temps qui précède.

Extrait du Corpus P3 *PAK: rah là là, je viendrait jamais chez toi si tu tiens 1 boutique 1 jour, PAS

Avec le passé composé, le procès nié par *ne…jamais* l'est également, par le biais d'une présupposition, pour le présent.

Extrait du Corpus F1 *PBW: et septembre est magnifique en septembre ;)... comme vous n'avez jamais vu ;)..... une splendeur fabuleuse!!!!!!!

Le tiroir plus-que-parfait n'aurait pas cette validité présente.

Avec un présent, *ne…jamais* construit une interprétation pantemporelle.

Extrait du Corpus F3 *PCA: on sait jamais Extrait du Corpus F6 *PAN: PEP je vais jamais au cine Extrait du Corpus F7 *PEI: PIW je ne mange jamais le midi :)

ne plus, employé avec le tiroir présent émet une présupposition sur le passé. Le procès dénoté a connu sa borne droite en un moment antérieur au moment de parole.

Extrait du Corpus P3 *PAG: il est pas loin %add: PAK *PAG: mais je le vois plus %add: PAK

ne plus, employé avec le tiroir imparfait émet également une présupposition sur le passé. Le procès dénoté a connu sa borne droite en un moment antérieur à un événement souvent implicite antérieur au moment de parole.

Extrait du Corpus P3 *PAM: PAY nah yen avais plis en stock :/

Avec un verbe télique, la lecture du procès est de plus itérative.

Extrait du Corpus F1 *PBW: mdrrrrrrrrrrr PAD ;).... ca faisait un moment que tu ne venais plus... tu cherches à te rattraper on dirait..... du calme ;)

Avec un tiroir futur, *ne…plus* fait une présupposition sur le présent, et produit une assertion pour l'avenir.

Extrait du Corpus F1 *PBW: quand PBD????? si tu attends mai.... y'en aura plus..... Extrait du Corpus F7 *PHQ: tu sera plus ds le noir %add: PBP Extrait du

[422] Le Querler (1996).

en vertu de la loi du droit d'auteur.

Corpus F8 *PBO: dans 2 min je suis plus la...

9.3.4. Adverbes temporels, tiroir et aspect lexical

Nous considérerons ici les cas de combinaison entre tiroir, aspect lexical et adverbes temporels. Nous présenterons les cas de *toujours, encore, déjà* et *tout de suite.*

toujours peut apporter une contribution itérative ou pantemporelle.

Les statifs combinés au présent et à toujours dénotent la continuité de l'événement sur le présent, autrement dit, le procès dénoté par le verbe occupe un portion temporelle antérieure au moment de parole incluant ce dernier.

Extrait du Corpus F1 *PBW: tu lui fais un gros bisou de ma part ;) ... sa blonde est toujours là ? Extrait du Corpus F1 *PBW: P96 est toujours là????????? 6 Extrait du Corpus F7 *PKA: PAK TU PUES TJS? :) %add: PAK Extrait du Corpus F7 *PJQ: PIW il a toujours été ainsi %add: PIW

Cependant, des spécifications temporelles additionnelles peuvent faire basculer l'interprétation du côté de l'itération, ainsi que l'illustre l'extrait suivant, dans lequel *le vendredi* spécifie un moment cyclique.

Extrait du Corpus F7 *PFN: a va tjrs le vendredi %add: PIW Extrait du Corpus P2 *PAG: on passe toujours apres un autre homme qui a dit les memes futilités a cette femme qu'on venere

Ainsi, la contribution de *toujours* ne semble pas strictement liée à l'interaction avec le tiroir mais davantage à la qualité aspectuelle du verbe.

Extrait du Corpus P3 *PAK: oui, avec option climatisation interne, pour que les branches des lunettes soient toujours à la bonne température Extrait du Corpus F7 *PJQ: je me suis toujours demandé comment on pouvait mettre la farce dedans sans faire de trous

Éventuellement *toujours* peut être intégré à un adverbial temporel introduit par une préposition.

Extrait du Corpus P8 *PAW: PBI je suis sans merci depuis toujours :)

Il n'a alors pas la même incidence, nous le verrons plus loin.

encore combiné à un état dénote la continuité.

Extrait du Corpus F1 *PAS: ben si la personne etait encore connecte je suppose que oui

Toutefois, si l'état n'est pas permanent, *encore* fait une présupposition sur la fréquence de cet état dans le passé. L'état dénoté s'est manifesté plusieurs fois dans un passé antérieur à celui dénoté pour l'événement.

Extrait du Corpus F1 *PBW: hummmmmmmmmm PAG.... t'étais encore bourré on dirait!!!!!! attention... tu vas rater ton bus!!!!!!! %add: PAG

Avec un passé composé et la négation, *encore,* indique que le locuteur fait l'hypothèse ou le vœu que l'événement nié pour le passé va se réaliser.

Extrait du Corpus F5 *PIW: j'ai pas encore prévenu ma chum.... mais je crois bien que je vais y aller..... j' ai trop de fourmis dans les jambes moi :) Extrait du Corpus F8 *PDT: PAS... j'ai pas encore prix l'ombre d'un.... ca chauffe ;(%add:

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

PAS

Cette valeur est également possible avec le présent, ainsi que l'illustrent les extraits suivants :

Extrait du Corpus F4 *PAG: PAJ le jour ou je t'appellerai XXX est pas encore en prevision à la meteo %add: PAJ Extrait du Corpus F6 *PBT: PAW par exemple.... pour la nouvelle année si tu souhaite bonnanné a qqun d'ici en tenant compte de ton heure ca gache tout le punch de la nouvelle année vu que c pas la nouvelle année encore ici

Avec un passé composé seul, *encore,* indique que le procès dénoté s'est déjà produit dans le passé.

Extrait du Corpus F8 *PDV: PDT je sais pas pkoi ce matin fallait po me chercher, ma maman m'a encore reveillé à 2h du mat %add: PDT

Avec le futur, *encore* fait une présupposition sur la fréquence de l'événement dans le passé et présente donc une itération, une occurrence d'un événement qui s'est déjà produit dans le passé.

Extrait du Corpus F3 *PCO: ca vas encore gueulles les voisins

Le présent combiné à un statif et *encore* affirme la continuité d'un événement. La présupposition est que la proposition est vraie sur le passé et sur le présent.

Extrait du Corpus F3 *PAC: t'est encore la? %add: PDL Extrait du Corpus F7 *PJQ: et tu y crois encore toi??? LOLOLOL %add: PEI Extrait du Corpus F8 *PAS: PDT fais comme moi... 5 cafés et je dors encore %add: PDT

Avec un non statif, l'interprétation est itérative.

Extrait du Corpus F7 *PCJ: bon, brb encore Extrait du Corpus F7 *PAK: chanserv s'est encore fait la malle

Suivant la combinaison, *encore* apporte donc soit une valeur de continuité d'un procès, soit d'itérativité.

Les valeurs de *déjà* sont nombreuses et elles ne sont pas seulement temporelles. Dans nos corpus, les valeurs temporelles se présentent comme suit.

La contribution de *déjà* est particulière. Elle indique que l'état dénoté survient plus tôt que ce qui était attendu.

Extrait du Corpus F3 *PCO: deja en we? Extrait du Corpus F3 *PBK: et la il est deja 17h45

Si combiné avec un adverbial spécifiant une durée, *déjà*, indique que cette durée est plus importante qu'attendu.

Extrait du Corpus F6 *PBP: PBT je sais.... j'suis pas pressé :) je grave deja depuis 7h....

Avec un non statif, l'interprétation est l'antériorité.

Extrait du Corpus F4 *PDL: PCO saute de la tour Eifel.. ca a deja ete fait? %add: PCO *PDR: oui PDL... ca a déjà été fait..... %add: PDL Extrait du Corpus F5 *PEF: t déjà venu ici avant? %add: PBX Extrait du Corpus F5 *PBX: j'ai essayé déjà et mon code ne marche pas Extrait du Corpus P3 *PAS: ben si la personne etait encore connecte je suppose que oui %add: PBZ (1) *PAS: mais elle avait deja quitte %add: PBZ

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

tout de suite ne semble pas être influencé par, ni influencer le temps verbal, il prend appui sur le tiroir verbal. *Tout de suite* signifie que le procès se réalise sans intervalle temporel, mais il nécessite donc qu'ait été repéré un autre procès auquel succède le second.

Extrait du Corpus F4 *PCA: PAB tu vas sur URL et tu le trouves tout de suite %add: PAB Extrait du Corpus F8 *PBL: g tout de suite reconnu P99 et son sal caractere

Un autre type de repérage est effectué par les unités *tôt* et *tard.* Dans nos corpus, *tôt* et *tard* réfèrent à du temps cyclique : des moments identifiables dans le déroulement d'une journée ; ou à du temps culturel rythmique (habitudes). Ces items peuvent ou non se baser sur une référence à l'époque (passé, présent, futur) établie.

tôt, dans nos corpus, lorsqu'il est combiné à une référence au temps cyclique, signifie le début de cette portion de temps cyclique.

Extrait du Corpus F4 *PDW: aller dans les jardins tôt le matin et piquer les oeufs des enfants

Dans d'autres cas, *tôt* présuppose un point du temps habituel ou conventionnel pour le déroulement d'un procès, et spécifie une occurrence particulière du procès comme se déroulant avant le point habituel ou conventionnel. Cette antériorité peut être accentuée par un modifieur, ainsi que l'illustre l'exemple ci-dessous.

Extrait du Corpus F4 *PJQ: le vendredi il parte plus tôt :)

Dans l'extrait ci-dessus, le déterminant de *vendredi* permet de spécifier ce dernier comme un type, et donc de spécifier l'événement représenté par un verbe au présent, comme une habitude.

tard connaît le même type d'effets de sens. Dans nos corpus, lorsqu'il est combiné à une référence au temps cyclique, signifie la fin de cette portion de temps cyclique. Dans d'autres cas, *tard* présuppose un point du temps habituel ou conventionnel pour le déroulement d'un procès, et spécifie une occurrence particulière du procès comme se déroulant postérieurement au point habituel ou conventionnel. Cette postériorité peut être accentuée par un modifieur, ainsi que l'illustre l'exemple ci-dessous.

Extrait du Corpus F6 *PEP: PCD non il fait jour hyper tard %add: PCD

9.3.5. Adverbes temporels déictiques, tiroir et aspect lexical

Nous présenterons ici les cas de *maintenant, aujourd'hui, hier, demain* et *bientôt* tels qu'ils se présentent dans nos corpus en combinaison avec tiroir verbal et aspect lexical.

maintenant réfère à un moment ponctuel contemporain du moment de parole, ou pose un point contemporain du moment de parole, à partir duquel un procès se déroule.

Extrait du Corpus F7 *PCG: j étais en privé avec PKY et il est partie maintenant :(surement au salon d Opium Extrait du Corpus F4 *PAG: hey!!! fo payer maintenant %add: PCA Extrait du Corpus F5 *PAG: Bon, c fini ! maintenant tous sur PAM %add: ALL

Dans l'extrait ci-dessous (Extrait du Corpus F3), *maintenant* réfère à un moment ponctuel, mais il présuppose également que le locuteur avait préalablement annoncé son départ.

Extrait du Corpus F3 *PDL: bon aller je me casse %add: ALL (1) *PDL: pour de

en vertu de la loi du droit d'auteur.

bon mnt %add: ALL

Toutefois, *maintenant* peut référer à un moment étendu, ainsi que l'illustrent les extraits suivants.

Extrait du Corpus F5 *PEP: le mercredi matin j'avais UNSS %add: PAW *PEP: mais maintenant ils bossent le mercredi et ils ont le samedi de libre %add: PAW Extrait du Corpus F7 *PCJ: lol PDS, tu as presque mon niveau en anglais mtn... :p %add: PDS

On constate également que *maintenant* combiné avec statif au futur s'interprète comme un point de départ (Extrait du Corpus F4).

Extrait du Corpus P4 *P17: tiens t'auras l'air d'un vré mnt

aujourd'hui, dans nos corpus, permet de repérer des événements dans un intervalle circonscrit (la journée). Les procès peuvent appartenir au passé, au présent, au futur, respectivement avec les tiroirs verbaux de passé composé, présent et futur et peuvent être mis en discours comme occupant la totalité ou une partie de l'intervalle. Cet intervalle peut être étendu comme on en trouve un seul exemple dans nos corpus.

Extrait du Corpus F6 *PBP: c'est lourd.... mais ya 2 ans c'etait le top :) %add: PBT (1) *PDH: ouais c clair %add: PBP (1) *PDH: maisaujourd'hui c 20 X %add: PBP

L'interprétation de *demain* combiné avec un tiroir verbal présent dépend de l'aspect lexical du verbe, éventuellement modifié par la structure participative, pour obtenir le mode de repérage temporel comme consumant une portion plus ou moins grande de temps.

Extrait du Corpus F3 *PAE: demain on se mange de l' auxerrois %add: PBU Extrait du Corpus F5 *PEB: mais demain j'ai bac blanc de francais!!! Extrait du Corpus F6 *PCD: et demain tu dis qu'il a pas survécu au contrôle doinier ;) %add: PEP

Un statif au présent, combiné à *demain* place le procès dans le futur, *demain* s'interprétant comme un point de départ.

Extrait du Corpus F7 *PHQ: Lol PBG moi ca va bien je suis en vacances demain %add: PBG

Une activité ou un accomplissement ne peuvent recevoir la même interprétation, sauf dans le cas où *demain* est introduit par une locution prépositionnelle (*à partir de*).

Dans nos corpus, le seul autre tiroir verbal employé avec *demain* est le futur périphrastique (une seule occurrence).

Éventuellement, un adverbial peut contribuer à la spécification temporelle en sus de *demain*.

Extrait du Corpus F6 *PDA: PBF, c demain a 14h %add: PBF

Dans nos corpus, *hier* est seulement attaché à des verbes présentant le tiroir passé composé.

Extrait du Corpus P3 *PAS: j'ai vu stalingrad hier %add: PAk

Il survient que des locuteurs jouent de l'imprécision de la portion temporelle à laquelle réfère *hier.*

Extrait du Corpus P4 *P19: P33 chaste? depuis hier? :) %add: P13 (4) *P13: nan chaste hier P19 %add: P19 (1) *P13: c ca qu'elle voulait dire %add: P19 *P19: ah ok %add: P13 *P13: elle était chaste %add: P19 *P19: mais la chastete c pas

en vertu de la loi du droit d'auteur.

durant toute la vie plutot qu'un seul jour %add: P13 *P19: ? %add: P13 *P19: :) %add: P13 *P13: oui tout sa vie jusqu'à hier %add: P19

Enfin, *hier* peut spécifier un nom.

Extrait du Corpus P4 *P33: g mal partt en + avec le sport d'hier %add: P13

Notons que pour *hier* et *demain*, des dérivés en *avant* et *après* se trouvent mais contrairement à ce qu'une simple composition de sèmes ferait naître, les composés conservent la même unité : "le jour qui précède strictement hier", "le jour qui succède strictement demain" et non pas "la portion de temps qui précède hier", "la portion de temps qui succède demain".

maintenant, aujourd'hui, hier et demain sont déictiques, c'est-à-dire que leur interprétation nécessite, dans les cas prototypiques et de façon systématique dans nos corpus, la prise en compte, la connaissance de la situation de communication. C'est également le cas de *bientôt* dans nos corpus. Le repérage temporel spécifié par *bientôt* est un point proche du moment de parole et lui succédant. Des conflits de valeurs apparaissent donc en combinaison avec certains tiroirs.

Employé avec le futur, il n'y a pas de conflit de valeurs si on considère que la valeur de base de *bientôt* est de spécifier un moment ultérieur prochain du moment de parole.

Extrait du Corpus F8 [423] *PDJ: PAH et P99, bientot on aura mazinger Z

On note que *bientôt* employé avec le passé composé, peut modifier le caractère d'accompli apporté par ce tiroir.

Extrait du Corpus F4 *PCC: elle a bientot fini je ta laisse apres mdr %add: PAG

bientôt peut également interdire le caractère d'actualité possible pour le tiroir présent, et translater le procès dans le futur.

Extrait du Corpus F3 *PBK: c est bientot le week end :D %add: PCO

9.3.6. Noms temporels, prépositions temporelles

Les noms dont le sémantisme contient une spécification primairement temporelle modifient la référence temporelle de l'énoncé dans lequel ils apparaissent en fonction de la structure à laquelle ils sont intégrés.

Ainsi *fois* est un nom qui devrait permettre de référer à une (ou des) occurrence(s) particulière(s) d'un événement. Mais selon la structure à laquelle il est intégré, ses valeurs peuvent différer et modifier la référence temporelle d'un énoncé.

Extrait du Corpus F1 *PBW: allo PAB..... tu t'y prends à 2 fois pour venir? %add: PAB *PAB: PBW.. ben la oui %add: PBW

Dans cet extrait, l'énoncé de PBW dénote une manifestation itérative pantemporelle pour l'événement. La réponse de PAB spécifie que la validation est soumise à la considération du seul événement qui vient de se produire.

fois lorsqu'il est combiné à *prochain* ou *dernier* permet de spécifier une occurrence d'événement dans le futur ou dans le passé par rapport au moment de parole, avec la

[423] ***PAH et P99 ont des surnoms de héros de dessins animés.***

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

présupposition que l'événement s'est déjà produit dans le passé. Ainsi, dans l'énoncé ci-dessous, *fois* repère une occurrence d'événement par rapport au moment de parole dans le futur, tandis que le deuxième extrait permet de repérer une occurrence d'événement dans le passé.

Extrait du Corpus F3 *PCC: je ferai mieux la prochaine fois %add: PDL Extrait du Corpus F7 *PJQ: enfin la dernière fois j'ai pris le TGV pour paris sur le quai D %add: PCJ

En revanche *l'autre fois*, *la dernière fois* ou *la première fois* sont neutres par rapport à la référence temporelle située [424] , ainsi que l'illustre l'extrait ci-dessous, dans lequel le repérage est contribué pour ce qui concerne l'époque par le tiroir verbal.

Extrait du Corpus F7 *PKE: le site que je t'ai donner laut fois %add: PBP Extrait du Corpus F7 *PKG: c la premiere fois que je me branche ici

Dans nos corpus, ce n'est que lorsque *fois* est déterminé par *un* qu'il n'y a pas de présupposition sur les occurrences précédentes de l'événement, contrairement à ce qu'il se passe pour *cette fois*, ainsi que l'illustrent les exemples ci-dessous.

Extrait du Corpus F8 *PDV: ben le fait que ce soit arrivé une fois ne justifie pas tes mots, tu aimes bien faire ça %add: PDY Extrait du Corpus P2 *PAP: PAG jvais en acheter a la framboise cette fois.. tu aime? :PP %add: PAG

des fois repère plusieurs occurrences d'un événement. On remarque ici que *fois* repère des points dans le temps chronologiques, en certains de ces points, la proposition se vérifie, tandis qu'elle ne le fait pas dans d'autres.

Extrait du Corpus F3 *PAJ: hehehe je m'épate des fois %add: PCC Extrait du Corpus P3 *PAB: PAJ faut se serrvire de la frande touche qui a sur ton clavier des foix !! %add: PAJ

En revanche, la structure n *fois* spécifie un nombre de points déterminés dans le temps chronologique pour lesquels la proposition se vérifie.

Extrait du Corpus F8 *PEN: l ai dis que 3 fois Extrait du Corpus P2 *PAP: tu la dit 25439583490 fois et personne ta repondu alors :) %add: PAK Extrait du Corpus P7 *PBB: il la juste fais 26 fois %add: PAL

C'est également le cas de *chaque fois* qui repère un nombre indéterminé de points dans le temps, pour lesquels la proposition est vraie.

Extrait du Corpus F6 *PAW: je sais, mais si je devais tenir compte du décalage horaire chq fois que je dis bonjour.... %add: PBT Extrait du Corpus F8 *PBL: Phoebe dans ce cas pkoi a chaque fois que tu viens tout le monde se met a s engueuler? %add: PDY

On remarque dans nos corpus que les points du temps dénotés par *chaque fois*, sont toujours les occurrences d'un événement.

Les noms qui sont des unités de mesure du temps ne sont, dans nos corpus, pas employés seuls. Ils dénotent des intervalles temporels de grandeur variable et devraient permettre de dénoter la durée, toutefois, selon la structure dans laquelle ils sont employés ces derniers peuvent référer du temps situé.

[424] Ces unités sont toutefois plus probablement aptes à spécifier une référence temporelle passée puisque le déterminant suppose que l'allocutaire a déjà construit une représentation de ce type d'événements comme pertinente à l'interprétation.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Dans nos corpus *en*+durée spécifie un intervalle temporel borné à droite au terme duquel un procès repéré chronologiquement par rapport au moment de parole par le tiroir verbal se réalise. Cet adverbial temporel a un sémantisme proche de *dans*+ durée, ainsi que le montre l'extrait ci-dessous.

Extrait du Corpus F7 *PCJ: PFN... tu crois que ds une heure, il sont regle leur pb? :) %add: PFN *PIB: non %add: PCJ *PIB: jamais en une heure %add: PCJ Extrait du Corpus F7 *PHH: vi, et j' arrive a Londres en moins d'une heure c vrai... mais c po là ke je vais %add: PJQ

Nous avons vu que l'insertion de *en*+durée est un test utilisé pour déterminer le type de procès. Toutefois, si les exemples donnés ci-dessus correspondent bien à la définition tentée pour *en*+ durée, on peut remarquer qu'un verbe non télique accepte l'adverbial, à condition qu'une caractérisation de télicité soit apportée par l'objet du verbe.

Extrait du Corpus F1 *PBW: Y aurait il quelqu'un intéressé à venir déblayer mon entrée d'auto et mon escalier?... en deux jours.... il est tombé assez de neige pour que j'en ai jusqu'à mi-cuisse !! %add: ALL

Enfin, on remarque que dans les énoncés suivants la définition donnée n'est pas valide, puisqu'ils dénotent un intervalle de temps pendant lequel plusieurs occurrences d'un même événement ont eu lieu.

Extrait du Corpus P3 *PAG: pis en 5 semaine j ai fais 15 h de cours %add: PBB Extrait du Corpus F7 *PBO: ça fait 3 interruptions de service en 3 jours.. ça comment à bien faire %add: PBD Extrait du Corpus F7 *PBO: j'ai jamais autant été spammé en un laps de temps si court %add: ALL

Lorsque la spécification temporelle introduite par *en* n'est pas intrinsèquement une durée, elle impose une lecture de repérage chronologique étendu.

Extrait du Corpus F1 *PBW: fiouuuuuuuuuu... ne parlez pas tous en même temps.... on ne s'entend plus icitte!!!!! Extrait du Corpus F1 *PAZ: PAY :) on y va peut etre aussi avec XXX en septembre %add: PAY Extrait du Corpus F1 *PAS: PAG Vu l'état en fin de nuit.. j' imagine pas ton état actuel.... mdrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr %add: PAG Extrait du Corpus F3 *PCC: je suis en ce moment a naton t a cote on dirai %add: PAJ

La simultanéité de deux événements peut être dénotée dans un énoncé dans lequel un gérondif apparaît.

Extrait du Corpus F3 *PAY: 'vais prendre une douche en écoutant du metal a fond, ca devrait me detendre... %add: ALL Extrait du Corpus F5 *PZZ: PAB donne une du sirop a PCZ en s'éloignant lentement a reculons..... "euh pas de rhume pour moi svp!!!" . j'ai suffisement donné ! %add: PCZ, ALL

Les exemples ci-dessus montrent qu'on ne peut prédire la façon dont se recouvrent les événements, le gérondif permet seulement de spécifier une partie de temps chronologique est occupée par les deux procès. Dans l'extrait du Corpus F5 ci-dessus, les événements ne peuvent se recouvrir totalement. Dans l'extrait du Corpus F3, un des événements commence avant l'autre logiquement, mais ce sont seulement les connaissances générales du monde qui permettent de le dire dans les deux cas.

La contrepartie du test permettant de déterminer un type de procès est le test d'insertion de *pendant*+durée. Il est deux façons de spécifier une durée après *pendant* et

en vertu de la loi du droit d'auteur.

ceci permet de postuler que cette unité contient en son sémantisme une indication de durée. En effet, soit *pendant* est suivi d'un item lexical temporel de durée dont nous avons présenté l'inventaire pour nos corpus, soit il est suivi d'un nom d'événement auquel peut être associé une durée.

Extrait du Corpus F6 *PCD: je suis quand mêem arrivé à 9h pendans trois semaines... %add: PEP Extrait du Corpus F3 *PAT: là y a pause pendant le rush :) %add: ALL

pendant spécifie donc un intervalle temporel occupé par un procès. Toutefois, ainsi que nous l'avons déjà mentionné, cette interprétation dépend du type de procès dénoté. En effet, seuls les procès non téliques peuvent être interprétés comme occupant tout l'intervalle dénoté par le circonstanciel.

Extrait du Corpus F6 *PBP: PBT je peux faire a peu pres n'importe quoi pendant que je grave du moment que ca accede pas trop a mon dur si je grave a partir du dur %add: PBT dans+durée, ainsi que nous l'avions annoncé, dans nos corpus, spécifie un intervalle temporel borné à droite au terme duquel un procès repéré chronologiquement par rapport au moment de parole comme futur se réalise. Extrait du Corpus F1 *PAY: mais dans 15 jours on part au soleil :) %add: PBW Extrait du Corpus F2 *PAD: bah en fait on verra dans une semaine lol %add: PBN Extrait du Corpus F3 *PBK: donc dans 2 heures on est en week end :D %add: ALL dans+durée n'est pas forcément rattaché au prédicat, il peut compléter un nom. Extrait du Corpus F7 *PBS: trop tard, le big boss a tout loggé ce soir en première instance je perd mon aop et du coup mes chances de founderage de dans deux ans %add: PBP (1) *PHQ: qui auraient lieux ds un an %add: PBS

depuis dans nos corpus peut être combiné à une durée ou à un point de repère. Il est le plus souvent attaché à un statif. *Depuis*+durée permet de spécifier un point dans le temps passé de façon déictique. Il spécifie une borne gauche pour un statif à partir duquel la proposition est vraie.

Extrait du Corpus F3 *(action): PCK doit aller chez l'ophtalmo depuis.. 2 ans %add: PCO Extrait du Corpus F5 *PCZ: p juste comme enfin on avait un jour de grand beau comme o, n les reve depuis 6 mois %add: PAB***

depuis+repère spécifie également une borne gauche dans le passé pour un statif à partir duquel la proposition est vraie. Le point du temps spécifié pour le repérage peut être déictique (F7) ou non (F6).

Extrait du Corpus F7 *PBO: pkoi tu crois que mon serveur est HS depuis ce matin?..... %add: PBD Extrait du Corpus F6 *PBP: mon joli teac 4x SCSI.... que j'ai depuis 99 lol %add: PBT

depuis+repère et *depuis*+*durée* combinés à des procès téliques au présent spécifient également une borne gauche dans le passé à partir duquel le procès se réitère, tandis qu'ils spécifient la même valeur que pour des procès statifs lorsque le tiroir est le passé composé.

Extrait du Corpus P3 *PAS: j'a banni ce chan de mon cerveau depuis 3 ans %add: ALL

Il faut noter que *depuis*+repère, peut comporter un repère qui est un événement identifiable dans le temps pour l'allocutaire.

en vertu de la loi du droit d'auteur.

Extrait du Corpus F5 *PEN: paske j'ai arreté depuis que je viens plus boire l'apéro chez vous... Extrait du Corpus F7 *PIO: PCJ depuis le début de la grève, y'a plus de monde que sur les quais ;) %add: PCJ

On pourrait faire l'hypothèse *depuis* est au passé ce que *jusqu'à* est au futur.

Extrait du Corpus P7 *PBD: moi suis en vacances juska octobre et depuis fevrier :/ %add: PAL

Toutefois, si nous ne rencontrons dans nos corpus que des exemples de référence temporelle contribuée par *depuis* pour un passé par rapport au moment de parole, il est possible que *depuis* contribue à de la référence temporelle relative à un événement présenté dans le futur. *Depuis* introduit alors en effet une antériorité par rapport à ce point de référence.

Jusqu'à spécifie une borne droite et introduirait une postériorité par rapport à un point de référence.

Extrait du Corpus P4 *P19: mais la chastete c pas durant toute la vie plutot qu'un seul jour %add: P13 *P19: ? %add: P13 *P19: :) %add: P13 *P13: oui tout sa vie jusqu'à hier %add: P19

Dans cet extrait, on comprend effectivement que la portion de temps dénotée comporte une borne droite par rapport à un point de référence, cette borne droite pouvant donc se situer dans le passé par rapport au moment de parole. Les exemples concernant *jusqu'à* dans nos corpus sont en très petit nombre. On peut toutefois ici affirmer que cette unité n'a pas les mêmes propriétés combinatoires que *depuis*, puisqu'elle ne peut se combiner à une durée sans modifier l'interprétation de façon considérable.

Les unités *avant* et *après* permettent de spécifier une relation d'ordre temporel. Elles ne font pas naître de conflit particulier. Dans nos corpus, ils permettent le plus fréquemment d'établir une relation d'ordre entre deux événements.

Extrait du Corpus F3 *PCO: il parait que les joueurs ont eu moins d egueulle apres le match contre l'Espagne %add: PBU, PAE Extrait du Corpus F5 *PBY: mais je le fait apres l avoir annoce %add: PDL Extrait du Corpus F7 *PJM: ce weekend la non mais celui d'apres ca reste ouvert ;))) %add: PIW

Dans nos corpus, seul *avant,* connaît quelques emplois déictiques. Dans l'extrait (F5) ci-dessous, le moment chronologique dénoté pour l'événement précède le moment de parole.

Extrait du Corpus F5 *PEF: t déjà venu ici avant %add: PBX

Dans nos corpus, seul *après* est combiné à une durée. Il ne spécifie alors plus une relation d'ordre entre deux événements, mais un intervalle temporel précédent un événement.

Extrait du Corpus F7 *PKE: ben les vielles font rires apres longtemps %add: PBP (2) *PAT: PKE non le svielles bonne font rire apre slongtps %add: PKE Extrait du Corpus P7 *PBD: non j'ai arreter mon année de deug bio apres le premier semestre %add: PBB *PBB: bien joué %add: PBD

avant et *après* dans nos corpus sont aussi parfois combinés à un point dans le temps (repère). Ils ne spécifient alors pas une relation d'ordre entre deux événements, mais l'antériorité ou la postériorité d'un événement par rapport à un point de repère temporel chronologique.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Extrait du Corpus F6 *PCD: moi j'ai décidé d'être à l'heure au boulot... en gros arriver avant 10h30 %add: PEP Extrait du Corpus F6 *PEP: j'ai aussi décidé de ne plus être en retard à mon cours de muscu %add: PCD *PEP: donc de partir avant 16H45 %add: PCD

après peut spécifier la réitération d'un événement qui n'implique pas les mêmes participants. Dans ce cas, *après* introduit un syntagme nominal référant au participant qui, le premier, était engagé dans l'événement.

Extrait du Corpus P2 *PAG: on passe toujours apres un autre homme qui a dit les memes futilités a cette femme qu'on venere %add: PAY

Une autre préposition temporelle se trouve dans nos corpus. *il y a* introduit un repérage temporel toujours déictique. Il se combine à une durée.

Extrait du Corpus F4 *PDL: PCD sauf ke il a dit ca y a2 semaines lol %add: PCD Extrait du Corpus P9 *PBD: chuis allé en savoie y a 1 mois :/ %add: PAL

Certaines des prépositions que nous avons décrites pour nos corpus peuvent faire partie de locutions conjonctives. Nous n'en reprendrons pas ici l'analyse. Les faits de subordination ici présentés sont ceux qui n'ont pas encore été discutés à l'occasion d'une préposition, pour la raison qu'ils n'apparaissent qu'en subordination.

Dans nos corpus, l'emploi de loin le plus fréquent des subordonnées introduites par *quand* a les caractéristiques suivantes. Et la phrase matrice, et la phrase enchâssée, comportent un verbe au présent. L'effet qui s'ensuit est que les événements sont présentés comme co-occurrents de façon récurrente.

Extrait du Corpus P3 *PAK: ben si, quand les airbags s'ouvrent, il faut que les lunettes restent en place %add: PAS

Un certain nombre d'adverbiaux temporels sont indifférents à la spécification en époque impliquée par le tiroir verbal.

C'est par exemple le cas des éléments spécifiant le rythme d'un événement.

Extrait du Corpus P3 *PAS: PAG sur la couche c marque 6 mois 1 an ca veux pas dire que tu dois les changer tous les 6 mois 1 an Extrait du Corpus P3 *PAK: Moi J V parfois pour me marrer... Mais ça fait carrément longtemps que J suis pas allé Extrait du Corpus F7 *PFL: un whois PBP c'est util parfois =)

Les références au temps cyclique chronologique acceptent aussi bien les repérages dans le passé que dans le futur.

Extrait du Corpus P3 *PAK: PAY, 7 semaine G relu les 5 1ers métabarrons... je crois qu'il y en a 6, C ça ¿

La semaine peut être conçue comme un intervalle pendant lequel une activité s'est déroulée avec arrêts et reprises, mais dans l'extrait du Corpus P3, elle est aussi un repérage d'un intervalle passé par rapport au moment de l'énonciation. Dans cet énoncé, c'est le tiroir verbal qui détermine la référence temporelle chronologique.

Extrait du Corpus P3 *PAG: euh le we des vacances ca le fait po Extrait du Corpus P3 *PAM: PAB c gentil:) mais.. faudrait en faire un an aout un GT.... la jserais la!:) %add: PAB (6) *PAB: PAM y en aura surment un en aout! %add: PAM Extrait du Corpus P3 *PAK: rah là là, je viendrait jamais chez toi si tu tiens 1 boutique 1 jour, PAS Extrait du Corpus P3 *PAS: l'autre jour j'ai casse un vert

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Certains items ne peuvent s'employer que pour référer à un temps chronologique incluant le présent contemporain du moment de parole.

Extrait du Corpus P3 *PAS: c cher les lunettes de nos jours

La référence temporelle dans nos corpus présente un caractère indiscutablement déictique. Nous avons également noté qu'elle est le plus souvent imprécise. Les points du temps chronologiques dénotés ne sont que rarement précisément déterminés.

On aura noté, dans cet exposé, que si on écarte la morphologie verbale, parmi les moyens de l'expression de la temporalité, un grand nombre d'éléments permettant la référence temporelle peuvent également servir à la référence spatiale. On notera qu'en français, *précéder* et *succéder* initialement dédiés à l'expression de rapports spatiaux [425] conviennent à la fois pour l'expression de rapports temporels et de rapports spatiaux. Tandis que la simultanéité s'exprime avec des lexèmes dont l'origine semble nettement temporelle – le latin *simul* signifie "dans le même temps, ensemble". Si l'on examine les prépositions [426] du français qui portent des spécifications temporelles, on se trouve encore devant une évidente relation du temps et de l'espace. *en*, *dans* et *à* conviennent pour des spécifications aussi bien temporelles que spatiales. La spécification d'un intervalle temporel reçoit le même marquage que la spécification d'un intervalle spatial :

(99) entre 8 heures et midi (100) entre Paris et Lyon (101) de 8 heures à midi (102) de Lyon à Paris

Avant, après, depuis, jusque… permettent aussi indifféremment les spécifications temporelles ou spatiales.

Extrait du Corpus P3 *PAB: trop tard :P (1) *PAB: on peut pas revenir en arrieur

Les intervalles utilisent également la métaphore, ainsi que l'illustre l'extrait suivant.

Extrait du Corpus F7 *PBO: j'ai jamais autant été spammé en un laps de temps si court

Ces spécifications, pour le temps, comme pour l'espace sont relatives à un point de référence qui peut être un centre déictique [427] . Ce qui conduit nombre de linguistes à affirmer que le temps est métaphore de l'espace.

On reconnaît également que des termes auxquels une signification spatiale est attachée reçoivent une interprétation temporelle dans certains cas.

(103) D'ici là j'aurais fini mon devoir (104) D'ici demain je te confirme l'heure de la réunion

Pour ce qui concerne le type de repérage déictique de nos corpus, on note que par endroits, pour produire un message un locuteur adopte la deixis de l'allocutaire.

Extrait du Corpus P5 *(action): P98 is away : pouf pu la, fo ke j'm'aere

En effet, au moment où le locuteur P98 produit cet énoncé, il est encore sur le canal, mais son énoncé tient compte du fait que lorsqu'il sera interprété, lui-même ne sera plus

[425] Picoche & Firmin (1992)

[426] Cadiot (1997);Dessaux-Berthonneau, Cadiot et Anscombre (1993) ; Dessaux-Berthonneau (1993)

[427] Comrie (1976)

en vertu de la loi du droit d'auteur.

connecté au canal.

Nous nous attarderons maintenant sur des phénomènes de ce type, lié à la spécificité de la deixis induite par le dispositif IRC.

9.4. Phénomènes de construction narrative coopérative.

Les locuteurs engagés dans les conversationnelles de nos corpus sont dans des espaces distants. Ils construisent ensemble un espace virtuel commun dans lequel ils interagissent.

Cet espace virtuel est physiquement est un ensemble d'informations binaires dont certaines sont reproduites de façon quasi-simultanée sur des supports informatiques d'ordinateurs connectés par le biais d'Internet à un canal d'un réseau IRC puis traduites vers les écrans. La perception des utilisateurs de cet espace – hormis les connaissances implicites et explicites qu'ils ont du réseau – est matérialisée par une fenêtre sur leur écran. On remarque que les moyens de référer à l'espace virtuel se basent sur cette représentation et sont les moyens de la référence spatiale ordinaire.

Extrait du Corpus F4 *PCA: elle etait la tout a l'heure

Ainsi des indications de localisations spatiales se basent sur les manifestations de la présence des locuteurs dans la zone "liste des participants" [428] .

Extrait du Corpus F1 *PBW: PAG ;)))))))))... pourquoi t, es pas en haut? %add: PAG

L'extrait du Corpus F1 présenté, bien que manifestant une interrogation sur la position spatiale de l'allocutaire est une interrogation sur le statut de l'allocutaire. En effet, les surnoms des utilisateurs qui ont des statuts particuliers (opérateurs et *voice*) apparaissent en début de liste, c'est-à-dire, sur l'écran, dans la partie supérieure de la zone "liste des participants". L'extrait du Corpus P3 ci-dessous est du même type.

Extrait du Corpus P3 *PAS: bcp de tetes connues pour moi ici dans le haut peuple...

La localisation spatiale des participants est en fait une indication de statut des participants.

Des indications de localisation spatiale apparaissent pour d'autres éléments. L'extrait du Corpus F3 ci-après contient une intervention qui indique à l'allocutaire que l'information qu'il cherche est présente sur son écran, dans la partie supérieure. L'émetteur produit donc son message comme s'il partageait la deixis de l'allocutaire, ce qui n'est évidemment pas le cas. Pour ce faire, il postule ou sait que l'interface de l'allocutaire est telle que l'information *topic* est présentée dans la partie supérieure de l'écran.

Extrait du Corpus F3 *PIW: regarde en haut PJQ.... tu trouveras le site :) %add:

[428] Cf. Chapitre 2, section 2.2.4, Illustration (2-8)

en vertu de la loi du droit d'auteur.

PJQ

La perception/représentation d'un canal IRC est similaire à celle d'un espace délimité. Il existe un *dehors* et un *dedans,* ainsi que les extraits suivants permettent de le déduire.

Extrait du Corpus F2 *PAD: ah ok VVV qui se fait foutre dehors à chaque fois qu'elle vient c un ragot %add: PAG Extrait du Corpus F8 *PDY: ben si vous voulez me foutre dehors ok, je reviendrais pas, sauf que j'aurais pas compris en quoi je foutais la merde

On note également une dissociation de la deixis, entre un *ici*, référant à la localisation spatiale réelle des locuteurs et un *ici* référant à l'espace virtuel. Les deux exemples suivant présentent le contraste.

Extrait du Corpus F1 *PBW: veinarde!!!!!!! tu devais venir ici... tu aurais vu de la neige à volonté :) %add: PAY Extrait du Corpus P9 *PAF: icci c pas toi qui decide fdes aop %add: PAB

Dans nos corpus, on doit signaler qu'un certain nombre d'énoncés font référence à des éléments de situation non partagés. Il s'agit de description des actions, activités ou environnement du locuteur. Quelques-unes seulement de ces interventions se produisent parce que l'événement décrit a une incidence sur la conversation ainsi que l'illustre l'extrait suivant.

Extrait du Corpus P8 *PBB: brb lgros qui me sert de chum mapelle :)

Le locuteur, ici, explicite la raison de son absence délimitée prochaine. Les plus nombreuses sont toutefois purement "gratuites" et ne s'insèrent pas dans une unité conversationnelle en cours. Ces énoncés ne répondent pas à la loi de pertinence, à moins de considérer que la pauvreté de l'information annexe transmise par le dispositif nécessite des palliatifs.

Extrait du Corpus F6 *(action) PCN écoute -#- DJ Taylor & Flow Dieu Dansa.MP3 -#- 3:29 (128kbps 44.1kHz Stereo) -#- (1) *(action) PCD écoute ma main dans ta gueulle.mp3

L'extrait présenté ci-dessus montre l'agacement que peut causer ce type d'interventions. Cependant dans une grande partie des cas, il ne suscite aucun commentaire, ou des commentaires manifestant de l'intérêt.

Extrait du Corpus F8 *(action): PAH se reprend une meringue au chocolat, hummmm (3) *PDT: PAH... moi je préfère les meringues nature ;)... tu me les laisses ;) %add: PAH

Les conversations menées, outre leur grande liberté d'expression, semblent banales. Il est toutefois un phénomène que nous ne trouvons mentionné nulle part ailleurs et qui semble lié au dispositif de communication. Ce phénomène semble encore dédoubler la deixis de l'espace virtuel.

On remarque en effet ça et là au long du corpus des énoncés qui font référence à des situations qui n'existent que par des énoncés – ainsi que dans la fiction. Certaines toutefois sont tout à fait banales et peuvent se rencontrer dans d'autres types d'interactions verbales. Ainsi, au téléphone, lorsque le faire n'est pas possible, le dire s'y substitue.

Extrait du Corpus F8 *PDE: je t'embrasse PAH :)) %add: PAH

Ces énoncés semblent très circonscrits et apparaissent essentiellement dans les

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

séquences d'ouverture et de clôture. On remarquera que les locuteurs choisissent pour ce faire le plus souvent le mode de message actions de la façon dont on trouve un exemple ci-dessous.

Extrait du Corpus P9 *(action): PBG fait de gros bisous a PAH :))

C'est là un pas vers des énoncés qui ne font en rien référence à la réalité, ces derniers se situant essentiellement dans des messages de type action.

L'extrait ci-dessous présente un premier type.

Extrait du Corpus F8 *PDT: PAH... moi je préfère les meringues nature ;)... tu me les laisses ;) %add: PAH (6) *(action): PAH offre des meringues nature a PDT

L'extrait du Corpus P3 ci-dessous est du même type. Le locuteur PAS répond à une intervention descriptive problématique de PAJ en apportant une solution "virtuelle" : il dit envoyer un fichier de nourriture.

Extrait du Corpus P3 *(action): PAJ a FaIm *(action): PAS DCC send almorino Nourriture.ZIP %add: PAJ

Il survient également que ces interventions obtiennent leur pertinence d'une façon plus complexe. Dans l'extrait ci-dessous un locuteur produit un énoncé évaluatif, axiologique, au bénéfice de sa propre face. PAD manifeste son accord à l'égard du contenu propositionnel, mais le minimise immédiatement par un message de type action qui met en scène l'allocutaire dans une situation que les connaissances extralinguistiques permettent d'interpréter comme un rapport entre un humain et un chien. L'énoncé devient ainsi potentiellement menaçant pour la face de l'allocutaire.

Extrait du Corpus F1 *PAD: 'tain putain tu as fais quoi aux zautre aop? lol %add: PCJ *PAD: ils sont pas tous deop par ma faute qd meme mdr %add: PCJ *PCJ: ben kestu veux, je suis le seul gars irréprochable sur ce chan, c pas ma faute! LOL %add: PAD *PAD: c vrai %add: PCJ *(action): PAD donne un susucr a PCJ lol gentil garcon gentil ;) %add: PCJ, ALL *PCJ: j'av lu un suçon lolololololololol %add: PAD (1) *PCJ: POUAHAHAHAHA PTDR %add: P99 *PAD: pfuuu..... %add: PCJ

Toutefois, l'emploi de *garçon* dans la seconde partie de l'énoncé vient en contradiction avec cette hypothèse et adoucit la menace.

Le type d'interventions discuté jusqu'ici est ponctuel. Le second type s'inscrit dans de véritables unités conversationnelles développées. Il implique le plus souvent plus de deux locuteurs. Ces derniers construisent ensemble, à la volée, une narration. Là, le mode de message action n'est plus tant préférentiel.

Il existe plusieurs types de déclencheurs de ces unités. Les cas sont catégorisables. Le déclencheur peut être un FTA produit envers un allocutaire, qui réagit par un message action. Il s'agit du cas le plus fréquent. Le déclencheur peut également être un énoncé (ou une partie d'énoncé) métaphorique, feint(e) d'être interprété(e) dans son sens littéral. Il peut être encore un énoncé injonctif avec impossibilité pour l'allocutaire de répondre à l'ordre ainsi formulé, c'est-à-dire, assumer le rôle que le verbe assigne en principe au sujet. Enfin, le déclencheur peut être une erreur commise par un participant.

Les unités conversationnelles fictionnelles qui font suite à un FTA, mettent en scène des combats. Dans l'extrait du Corpus F3 ci-dessous, c'est sur le surnom de PBZ que se

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

produit le FTA.

Extrait du Corpus F3 *(action): PBZ se retient pour pas frapper PAE %add: ALL *PCM: non non retient toi pas vas y on na rien vu %add: PBZ *PCM: lol %add: PBZ *PBZ: lol %add: PCM *PDY: lolol %add: PCM *PAE: pkoi PDY? %add: PDY *PDY: pkoi koi PAE? %add: PAE *pAE: pkoi tu veux me frapper %add: PDY *(action): PCM donne un grand baton en mousse a PBZ tien vas y gene toi pas %add: ALL *PCM: lol %add: PDY, PAE, PBZ, *PDY: lol PAE, j'ai rien dit moi %add: PAE *PDY: c PBZ le méchant lol :p %add: PAE *PAE: oops j'ai mal lu %add: PDY *PBZ: une envie soudaine PAE, je voulais des fraises mais j'en ai pas sous la main alors je t'ai choisi %add: PAE (15) *(action): PBZ saute sur PAE et la baillonne au radiateur %add: ALL *(action): PAE sort son armure de gladiateur %add: ALL *PCM: jespere que le radi est ferme au moins arf %add: PBZ, PAE, *PCM: lol %add: PBZ, PAE, (1) *(action): PAE coupe ses attaches %add: ALL *PAE: VIENS TE BATTRE %add: PBZ

Les unités conversationnelles fictionnelles dont le déclencheur est un énoncé métaphorique sont bien moins fréquentes mais manifestent une fois de plus la dimension ludique des interactions dans les IRC.

Extrait du Corpus P3 *(action): PAG s'est autodétruite, présentement (4) *(action): PAS rammasse les miettes de PAG

Ce que nous avons appelé les défis improbables sont les cas ou un locuteur demande à un allocutaire de réaliser quelque chose qu'il n'est pas possible de réaliser immédiatement par le biais du dispositif. L'extrait ci-dessous en est un exemple.

Extrait du Corpus F3 *PCK: et la toupie chipriote, vs connaissez? %add: ALL (3) *PCA: la pasteque indonesiene %add: ALL *PAE: PCK montre moi lol %add: PCK (1) *PBK: tu me montres PCK? %add: PCK (2) *PCA: la choucroute marsienne %add: ALL *PCK: je vs préviens faut s' accrocher sinon gare au mal de coeur %add: PAE, PBK *(action): PAE s'accroche fort a PCK %add: ALL (1) *PCK: mmmmmhh la choucroute :)))) %add: ALL *(action): PBK s accroche tres fort a PAE %add: ALL *PAE: LOLOLOLOL %add: ALL *(action): PCK attention les p'tits loups accrochez vous, le manège va partir %add: ALL (2) *(action): PAE lache %add: ALL (7) *PAE: A mon signal dechainez les enfers

PAE et PBK demandent à PCK de lui montrer quelque chose, ce qui n'est pas possible réellement via le dispositif. PAE entame donc une mise en scène, à laquelle adhèrent les requérants et les autres participants de l'unité conversationnelle.

Le dernier cas est lorsque le déclencheur est une erreur commise par un participant. Celle-ci peut être une erreur de manipulation (commandes des opérateurs), une erreur dans un énoncé (syntaxique, lexicale, morphologique…), ou une erreur de décodage (lecture ou interprétation) d'un énoncé.

Extrait du Corpus P3 *PAB: j'ai mal viser! %add: ALL (4) *(action): PAS offre une paire de lunette a PAB %add: PAB (3) *(action): PAB remerci PAS ca fait longtemp que j'en voulait!!! %add: PAS (1) *(action): PAB tombe en senaglo dans les bras de PAS %add: PAS *(action): PAS de rien ca me fait plaisir j'insiste ca fait 200 euros %add: PAS (1) *PAS: lol PAB %add: PAB *PAB: heh %add: PAS *(action): PAS courage PAB tu vas tenir le coup %add: PAB *PAG: il se passe quoi tit PAB? %add: PAB *(action): PAS tape dans le dos de PAB %add: ALL (3) *(action): PAB

en vertu de la loi du droit d'auteur.

tiendra le cup coute que coute! que PAK d'ailleur! ;) %add: ALL

Si on parvient à peu près à déterminer les déclencheurs de ces unités, il n'en va pas de même de leur arrêt. L'arrêt est rarement lié à l'expression d'agacement d'un locuteur. Deux autres cas se présentent plus couramment : soit un dénouement est atteint, soit une autre unité conversationnelle attire progressivement les locuteurs.

Le tiroir verbal utilisé dans ces unités est essentiellement le présent, qui a l'avantage de simplifier la construction de la trame et pose les événements comme indiscutables. Il n'y a pas de négociation explicite sur la construction de la trame narrative : on ne remet pas en cause l'élément de trame narrative contribué par autrui explicitement. Il survient toutefois qu'un locuteur utilise des structures permettant de modifier ce qui a été construit auparavant.

9.5. Conclusion

Ce chapitre a montré, après quelques réflexions sur le concept de temps et son traitement linguistique, que nos corpus présentent une utilisation de la morphologie verbale pour la référence temporelle essentiellement déictique. L'inventaire des tiroirs utilisés le laissait déjà supposer, toutefois cet inventaire présente quelques spécificités par rapport à ce qu'on attendrait d'un oral prototypique (par ex. présence de formes de passé simple). Nous avons observé que les combinaisons d'éléments de l'énoncé portant des spécifications temporelles sont nombreuses et que pour nos corpus, elles aboutissent le plus souvent à un calcul déictique.

Nous avons également, à l'occasion de l'observation des formes de présent, mis en évidence un mode d'utilisation particulier de ce dernier. Nous l'avons qualifié de présent fictionnel, dans la mesure où les événements dénotés le sont comme le seraient des événements contemporains du moment de parole, mais qui toutefois ne sont d'aucune réalité effective. Les énoncés affectés sont comparables à des énoncés produits pour remplacer le *faire* lorsque le *faire* n'est pas envisageable (communications téléphoniques ou épistolaires). Une sous-catégorie se distingue cependant. Elle se dégage à partir d'énoncés utilisant le présent qui, d'une part, impliquent le plus souvent des situations improbables et relevant d'univers de référence non actuels, et d'autre part génèrent la production coopérative d'une narration. De telles séquences ne sont véritablement pas surprenantes dans nos corpus. Le dispositif ne permet d'échanger de façon synchrone que des énoncés dactylographiés. Ce fait engage à une exploitation de possibilités nouvelles. Le dispositif n'implique pas non plus une situation de communication pleinement partagée (les locuteurs sont engagés dans des espaces-temps différents, et dans des canaux, des unités conversationnelles différents), on peut alors faire l'hypothèse que ces séquences sont le résultat d'un besoin de rétablir le déséquilibre, en créant un univers fictionnel commun. Cet état de faits est de plus favorisé par le fait qu'une partie de la population internaute est reconnue pour avoir une culture du jeu en réseau et particulièrement des MUD (Multi-User Dungeons, *i. e.* jeux de rôles en réseau).

en vertu de la loi du droit d'auteur.

Conclusion

Cette étude s'était donné pour but de caractériser l'usage linguistique dans les IRC, à partir d'un corpus naturel en fonction des propriétés du dispositif susceptible d'imposer une contrainte. Il ne faut cependant pas écarter le fait qu'il s'agit d'une étude réalisé sur un corpus recueilli sur quelques mois, et dans des canaux particuliers. Ainsi, on peut envisager qu'une étude réalisée sur un plus long terme ou sur d'autres types de canaux apporterait un éclairage plus précis.

Nous avons essayé de mettre en place une méthodologie de recueil et d'analyse des données qui puisse servir à d'autres études de ce type de corpus. Ainsi, le chapitre 4, bénéficiant de la description du corpus brut, a permis de mettre en perspective les différents traitements des corpus qui ont été nécessaires d'un point de vue déontologique et pour les analyses linguistiques. Ces traitements ont été effectués, il ne faut pas l'oublier, dans le souci de correspondre à un format compatible avec un logiciel de traitement des corpus. Nous avons évidemment pesé le choix du logiciel, en fonction du corpus lui-même, et également en fonction des traitements permis par le logiciel. Ce choix a donc été un compromis à plusieurs points de vue. Certes il aurait été préférable de concevoir un système complètement adapté à la spécificité de nos corpus, nous avons toutefois choisi d'utiliser un standard, ce qui présente l'avantage non seulement de bénéficier de l'expérience acquise d'autres chercheurs sur d'autres types de corpus, mais également de ne pas multiplier les formats de données utilisés dans la communauté scientifique et autorise donc les comparaisons entre des corpus de types différents. De plus, l'utilisation par des chercheurs d'horizons différents d'un même outil, en permet le perfectionnement.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Les procédures de standardisation et de lemmatisation appliquées aux corpus sont également contestables. Cependant, nous ne pouvions étudier le code écrit des IRC sans avoir recours à la norme, et nous ne pouvions quantifier l'écart à la norme sans adopter une procédure de normalisation systématique sur tout le corpus. Malgré tout, nos corpus restent adaptables à d'autres types de logiciels.

La méthodologie nécessite également de présenter de la façon la plus explicite possible le dispositif ayant généré le corpus. Nous avons choisi pour ce faire de replacer les IRC dans le contexte du réseau Internet et des différents types de communication humaine qu'il permet. Cette première étape est importante dans la mesure où ce milieu est en constante évolution à tous les niveaux. Nous avons, à cette occasion, en effet, entrevu que, du point de vue linguistique, si l'anglais en était le maître, des signes montrent qu'il se peut qu'il n'en soit pas toujours le cas dans le futur, et donc que les langues les plus parlées dans le monde en tout cas, puissent connaître des phénomènes liés à l'utilisation du réseau pour la communication humaine, qui ne sauraient être écartés de l'analyse. Des études de l'influence des contraintes d'un dispositif de communication sur différentes langues devraient permettre d'évaluer plus objectivement la spécificité d'un dispositif.

Ainsi, avons-nous présenté dans le détail, en évitant les descriptions techniques et en ne retenant que les dimensions nécessaires à la compréhension du déroulement des comportements communicatifs, les caractéristiques du dispositif IRC. Nous avons relevé celles qui étaient susceptibles d'avoir des effets particuliers sur la production de messages linguistiques. Les paramètres majeurs à prendre en considération sont que le dispositif est basé sur le texte dactylographié, et qu'il impose une contrainte temporelle. Mais le dispositif se définit par un ensemble de propriétés. Ainsi, un canal se matérialise par une fenêtre sur l'écran, dans laquelle les messages des participants s'inscrivent selon leur ordre d'arrivée dans le canal. Un participant a accès à l'historique de la communication sur le canal depuis son arrivée dans le canal. Nous avons vu également que la fenêtre d'activité d'un canal présente des messages de types différents. Les participants ont en effet nécessairement un statut dans le canal. Nous avons à ce sujet observé le fait que suivant son statut, un participant peut avoir certains privilèges (suspendre la communication, déconnecter un utilisateur…). Au niveau de la charge cognitive engendrée par le dispositif, nous avons mentionné qu'elle peut être augmentée par le fait d'une part que les participants peuvent être engagés simultanément dans plusieurs canaux publics et/ou privés, et d'autre part, par le nombre de participants actifs, puisque des unités conversationnelles impliquant des ensemble de participants différents et/ou des thématiques différentes peuvent avoir lieu en parallèle dans un même canal.

Nous supposons que ces contraintes se présentent également pour toutes les langues (sauf systèmes d'écriture différents) et il serait intéressant de comparer l'usage linguistique dans les IRC que font les locuteurs de langues différentes de celles les plus couramment étudiées.

Ainsi, au niveau de l'écart à la norme de l'écrit, nous avons supposé que la contrainte temporelle imposée par le dispositif aurait pour conséquence un emploi fréquent de procédés d'abréviation. Il apparaît que le non respect de la norme n'entrave pas la compréhension, nous n'avons relevé aucune marque d'incompréhension dans les corpus

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

suite à l'emploi d'abréviations non conventionnelles. La contrainte temporelle a donc une influence sur le code écrit, ainsi qu'on s'y attendait. Toutefois, nous avons montré que le raccourcissement de la forme n'est pas le seul écart à la norme écrite. Nous avons relevé des phénomènes qui ne sont pas conformes à la norme sans pour autant produire une réduction formelle et des phénomènes qui au contraire présentent plus de matériel que la forme standard. Nous avons ainsi pu établir que le code écrit dans les IRC est modifié (par rapport à la norme) pour représenter des caractéristiques de l'oral. La chaîne orale est ainsi transposée par projection des différentes dimensions acoustiques sur les dimensions signifiantes de l'écrit. Nous avons également mis en évidence des procédés présentant un degré, certes faible, de cryptonymie, mais suffisant pour laisser supposer l'intention des internautes d'établir des signes d'appartenance à une communauté. Cette caractéristique est également abordée au chapitre concernant le lexique, mais, si des phénomènes similaires ont été relevés pour l'anglais, pour d'autres types de canaux, d'autres types de dispositifs de communication en réseau basés sur l'écrit, on peut se demander si cette caractéristique est valable pour tous les canaux, pour toutes les langues et pour tous les dispositifs.

Concernant le lexique, nous avons pu mettre en évidence une diversité lexicale importante. On trouve un certain nombre de termes dont la fréquence est peu élevée ou est absente en français standard. Les termes de faible fréquence standard dans le corpus sont essentiellement des noms d'armes ou d'êtres fictionnels. Ceux-ci apparaissent dans les unités conversationnelles de nos corpus discutées au dernier chapitre, et font penser au lexique des jeux de rôles. Les formes absentes du français standard sont essentiellement des formes utilisées pour référer à l'environnement IRC et au champ de l'informatique et des réseaux. Ces unités sont essentiellement des emprunts à l'anglais. Il apparaît également des jurons, des termes référant à des tabous sexuels, des doublons, des éléments de verlan, et des termes de variétés régionales du français. On ne peut non plus omettre le fait qu'un certain nombre d'unités appartiennent à des sociolectes de jeunes (15-25 ans).

Au point de vue de la syntaxe, nous avons pu constater qu'elle reflète la syntaxe de l'oral, hormis les phénomènes de construction en grille relevés par Blanche-Benveniste (1997). Les structures disloquées, par exemple, sont en nombre important. Les phénomènes d'intégration syntaxique se présentent globalement de façon similaire à ce qui est relevé pour l'oral. Toutefois, le dispositif n'interdit pas de recourir sporadiquement à des constructions habituellement qualifiées de plus littéraires. Il est là aussi problable que des canaux sur lesquels des thématiques de discussion précis sont établis manifestent une autre utilisation de la syntaxe.

Au plan de la temporalité, nous avons mis en évidence son mode de référence est essentiellement déictique. Une fois de plus, nous ne disons pas que les références non déictiques ne sont pas possibles, mais qu'elles sont peu fréquentes. Nous avons constaté que toutes les formes verbales ne sont pas utilisées dans nos corpus, et que, celles qui le sont, ne le sont pas également. Nous avons ainsi montré l'importance en nombre du tiroir présent. D'autre part, les références temporelles ne sont pas précises, elles sont même fréquemment vagues, et on note également des mentions fréquentes de la durée et de la vitesse. Ceci porterait à penser que les internautes sont moins préoccupés par le temps

en vertu de la loi du droit d'auteur.

chronologique (calendrier) que par l'occurrence des événements les uns par rapport aux autres et à la quantité de temps que consomment les événements. Mais cette particularité n'est sans doute pas entièrement contribuée par le dispositif. Certes, le temps partagé des locuteurs ne peut être directement qualifié qu'en référence au temps GMT (les locuteurs sont engagés dans des temporalités cycliques décalées ou culturellement différentes), mais on peut supposer que les buts communicatifs et le type de locuteur joue sur cette facette. En effet, nous avons ensuite observé des phénomènes de construction narrative fictionnelle coopérative qui ne peuvent pas non plus être le seul fait du dispositif (culture des jeux de rôles, absence d'histoire conversationnelle commune…).

Tout au long de ce travail, nous avons sporadiquement relevé la ludicité en œuvre dans l'échange de message linguistique. Celle-ci se manifeste à divers niveaux. Ainsi, les internautes jouent-ils du linguistique (polysémie et ambiguïtés), du métalinguistique (corrections orthographiques, négociations sur le code) autant que de l'extralinguistique (événements survenant sur le canal, situations des participants, etc.) et du dispositif (possibilités de changer de surnom, d'exclure des participants, etc.). Cette dimension semble une caractéristique importante des messages linguistiques sur les IRC.

L'usage linguistique ici examiné présente sans doute un certain degré d'instabilité. Ayant considéré des corpus antérieurs (fin des années 90), il nous semble en effet que des différences s'observent (au niveau du lexique et des sigles IRC en particulier). Ce constat demanderait bien sûr une analyse complète et la possibilité de mettre en parallèle ces évolutions/différences à celles observées pour d'autres dispositifs et dans les années à venir.

en vertu de la loi du droit d'auteur.

Bibliographie

Abeillé, A. et Godard, D., à paraître, "Deux types de prédicats complexes dans les langues romanes", in *LINX.*

Abeillé, A., Godard, D. et Sag, I. A., 1998, "Two kinds of composition in French complex predicates", in *Complex predicates in nonderivational syntax*, A. Kathol, T. Nakazawa and E. W. Hinrichs (eds), Academic Press, San Diego, pp. xviii, 451.

Aissen, J. et Perlmutter, D. M., 1983, "Clause reduction in Spanish", in *Studies in relational grammar*, D. M. Perlmutter (ed), University of Chicago Press, Chicago, pp. 360-403.

Amadasi Guzzo, M. G., 1995, "L'écriture dans l'histoire", in L'archéologie nouvelle, Vol. 4, pp. 15-61.

Andersen, H. L., 1995, "Dépendance entre propositions", in *Etudes Romanes,* Vol. 34, pp. 43-56.

Andersson, S.-G., 1972, *Aktionalität im Deutschen. Eine Untersuchung unter Vergleich mit dem russischen Aspektsystem*, univ. Almqvist & Wiksell (distr.), Uppsala, Stockholm.

André-Larochebouvy, D., 1984, *Introduction à l'analyse sémio-linguistique de la conversation*, Didier Crédif, Paris Saint-Cloud.

Anis, J., 1998, *Texte et ordinateur : l'écriture réinventée ?*, De Boeck Université, Paris, Bruxelles.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Anis, J., Chiss, J.-L. et Puech, C., 1988, *L'Écriture : théories et descriptions*, Paris, Bruxelles.

Aristote, 1983, *Premiers Analytiques (Organon III)*, Vrin, Paris.

Aristote, 1984, *De l'interprétation (Organon II)*, Vrin, Paris.

Arrivé, M., Gadet, F. et Galmiche, M., 1986, *La grammaire d'aujourd'hui : guide alphabétique de linguistique française*, Flammarion, Paris.

Arsac, J., Fluhr, C. et Bériot, A., "Informatique documentaire", in *Techniques de l'ingénieur.*

Attal, P., 1999, *Questions de grammaire*, Presses universitaires du Septentrion, Villeneuve d'Ascq.

Austin, J. L. et Lane, G., 1970, *Quand dire, c'est faire*, Éditions du Seuil, Paris.

Bailly, A., 1935, *Dictionnaire grec-français : rédigé avec le concours de E. Egger à l'usage des élèves des lycées et des collèges*, Hachette, Paris.

Bally, C., 1965, *Linguistique générale et linguistique française*, Éditions Francke, Berne.

Barberis, J. M. (ed) 1999, "*Le français parlé : variétés et discours*", Praxiling Université de Montpellier III, Montpellier.

Barberis, J. M., 1992, "Onomatopées, interjection, un défi pour la grammaire", in *L'information Grammaticale,* Vol. 53, pp. 52-57.

Bar-Hillel, Y., 1953, "Some linguistic problems connected with machine translation", in *Philosophy of Science,* Vol. 20, pp. 217-223.

Bar-Hillel, Y., 1960, "The present status of automatic translation of languages", in *Advances in Computers*, F. L. Alt (ed), Academic Press, New York, pp. 91-163.

Basset, L., 1989, *La syntaxe de l'imaginaire. Etude des modes et des négations dans l'Illiade et L'Odyssée*, Lyon.

Beaman, K., 1984, "Coordination and subordination revisited: Syntactic complexity in spoken and written narrative discourse," in *Coherence in spoken and written discourse*, T. D. (ed), pp. 45-80.

Bennett, M. et Partee, B., 1978, *Toward the logic of tense and aspect in English*, Indiana University Linguistics Club.

Benveniste, E., 1974, *Problèmes de linguistique générale*, Gallimard, Paris, Vol. 2/2.

Béraud, A., Euzen-Dague, M.-G. et Rémi-Giraud, S., 1988, *Le Taste-mots dans les arbres : étude systématique du lexique français à l'usage des lycées, grandes écoles et universités*, Centre régional de documentation pédagogique, Lyon.

Berger, P., 1891, *Histoire de l'écriture dans l'antiquité*, Imprimerie nationale, Paris.

Berman, R. A. et Slobin, D. I., 1994, *Relating events in narrative : a crosslinguistic developmental study*, L. Erlbaum Associates, Hillsdale N.J.

Berthonneau, A.-M. et Kleiber, G., 1993, "Pour une nouvelle approche de l'imparfait : l'imparfait un temps anaphorique méronimique", in *Langages,* Vol. 112, pp. 55-73.

Biber, D., 1988, *Variation across speech and writing*, Cambridge University Press, Cambridge England ; New York.

Biber, D., Conrad, S. et Reppen, R., 1998, *Corpus linguistics : investigating language*

 Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

structure and use, Cambridge University Press, Cambridge ; New York.

Bilger, M., Eynde, K. v. d., Blanche-Benveniste, C. et Gadet, F., 1998, *Analyse linguistique et approches de l'oral : recueil d'études offert en hommage à Claire Blanche-Benveniste*, Peeters, Leuven.

Blake, B. J., 1990, *Relational grammar*, Routledge, London ; New York.

Blanche-Benveniste, C. et Chervel, A., 1978, *L'orthographe*, F. Maspero, Paris.

Blanche-Benveniste, C., 1977, "L'un chasse l'autre, le domaine des auxiliaires", in *Recherches sur le français parlé,* Vol. 1, pp. 100-148.

Blanche-Benveniste, C., 1984, "Commentaires sur le passif en français", in *Travaux du Cercle linguistique d'Aix-en-Provence,* Vol. 2, pp. 1-23.

Blanche-Benveniste, C., 1985, "La langue du dimanche", in *Reflets,* Vol. 14, pp. 42-43.

Blanche-Benveniste, C., 1991, "Les études sur l'oral et le travail d'écriture de certains poètes contemporains", in *Langue Française*:89, pp. 52-71.

Blanche-Benveniste, C., 1992, "A propos des énoncés sans verbe: Les Énoncés réponses." in *Recherches sur le Français Parlé,* Vol. 11, pp. 57-85.

Blanche-Benveniste, C., 1997, *Approches du français parlé.*

Boas, F. et Benedict, R., 1965, *General anthropology*, Heath reprinted by Johnson Reprint Corp. New York, Boston.

Boissel, P., Darbord, B., Devarrieux, J., Fuchs, C. et Guimier, C., 1989, "paramètres énonciatifs et interprétations de *pouvoir*", in *Langue française,* Vol. 84, pp. 24-69.

Borillo, A., 1982, "deux aspects de la modalisation assertive : *croire* et *savoir*", in *Cahiers de grammaire,* Vol. 4, pp. 33-53.

Bottero, J., 1983, *Mésopotamie - l'écriture, la raison, les dieux*, Gallimard.

Bouchard, 2000, "M'enfin !!! des 'petits mots' pour les "petites" émotions ?" in *Les émotions dans les interactions*, C. Plantin, M. Doury and V. Traverso (eds), PUL/ARCI, Lyon, pp. 223-238.

Branca-Rosoff, 1998, "Abréviations et icônes dans les prises de notes", in *Analyse linguistique et approches de l'oral : recueil d'études offert en hommage à Claire Blanche-Benveniste*, M. Bilger, K. v. d. Eynde, C. Blanche-Benveniste and F. Gadet (eds), Peeters, Leuven.

Brown, E. K. et Miller, J., 1999, *Concise encyclopedia of grammatical categories*, Elsevier, Amsterdam ; New York.

Brown, P. et Levinson, S. C., 1987, *Politeness : some universals in language usage*, Cambridge University Press, Cambridge Cambridgeshire ; New York

Bruillard, É. 1997, *Les machines à enseigner*, Paris.

Brunet, E., 2000, "Qui lemmatise dilemme attise", in *lexicométrica,* Vol. 2, pp. 1-19.

Brunot, F. et Bruneau, C., 1905, *Histoire de la langue française : des origines à 1900*, A. Colin, Paris.

Brunot, F., 1922, *La pensée et la langue; méthode, principes et plan d'une théorie nouvelle du langage appliquée au français*, Masson et cie, Paris.

Bruxelles, S. et Traverso, V., 2001, "*Ben* : apport de la description d'un « petit mot » du

en vertu de la loi du droit d'auteur.

discours à l'étude des polylogues." in *Marges Linguistiques,* Vol. 2, pp. 1-18.

Bush, V., 1945, "As We May Think", in *The Atlantic Monthly,* Vol. 2001.

Cadiot, P. et Furukawa, N., 2000, "La prédication seconde, présentation", in *Langue Française*:127, pp. 3-5.

Cadiot, P., 1997, *Les prépositions abstraites en français*, Armand Colin, Paris.

Calvet, L. J., 1987, *La guerre des langues et les politiques linguistiques*, Payot, Paris.

Calvet, L. J., 1999, *Pour une écologie des langues du monde*, Plon, Paris.

Calvet, L. J., 2002, *Le marché aux langues : essai de politologie linguistique sur la mondialisation*, Plon.

Carré, R., Dégremont, J.-F., Gross, M., Pierrel, J.-M. et Sabah, G., 1991, *Langage humain et machine*, Presses du CNRS.

Carton, F., 1983, *Les Accents des francais*, Hachette, Paris.

Catach, N., 1980, "La ponctuation", in *Langue Française,* Vol. 45, pp. 16-27.

Catach, N., 1995, *L'orthographe française traité théorique et pratique*, Nathan, Paris.

Chafe, W., 1982, "Integration and involvment in speaking, writing and oral literature", in *Spoken and written language: Exploring orality and literacy*, T. D. (ed), pp. 35-53.

Chang, N., 1997, "A cognitive approach to Aspectual Composition".

Chaudenson, R., 2000, *Mondialisation : la langue française a-t-elle encore un avenir ?*, Didier Erudition.

Chenon, C., Chenu, F. et Gabal, A., 1998, *Correction de textes issus d'OCR*, rapport de stage CEA, Saclay.

Choi, E.-J., 2000, "*Contribution à une étude comparative des systèmes verbaux français et coréen : suggestion d'une structure logico-sémantique pour l'analyse du temps*", Nouveau Régime, Sciences du Langage, Université lumière Lyon2, Lyon.

Chomsky, N., 1968, *Syntactic structures*, Mouton, The Hague.

Chomsky, N., 1987, *La nouvelle syntaxe : concepts et conséquences de la théorie du gouvernement et du liage*, Éditions du Seuil, Paris.

Clément, D., 1991, "réflexions sur la notion d'intégration en syntaxe ou qu'entend-on par "l'expression X est plus ou moins 'intégrée' à la phrase" ?" in *Langage,* Vol. 104, pp. 7-21.

Coates, J., 1994, "No Gap, Lots of Overlap: Turn-taking Patterns in the Talk of Women Friends", in *Language and Literacy in Social Context*, D. Graddol, J. Maybin and B. Stierer (eds), Clevedon: Multilingual Matters and the Open University Press.

Cohen, D., 1989, *L'aspect verbal*, Presses universitaires de France, Paris.

Cohen, M. S. R., 1953, *L'écriture*, Éditions sociales, Paris.

Comrie, B., 1976, *Aspect : an introduction to the study of verbal aspect and related problems*, Cambridge University Press, Cambridge ; New York.

Comrie, B., 1985, *Tense*, Cambridge University Press, Cambridge Cambridgeshire ; New York.

Confais, J.-P., 1995, *Temps, mode, aspect : les approches des morphèmes verbaux et leurs problèmes a l'example du francais et de l'allemand*, Presses Universitaires du

en vertu de la loi du droit d'auteur.

Mirail, Toulouse.

Corblin, F., Godard, D. et Marandin, J.-M. (eds), 1999, "*Empirical issues in formal syntax and semantics 2*", P. Lang, Bern ; New York.

Craig, C. G. (ed) 1986, "*Noun classes and categorization : proceedings of a symposium on categorization and noun classification, Eugene, Oregon, October 1983*", Typological studies in language v. 7., J. Benjamins, Amsterdam ; Philadelphia.

Creissels, D., 1995, *Eléments de syntaxe générale*, Presses universitaires de France, Paris.

Creissels, D., 2000a, *Syntaxe - première partie*, Cours de Licence, Univeristé Lumière Lyon2, Lyon.

Creissels, D., 2000b, *Syntaxe - deuxième partie*, Cours de Licence, Université Lumière Lyon2, Lyon.

Creissels, D., 2000c, *Syntaxe*, Cours de maîtrise, Université lumière Lyon2, Lyon.

Creissels, D., 2000d, "L'emploi résultatif de 'être+participe passé' en français", in *Cahiers Chronos n°6 Passé et parfait*, pp. 133-142.

Crowley, T., 1997, *An introduction to historical linguistics*, Oxford University Press, Auckland N.Z. ; New York.

Crystal, D., 2001, *Language and the Internet*, Cambridge University Press, Cambridge UK ; New York.

Culicover, P. W. et McNally, L. (eds), 1997, "*The limits of syntax*", Syntax and semantics ; v. 29., Academic Press, San Diego.

Culioli, A., 1980, "Valeurs aspectuelles et opérations énonciatives : l'aoristique", in *La notion d'aspect*, J. David and R. Martin (eds), Klincksieck, Paris, pp. 181-194.

Culioli, A., 1990, *Pour une linguistique de l'énonciation : opérations et représentations*, Ophrys, Paris.

Damourette, J. et Pichon, E., 1911-1936, *Des mots à la pensée; essai de grammaire de la langue française*, Collection des linguistes contemporains, Paris.

David, J. et Kleiber, G., 1983, *La Notion sémantico-logique de modalité : colloque*, Klincksieck, Metz, Paris.

David, J., Martin, R. et Pottier, B. (eds), 1980, "*La notion d'aspect : Colloque organisé par le Centre d'Analyse Syntaxique de l'Univ. de Metz*", Klincksieck, Paris.

Davis, P. W. et Rice University. Dept. of Linguistics. Symposium, 1995, *Alternative linguistics : descriptive and theoretical modes*, J. Benjamins, Amsterdam ; Philadelphia.

De Swart, H., 2000, "Tense, aspect an coercion in a cross-linguistic perspective", M. Butt and T. Holloway King (eds), *Berkeley Formal Grammar conference*, University of California, Berkeley, CSLI Publications.

Delattre, P., 1966, *Studies in French and comparative phonetics; selected papers in French and English*, Mouton, The Hague.

Dessaux-Berthonneau, A. M., Cadiot, P. et Anscombre, J.-C., 1993, *Les Prépositions : méthodes d'analyse*, Presses universitaires de Lille, Villeneuve d'Ascq.

Dessaux-Berthonneau, A.-M., 1993, "Avant/après: De l'espace au temps", in *Lexique,*

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Vol. 11, pp. 41-109.

Di Cristo, A., 1998, "Intonation in French", in *Intonation systems : a survey of twenty languages*, Cambridge University Press, Cambridge U.K. ; New York NY, pp. 195-218.

Diop, C. A., 1988, *Nouvelles Recherches sur l'Egyptien Ancien et les langues Negro-Africaines modernes*, Présence Africaine.

distributed by Halsted Press, Hillsdale N.J., New York.

Dixon, R. M. W., 1977, *A grammar of Yidiny*, Cambridge University Press, Cambridge ; New York.

Douglas-Cowie, E., 1978, " Linguistic code-switching in a Northern Irish village: social interaction and social ambition." in *Sociolinguistic Patterns in British English*, P. Trudgill (ed), Edward Arnold, London, pp. 37-51

Dowty, D. R., 1977, "Toward a semantic analysis of verb aspect and the English "imperfective" progressive", in *Linguistics and Philosophy,* Vol. 1, pp. 45-77.

Dowty, D. R., 1979, *Word meaning and Montague grammar : the semantics of verbs and times in generative semantics and in Montague's PTQ*, D. Reidel Pub. Co., Dordrecht ; Boston.

Dowty, D. R., 1986, "The effects of aspectual class on the temporal structure of discourse : semantics or pragmatics", in *Linguistics and Philosophy,* Vol. 9:1, pp. 37-62.

Drescher, M., 2000, "Eh tabarnouche! c'etait bon. Pour une approche communicative des jurons en francais quebecois", in *Cahiers-de-praxematique*:34, pp. 133-160.

Dubois, D. (ed) 1991, *Sémantique et cognition : catégories, prototypes, typicalité*, Editions du Centre national de la recherche scientifique : Diffusion Presses du CNRS, Paris.

Dubois, D., 1991, "Catégorisation et cognition : "10 ans après", une évaluation des concepts de Rosch", in *Sémantique et cognition : catégories, prototypes, typicalité*, D. Dubois (ed), Editions du Centre national de la recherche scientifique : Diffusion Presses du CNRS, Paris, pp. 31-54.

Dubois, J., 1967, *Grammaire structurale du français : le verbe*, Librairie Larousse, Paris.

Dubois, J., 1971, *Dictionnaire du français contemporain*, Larousse, Paris.

Ducrot, O. et Wahl, F., 1968, *Qu'est-ce que le structuralisme?*, Éditions du Seuil, Paris.

Dufour, A., 1995, *Internet*, Presses universitaires de France, Paris.

Dupont, P., 1983, "*Modèles mathématiques pour l'analyse linguistique*", Thèse de Doctorat, Sciences du langage, Lumière, Lyon.

Edmonds, J., 1978, "The verbal complex V'-V in French", in *Linguistic Inquiry,* Vol. 9, pp. 151-175.

Encrevé, P., 1988, *La liaison avec et sans enchaînement : phonologie tridimensionnelle et usages du français*, Seuil, Paris.

Engelbart, D. C., 1963, "A Conceptual Framework for the Augmentation of Man's Intellect." in *Vistas in Information Handling.*

Engwall, G., 1984, *Vocabulaire du roman français (1962-1968) : dictionnaire des*

 Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

fréquences, Distribution Almqvist & Wiksell International, Stockholm Suède.

Ewing Duncan, D., 1999, *Le temps conté*, Paris.

Fayol, M., 1997, *Des idées au texte : psychologie cognitive de la production verbale, orale et écrite*, Presses Universitaires de France.

Feuillet, J., 1988, *Introduction à l'analyse morphosyntaxique*, Presses universitaires de France, Paris.

Fluhr, C., 1997, *Informatique documentaire*, Cours de DEA Information Scientifique et Technique, INSTN, CEA Saclay, Saclay.

Fluhr, C., Mordini, P., Moulin, A. et Stegentritt, E., 1994, *EMIR Final report, ESPRIT project 5312, DG III*, Commission of the European Union.

Fodor, J. A., 1983, *The modularity of mind : an essay on faculty psychology*, MIT Press, Cambridge Mass.

Foley, W. A. et Van Valin, R. D., 1984, *Functional syntax and universal grammar*, Cambridge University Press, Cambridge Cambridgeshire ; New York.

Foley, W. A. et Van Valin, R. D., 1985, "Information parckaging in the clause", in *Typology and syntactic description*, T. Shopen (ed), Cambridge University Press, Cambridge ; New York, pp. 282-363.

Fonagy, I., 1989, "Le français change de visage ?" in *revue romane,* Vol. 24:2, pp. 225-254.

Foulet, 1925, "Le développement des formes surcomposées", in *Romania*, pp. 203-252.

Fox, B. et Jasperson, R., 1995, "A syntactic exploration of repair in english conversation", in *Alternative linguistics : descriptive and theoretical modes*, P. W. Davis (ed), J. Benjamins, Amsterdam ; Philadelphia, pp. 77-134.

Fuchs, C. et Guimier, C., 1989, "La polysémie de *pouvoir*", in *Langue Française,* Vol. 84, pp..

Fuchs, C., Lacheret-Dujour, A. et Victorri, B., 1993, *Linguistique et traitements automatiques des langues*, Hachette supérieur, Paris.

Gaatone, D., 1998, *Le passif en français*, Duculot, Paris.

Gadet, F., 1989, *Le français ordinaire*, A. Colin, Paris.

Gadet, F., 1996, "Une distinction bien fragile: oral/ecrit", in *Travaux neuchatelois de linguistique-(TRANEL)*:25, pp. 13-27.

Gardies, J.-L., 1979, *Essai sur la logique des modalités*, Presses universitaires de France, Paris.

Gardiner, A. H., 1950, *Egyptian grammar; being an introduction to the study of hieroglyphs*, Published on behalf of the Griffith Institute Ashmolean Museum Oxford by Oxford University Press, London.

Givón, T., 1984, *Syntax : a functional-typological introduction*, J. Benjamins Pub. Co., Amsterdam ; Philadelphia.

Givón, T., 2001, *Syntax : an introduction*, J. Benjamins, Amsterdam ; Philadelphia, Vol. 1/2.

Givón, T., 2001, *Syntax : an introduction*, J. Benjamins, Amsterdam ; Philadelphia, Vol. 2/2.

en vertu de la loi du droit d'auteur.

Godard, D., 1988, *La syntaxe des relatives en français*, Editions du Centre national de la recherche scientifique, Paris.

Goffman, E., 1973, *La mise en scène de la vie quotidienne : la présentation de soi*, Minuit, Paris, Vol. 1/2.

Goffman, E., 1973, *La mise en scène de la vie quotidienne : Les relations en public*, Minuit, Paris, Vol. 2/2.

Goffman, E., 1974, *Les rites d'interaction*, Minuit, Paris.

Goldberg, A. E., 1995, *Constructions : a construction grammar approach to argument structure*, University of Chicago Press, Chicago.

Goldberg, A. E., 1996, *Conceptual structure, discourse, and language*, CSLI Publications, Stanford Calif.

Goldsmith, J. A. (ed) 1995, "*The handbook of phonological theory*", Blackwell handbooks in linguistics ; 1., Blackwell, Cambridge Mass. USA.

Gosselin, L., 1996, *Sémantique de la temporalité en français : un modèle calculatoire et cognitif du temps et de l'aspect*, Duculot, Louvain-la-Neuve.

Gosselin, L., 1998, "Modalité et temporalité : le rôle de l'aspect", B. Victorri and J. François (eds), *Sémantique du lexique verbal*, Caen, Presses Universitaires de Caen, pp. 87-97.

Gougenheim, G., 1964, *L'élaboration du français fondamental (1er degré) : étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, Chilton Books, Philadelphia.

Graddol, D., Maybin, J. et Stierer, B., 1994, *Researching language and literacy in social context : a reader*, Multilingual Matters in association with the Open University, Clevedon England ; Philadelphia.

Gregory, M. L. et Michaelis, L. A., 2001, "Topicalization and left-dislocation: a functional opposition revisited", in *Journal of Pragmatics,* Vol. 33:11, pp. 1665 -1706.

Grenoble, L. A. et Whaley, L. J., 1998, *Endangered languages : language loss and community response*, Cambridge University Press, Cambridge England ; New York.

Grevisse, M. et Goosse, A., 1993, *Le bon usage : grammaire française*, Duculot, Paris.

Guillaume, G., 1929, *Temps et verbe; théorie des aspects, des modes et des temps*, H. Champion, Paris.

Guillaume, G., 1938, "Théorie des auxiliaires et examen de faits connexes", in *Bulletin de la Société de Linguistique de Paris,* Vol. 39:I, pp. 5-23.

Guimier, C., 1989, "Constructions syntaxiques et interprétations de *pouvoir*", in *Langue Française,* Vol. 84, pp. 9-23.

Gülich, E., 1999, "Les activités de structuration dans l'interaction verbale", in *Le français parlé : variétés et discours*, J. M. Barberis (ed), Praxiling Université de Montpellier III, Montpellier.

Habert, B., Nazarenko, A. et Salem, A., 1997, *Les linguistiques de corpus*, A. Colin.

Hagège, C., 1985, *L'homme de paroles : contribution linguistique aux sciences humaines*, Fayard, Paris.

Hagège, C., 2000, *Halte à la mort des langues*, Editions Odile Jacob, Paris.

en vertu de la loi du droit d'auteur.

Halliday, M. A. K. et Hasan, R., 1985, *Language, context and text : aspects of language in a social-semiotic perspective*, Deakin University Press, Victoria Australia.

Halliday, M. A. K., 1985, *An introduction to functional grammar*, E. Arnold, London ; Baltimore Md. USA.

Halliday, M. A. K., 1985, *Intonation et rythme : suppléments à la proposition*, Groupe de recherches sémio-linguistiques (U.R.L. 7 de l'Institut national de la langue française) Ecole des hautes études en sciences sociales Centre nationale de la recherche scientifique, Paris.

Hansen, A. B., 1991, "The covariation of [e] with style in Parisian French : an empirical study of E caduc and pre-pausal [schwa]", *ERTW Phonetics and phonology of speaking Styles*, Barcelona, pp. 30-1-30-7.

Hansen, M.-B. M., 1995, "Marqueurs metadiscursifs en francais parlé : l'exemple de bon et de ben", in *Le français moderne,* Vol. 63:1, pp. 20-41.

Hansen, M.-B. M., 1997, "*alors* and *donc* in spoken french : a reanalysis", in *Journal of Pragmatics*:28, pp. 153-187.

Herdan, G., 1966, *The advanced theory of language as choice and chance*, Springer-Verlag, Berlin ; New York.

Hirst, D. et Di Cristo, A., 1998, "A survey of intonation systems", in *Intonation systems : a survey of twenty languages*, Cambridge University Press, Cambridge U.K. ; New York NY, pp. 1-44.

Hirst, D. et Di Cristo, A., 1998, *Intonation systems : a survey of twenty languages*, Cambridge University Press, Cambridge U.K. ; New York NY.

Hutchins, J., 1996, "ALPAC: The (In)famous Report", in *MT News International,* Vol. 14, pp. 9-12.

Imprimerie_nationale, 1990, *Lexique des règles typographiques en usage à l'Imprimerie nationale*, Imprimerie nationale, Paris.

INALF, 1971, *Dictionnaire des fréquences du Trésor de la langue française*, Didier, Paris/6.

Jacquard, A., 2001, *La science à l'usage des non-scientifiques*, Calmann-Lévy, Paris.

Jakobson, R., 1963, *Essais de linguistique générale*, Les Editions de Minuit, Paris.

Jeanneret, T., 1999, *La coénonciation en français : approches discursive, conversationnelle et syntaxique*, P. Lang, Bern.

Jefferson, G., 1973, "A Case of Precision Timing in Ordinary Conversation: Overlapped Tag-Positioned Address terms in Closing Sequences", in *Semiotica,* Vol. 9, pp. 47-96.

Juilland, A. G., Brodin, D. R. et Davidovitch, C., 1970, *Frequency dictionary of French words*, Mouton, The Hague.

Kahn, F., 1954, *Le système des temps de l'indicatif, chez un Parisien et chez une Bâloise*, Droz, Genève.

Kamio, A. et Dokky*o Daigaku, 1997, *Directions in functional linguistics*, J. Benjamins, Amsterdam ; Philadelphia.

Kamp, H. et Reyle, U., 1993, *From discourse to logic : introduction to model theoretic semantics of natural language, formal logic and discourse representation theory*,

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Kluwer Academic, Dordrecht ; Boston.

Kathol, A., Nakazawa, T. et Hinrichs, E. W., 1998, *Complex predicates in nonderivational syntax*, Academic Press, San Diego.

Keenan, E. O. et Bennett-Kastor, T., 1977, *Discourse across time and space*, Department of Linguistics University of Southern California, Los Angeles Calif.

Keenan, E. O., 1977, "Why look at planned and unplanned discourse ?" in *Discourse across time and space*, E. O. Keenan and T. Bennett-Kastor (eds), Department of Linguistics University of Southern California, Los Angeles Calif., pp. 304.

Kenny, A. J. P., 1963, *Action, emotion and will*, Routledge & K. Paul Humanities Press, London New York.

Kerbrat-Orecchioni, C. et Plantin, C. (eds), 1995, *Le trilogue*, Presses Universitaires de Lyon, Lyon.

Kerbrat-Orecchioni, C., 1990, *Les Interactions verbales Tome I*, A. Colin, Paris.

Kerbrat-Orecchioni, C., 1992, *Les Interactions verbales Tome II*, A. Colin, Paris.

Kerbrat-Orecchioni, C., 1995, "La construction de la relation interpersonnelle : quelques remarques sur cette dimension du dialogue", in *Cahiers de linguistique française,* Vol. 16:acte du VIème Colloque de pragmatique de Genève, 1e partie, pp. 69-88.

Kerbrat-Orecchioni, C., 1996, *La Conversation*, Seuil, Paris.

Kern, R., 1995, "Restructuring Classroom Interaction with Networked Computers: Effects on Quantity and Quality of Language Production", in *Modern Language Journal,* Vol. 79:4, pp. 457-476.

Klein, E. et Spiro, M. (eds), 1996, "*Le temps et sa flèche*", Champs, Flammarion, Paris, Vol. 339.

Klum, A., 1961, "*Verbe et adverbe; étude sur le système verbal indicatif et sur le système de certains adverbes de temps à la lumière des relations verbo-adverbiales dans la prose du français contemporain*", Almqvist & Wiksell, Stockholm.

Klum, A., 1975, "Rapports temporels, modes d'action et problèmes connexes", in *Studia Neopphilologica*, pp. 213-241.

Koch, P., 1995, "Subordination, intégration syntaxique et oralité", in *Etudes Romanes,* Vol. 34, pp. 13-42.

Koenig, J.-P. et Lambrecht, K., 1999, "French relative clauses as secondary predicates : a case study in construction theory", in *Empirical issues in formal syntax and semantics 2*, F. Corblin, D. Godard and J.-M. Marandin (eds), P. Lang, Bern ; New York, pp. 191-214.

Kroll, B., 1977, "Combining ideas in written and spoken English: A look at subordination and coordination", in *Discourse across time and space*, K. E. and B. T. (eds), pp. 69-108.

Labat, R., 1963, *Manuel d'épigraphie Akkadienne : signes, syllabaire, idéogrammes*, imprimerie nationale, Paris.

Labov, 1978, *Le Parler ordinaire : la langue dans les ghettos noirs des États-Unis*, Paris.

Lakoff, G., 1987, *Women, fire, and dangerous things : what categories reveal about the*

 Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

mind, University of Chicago Press, Chicago.

Lakoff, R., 1979, "Stylistic strategies within a grammar of style", in *Language, sex and gender*, O. J. e. al. (ed), pp. 53-78.

Lambert, M., 1967, "Ecritures idéographiques et syllabiques en Mésopotamie et en Islam", in *Comptes rendus du GLECS*, Geuthner, Paris.

Lambrecht, K., 1994, *Information structure and sentence form : topic, focus, and the mental representations of discourse referents*, Cambridge University Press, Cambridge ; New York NY USA.

Lambrecht, K., 1996, "On the Formal and Functional Relationship between Topics and Vocatives. Evidence from French", in *Conceptual structure, discourse, and language*, A. E. Goldberg (ed), CSLI Publications, Stanford Calif., pp. 267-288.

Lambrecht, K., 1998, "Sur la relation formelle et fonctionnelle entre topiques et vocatifs", in *Langues,* Vol. 1:1, pp. 34-45.

Lambrecht, K., 2000, "Prédication seconde et structure informationnelle : la relative de perception comme construction présentative", in *Langue française,* Vol. 127, pp. 49-66.

Langacker, R. W., 1987, *Foundations of cognitive grammar*, Stanford University Press, Stanford Calif.

Langacker, R. W., 1991, *Foundations of cognitive grammar*, Stanford university press, Vol. 2/2.

Langacker, R. W., 1995, "Viewing in cognition and grammar", in *Alternative linguistics : descriptive and theoretical modes*, P. W. Davis (ed), J. Benjamins, Amsterdam ; Philadelphia, pp. 153-212.

Langacker, R. W., 1999, *Grammar and conceptualization*, Mouton de Gruyter, Berlin ; New York.

Laufer, R. et Scavetta, D., 1992, *Texte, hypertexte, hypermédia*, Presses universitaires de France, Paris.

Le Goffic, P., 1993, *Grammaire de la phrase française*, Hachette supérieur, Paris.

Le Guern, M. et Rémi-Giraud, S. (eds), 1986, "*Sur le verbe*", Linguistique et sémiologie, Presses universitaires de Lyon, Lyon.

Le Guern, M., 1986, "Notes sur le verbe français", in *Sur le verbe*, M. Le Guern and S. Rémi-Giraud (eds), Presses universitaires de Lyon, Lyon, pp. 357.

Le Querler, N., 1996, *Typologie des modalités*, Presses universitaires de Caen, Caen France.

Lebelle, C., 1938, *le latin en 100 pages. Grammaire Syntaxe Vocabulaire systématique*, Paul Ferran, Marseille.

Leeman-Bouix, D., 1994, *Grammaire du verbe français : des formes au sens*, Nathan, Paris.

Léon, M., 1984, "Erreurs et normalisation, les liaisons fautives en français contemporain", in *revue de phonétique appliquée*:69, pp. 1-9.

Léon, P. R., 1992, *Phonétisme et prononciations du français*, Nathan, Paris.

Léon, P., 1987, "E caduc, facteurs distributionnels et prosodiques dans deux types de

en vertu de la loi du droit d'auteur.

discours", *XI international congress of phonetic sciences*, Talinn, Estonia, Vol. 3, pp. 109-112.

Lerner, G., 1991, "On the syntax of sentences-in-progress", in *Language in Society,* Vol. 20, pp. 441-458.

Levelt, W. J. M., 1989, *Speaking : from intention to articulation*, MIT Press, Cambridge Mass.

Levelt, W. J. M., 1993, *Lexical access in speech production*, Blackwell, Cambridge Mass.

Levine, J. R. et Baroudi, C., 1994, *Internet pour les nuls*, Sybex.

Littré, E. et Beaujean, A., 1960, *Dictionnaire de la langue française : abrégé du dictionnaire de Littré*, Gallimard Hachette, Paris.

Long, M. H., 1983, "Linguistic and conversational adjustments to non native speakers", in *Studies in Second Language Acquisition,* Vol. 5:2, pp. 177-193.

Longacre, R. E., 1999, "Grammatical units", in *Concise encyclopedia of grammatical categories*, E. K. Brown and J. Miller (eds), Elsevier, Amsterdam ; New York, pp. 172-176.

Loufrani, C., 1981, "Locuteur collectif ou locuteur tout court", in *Recherches sur le français parlé,* Vol. 3, pp. 215-243.

Loufrani, C., 1985, "Le locuteur collectif : typologies de configurations discursives", in *Recherches sur le français parlé,* Vol. 6, pp. 169-193.

Lucci, V. et Millet, A. (eds), 1994, "*L'orthographe de tous les jours : enquête sur les pratiques orthographiques des Français*", Politique linguistique, Paris.

Luzzati, D., 1991, "Oralité et interactivité dans un écrit Minitel", in *Langue Française*:89, pp. 99-109.

Mac Luhan, M., 1968, *Pour comprendre les médias : les prolongements technologiques de l'homme*, Le Seuil, Paris.

MacWhinney, B., 1995, *The CHILDES project : tools for analyzing talk*, L. Erlbaum, Hillsdale N.J.

Maingueneau, D., 1986, *Eléments de linguistique pour le texte littéraire*, Bordas, Paris.

Malotki, E., 1983, *Hopi time : a linguistics analysis of the temporal concepts in the Hopi language*, Mouton, Berlin ; New York.

Mann, W. C. et Thompson, S. A., 1992, *Discourse description : diverse linguistic analyses of a fund-raising text*, J. Benjamins Pub. Co., Amsterdam ; Philadelphia.

Marchal, A., 1980, *Les sons et la parole*, Guérin, Montreal.

Martin, R., 1987, *Langage et croyance : les "univers de croyance" dans la théorie sémantique*, P. Mardaga, Bruxelles.

Martinet, A., 1986, *Des steppes aux océans : l'indo-européen et les "indo-européens"*, Payot, Paris.

Mayfield Tomokiyo, L., 2000, "Linguistic Properties of non-native speech", *ICASSP 2000*, Istambul, Turkey.

Mayfield Tomokiyo, L., 2001, "*Recognizing non-native speech: Characterizing and*

en vertu de la loi du droit d'auteur.

adapting to non-native usage in speech recognition", Interactive Systems Lab, Carnegie Mellon University, Pittsburgh.

Michaelis, L. A., à paraître, "Type Shifting in Construction Grammar: An Integrated Approach to Aspectual Coercion." in *Cognitive Linguistics.*

Milner, J. C., 1978, *De la syntaxe à l'interprétation : quantités, insultes, exclamations*, Seuil, Paris.

Milroy, L., 1987, *Language and social networks*, B. Blackwell, Oxford, Vol. 2.

Mondada, L., 2001, "Pour une linguistique interactionnelle", in *Marges Linguistiques,* Vol. 1, pp. 1-21.

Morel, M. A. et Danon Boileau, L., 1998, *Grammaire de l'intonation l'exemple du français*, Ophrys., Paris France.

Mortureux, M.-F., 1997, *La lexicologie entre langue et discours*, Sedes, Paris.

Muller, C., 1992, *Initiation aux méthodes de la statistique linguistique*, Champion.

Nelson, T. H., 1978, "Dream Machines: New Freedoms Through Computer Screens--A Minority Report." in *Computer Lib.*

Nelson, T. H., 1990, "On the Xanadu Project", in *BYTE Magazine,* Vol. 15:9, pp. 298-299.

Nelson, T., 1965, "A File Structure for the Complex, The Changing and The Indeterminate", *ACM 20th National Conference*, pp. 84-100.

Nettle, D. et Romaine, S., 2000, *Vanishing voices : The extinction of the World's Languages*, Oxford University Press.

Niklas-Salminen, A., 1997, *La lexicologie*, Armand Colin, Paris.

Ochs, E., 1979, "Planned and unplanned Discourse", in *Discourse and syntax*, T. Givón (ed), Academic Press, New York, pp. 51-80.

O'Donnell, R. C., 1974, "Syntactic difference between speech and writing", in *American Speech,* Vol. 49, pp. 102-110.

Palmer, F. R., 1986, *Mood and modality*, Cambridge University Press, Cambridge Cambridgeshire ; New York.

Parisse, C. et Le Normand, M.-T., 2000, "Automatic disambiguiation of morphosyntax in spoken language corpora", in *Behavioural Research Methods, Instruments, & Computers,* Vol. 32:3, pp. 468-481.

Parodi, C. (ed) 1996, "*Aspects of Romance linguistics : selected papers from the Linguistic Symposium on Romance Languages XXIV March 10-13, 1994*", Georgetown University Press, Washington D.C.

Payne, J. R., 1985, "Complex phrases and complex sentences", in *Language typology and syntactic description*, T. Shopen (ed), Cambridge University Press, Cambridge ; New York, pp. 3-41.

Perlmutter, D. M., 1983, *Studies in relational grammar*, University of Chicago Press, Chicago.

Picoche, J. et Firmin, G., 1992, *Dictionnaire étymologique du français*, Dictionnaires Le Robert, Paris.

Picoche, J. et Honeste, M.-L., 1992, "Un dictionnaire des mots français de haute

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

fréquence - Projet et bilan provisoire", *Actes du XXe congrès international de Linguistique et Philologie romanes*, Zurich, Vol. Tome IV - section VI -, pp. 269-281.

Picoche, J. et Marchello-Nizia, C., 1998, *Histoire de la langue française*, Nathan, Paris.

Pierce, J. R. et Carroll, J. B., 1966, "Language and Machines", in *Computers in Translation and Linguistics.*

Plantin, C., Doury, M. et Traverso, V. (eds), 2000, "*Les émotions dans les interactions*", PUL/ARCI, Lyon.

Prince, E., 1992, "The ZPG letter: subjects, definiteness, and information-status", in *Discourse description : diverse linguistic analyses of a fund-raising text*, W. C. Mann and S. A. Thompson (eds), J. Benjamins Pub. Co., Amsterdam ; Philadelphia.

Prince, E., 1997, "On the functions of left dislocation in english discourse", in *Directions in functional linguistics.*, A. Kamio (ed), John Benjamins, Philadelphia/Amsterdam, pp. 117-44.

Prince, E., 1997, "On the limits of syntax, with reference to Left-Dislocation and Topicalization", in *The limits of syntax*, P. Culicover and L. McNally (eds), Academic Press, New York, pp. 281-302.

Pujol, N., 1993, *Corpora : éléments pour un guide juridique*, CNRS.

Raible, W., 1992, "Junktion. Eine Dimension des Sprache und ihre Realisierungsformen zwischen Aggregation und Integration.".

Reichenbach, H., 1947, *Elements of symbolic logic*, Macmillan Co., New York.

Rey, A., 1977, *Le lexique : images et modèles, Du dictionnaire à la lexicologie.*

Riegel, M., Pellat, J.-C. et Rioul, R., 1994, *Grammaire méthodique du français*, Presses universitaires de France, Paris.

Robert, P., Rey-Debove, J. et Rey, A., 1993, *Le nouveau petit Robert : dictionnaire alphabétique et analogique de la langue française*, Dictionnaires Le Robert, Paris.

Rosch, E., 1976, "Classification d'objets du monde réel : Origines et représentation de la cognition," in *Bulletin de Psychologie,* Vol. Numéro spécial:La mémoire sémantique, pp. 242-235

Rosch, E., 1978, "Principles of Categorization", in *Cognition and categorization*, E. Rosch, B. B. Lloyd and Social Science Research Council (U.S.) (eds), L. Erlbaum Associates, distributed by Halsted Press, Hillsdale N.J., New York, pp. 28-47.

Rosch, E., Lloyd, B. B. et Social Science Research Council (U.S.), 1978, *Cognition and categorization*, L. Erlbaum Associates.

Ross, J. R., 1967, "*Constraints on variables in syntax*", Massachusetts Institute of Technology, Boston.

Rossi, M., 1981, *L'Intonation : de l'acoustique à la sémantique*, Klincksieck, Paris.

Roulet, E., 1981, "Echanges, interventions et actes de langage dans la structure de la conversation", in *Etudes de linguistique appliquée,* Vol. 44, pp. 7-39.

Roulet, E., 1986, "Complétude interactive et mouvements discursifs", in *Cahiers de linguistique française,* Vol. 7, pp. 189-206.

Roulet, E., 1987, "Complétude interactive et connecteurs reformulatifs", in *Cahiers de linguistique française,* Vol. 8, pp. 111-140.

 Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé

Roulet, E., 1989, "De la structure de la conversation à la structure d'autres types de discours", in *Modèles du discours. Recherches actuelles en suisse romande*, C. Rubattel (ed), Lang, Berne.

Roulet, E., Auchlin, A., Moeschler, J., Schelling, M. et Rubattel, C. (eds), 1987, "*L'Articulation du discours en français contemporain*", P. Lang, Bern ; New York.

Rude, N., 1986, "Graphemic Classifiers", in *Noun classes and categorization*, C. Craig (ed), John Benjamins, pp. 133-138.

Ruwet, N., 1982, *Grammaire des insultes et autres études*, Seuil, Paris.

Sacks, H. et Jefferson, G., 1992, *Lectures on conversation*, Blackwell, Oxford UK ; Cambridge Mass.

Sacks, H., Schegloff, E. et Jefferson, G., 1974, "A simplest systematics for the organization of turn-taking for conversation", in *Language*:50, pp. 696–735.

Sag, I. A. et Wasow, T., 1999, *Syntactic theory : a formal introduction*, Center for the Study of Language and Information, Stanford Calif.

Sapir, E., 1953, *Le langage : introduction à l'étude de la parole*, Payot, Paris.

Schachter, P., 1985, "Parts-of-speech systems", in *Language typology and syntactic description : Clause structure*, T. Shopen (ed), pp. 3-61.

Searle, J. R., 1979, *Expression and meaning : studies in the theory of speech acts*, Cambridge University Press, Cambridge Eng. ; New York.

Shopen, T. (ed) 1985, "*Language typology and syntactic description*", Cambridge University Press, Cambridge ; New York/3.

Sottas, H. et Drioton, E., 1987, *Introduction à l'étude des hiéroglyphes*, Paris.

Spradley, J. P., 1980, *Participant observation*, Holt Rinehart and Winston, New York.

Sten, H., 1952, *Les temps du verbe fini (indicatif) en francais moderne*, Munksgaard, København.

Steyer, R., 1998, *HTML 4 et Dynamique*, Micro Application PC Poche, Paris, Vol. 1/1

Sueur, J.-P., 1979, "Une analyse sémantique des verbes *devoir* et *pouvoir*", in *Le français Moderne,* Vol. 47:2, pp. 97-120.

Talmy, L., 2000a, *Toward a Cognitive Semantics : Concept Structuring Systems*, MIT Press, Cambridge, Mass., Vol. 1/2.

Talmy, L., 2000b, *Toward a Cognitive Semantics : Concept Structuring Systems*, MIT Press, Cambridge, Mass., Vol. 2/2.

Tanenbaum, A., 1997, *Réseaux*, Interéditions, Paris.

Tenny, C., 1994, *Aspectual roles and the syntax-semantics interface*, Kluwer Academic, Dordrecht ; Boston.

Tesnière, L., 1969, *Eléments de syntaxe structurale*, Paris.

Toffler, A., 1980, *La Troisième vague*, Denoël.

Touratier, C., 1996, *Le système verbal français : description morphologique et morphématique*, A. Colin, Paris.

Tranel, B., 1995, "Current issues in French phonology", in *The handbook of phonological theory*, J. A. Goldsmith (ed), Blackwell, Cambridge Mass. USA, pp.

en vertu de la loi du droit d'auteur.

798-816.

Tranel, B., 1996, "French liaison and elision revisited: A unified account within optimality theory", in *Aspects of Romance linguistics : selected papers from the Linguistic Symposium on Romance Languages XXIV March 10-13, 1994*, C. Parodi (ed), Georgetown University Press, Washington D.C., pp. 433-455.

Trask, R., 1999, "Parts of speech", in *Concise encyclopedia of grammatical categories*, E. K. Brown and J. Miller (eds), Elsevier, Amsterdam ; New York, pp. 278-284.

Traverso, V., 1995, "Gestion des échanges dans la conversation à trois participants", in *Le trilogue*, C. Kerbrat-Orecchioni and C. Plantin (eds), Presses Universitaires de Lyon, Lyon, pp. 29-53.

Traverso, V., 1996, *La conversation familière : analyse pragmatique des interactions*, Presses universitaires de Lyon, Lyon.

Trudgill, P. (ed) 1978, "*Sociolinguistic patterns in British English*", E. Arnold, London, 186 p.

Trudgill, P., 1983, *Sociolinguistics : an introduction to language and society*, Penguin, Harmondsworth Middlesex England ; New York N.Y. U.S.A.

Vendler, Z., 1967, *Linguistics in philosophy*, Cornell University Press, Ithaca N.Y.

Vendryes, J., 1923, *Le langage; introduction linguistique à l'histoire*, Albin Michel, Renaissance du livre, Paris.

Verkuyl, H. J., 1972, *On the compositional nature of the aspects*, Reidel, Dordrecht.

Vet, C., 1980, *Temps, aspects et adverbes de temps en français contemporain : essai de sémantique formelle*, Droz, Genève.

Walter, H., 1977, *La phonologie du français*, Presses universitaires de France, Paris.

Walter, H., 1984, "L'innovation lexicale chez les jeunes Parisiens", in *La Linguistique,* Vol. 20:2, pp. 69-84.

Walter, H., 1988, *Le français dans tous les sens*, R. Laffont, Paris.

Ward, G. et Prince, E., 1991, "On the topicalization of indefinite NPs." in *Journal of Pragmatics,* Vol. 16:2, pp. 167-78.

Weinrich, H., 1973, *Le Temps : le récit et le commentaire*, Éditions du Seuil, Paris.

Whorf, B. L., 1970, *Language, thought, and reality; selected writings*, Technology Press of Massachusetts Institute of Technology, Cambridge Mass.

Willems, D., 1998, "Données et théories en linguistiques : réflexions sur une relation tumultueuse et changeante", in *Analyse linguistique et approches de l'oral : recueil d'études offert en hommage à Claire Blanche-Benveniste*, M. Bilger, K. v. d. Eynde, C. Blanche-Benveniste and F. Gadet (eds), Peeters, Leuven.

Wilmet, M., 1980, "Aspect grammatical, aspect sémantique et aspect lexical : un problème de limites", in *La notion d'aspect*, J. David, R. Martin and B. Pottier (eds), Klincksieck, Paris.

Wilmet, M., 1998, *Grammaire critique du français*, Hachette supérieur Duculot, Paris Louvain-la-Neuve.

Witko-Commeau, A., 1995, "Du trilogue dans le polylogue", in *Le trilogue*, C. Kerbrat-Orecchioni and C. Plantin (eds), Presses Universitaires de Lyon, Lyon, pp.

en vertu de la loi du droit d'auteur.

284-305.
Wright, G. H. v., 1951, *An essay in modal logic*, North-Holland Pub. Co., Amsterdam.

Copyright CHENU Florence et Université Lumière - Lyon 2 - 2003.Ce document est protégé