

CHAPITRE II

LA BASE DE DONNÉES PROSODIQUE

1. OBJECTIFS

En premier lieu, la constitution d'une base de données permet, non pas de supprimer, mais de contraindre la subjectivité de l'analyse, et de la contrôler. Par ailleurs elle offre les conditions idéales de vérification des hypothèses. Elle doit donc faciliter la mise en oeuvre de l'analyse, en optimisant l'extraction des connaissances portant sur le contenu du corpus, prosodique et linguistique.

Ce chapitre a donc pour objectif de décrire l'organisation de la base de données conçue en fonction de notre projet. On considérera ainsi la structure du corpus, la méthode de segmentation et d'étiquetage, et le contenu des étiquettes.

2. REALISATION DU CORPUS

Le corpus constitue la base de l'expérimentation. Il est la résultante d'un ensemble de facteurs tels que le texte, les consignes de lecture (qui substituent, aux conditions naturelles d'élocution, d'autres conditions pragmatiques telles que la lecture de textes sous certaines contraintes explicites), l'enregistrement lui-même, le consentement des locuteurs à la réalisation de cette tâche, et le bon accomplissement de cette dernière.

2.1. LE TEXTE DU CORPUS

Une décision importante concerne le type de discours (lu ou spontané) à retenir pour l'expérimentation, décision qui suscite toujours des débats entre les chercheurs. Ces débats de toute façon n'ont pas grand intérêt car ces différents choix présentant chacun des avantages et des inconvénients, aucun n'est idéal.

Le principal avantage du discours spontané est sans doute le naturel de l'élocution, le locuteur adhérant au contenu qu'il exprime, même s'il s'agit, bien entendu, de l'expression d'un doute. Pour recueillir un échantillon suffisant du phénomène observable, il faut cependant recueillir des textes généralement longs, par autant de locuteurs voulus. Cette pratique se généralise de plus en plus, mais il faut bien dire que les opérations d'étiquetage en ralentissant beaucoup le processus d'analyse, pénalisent considérablement les recherches. En outre de tels corpus, étant par nature différents les uns des autres, dans leurs motivations, expression, style, contenu, destinataire, conditions pragmatiques de réalisation (etc.) peuvent limiter fortement la reproductibilité du processus soumis à l'étude.

A l'inverse, le choix d'un texte d'énoncé — ensemble textuel ou ensemble de phrases sans lien sémantique —, a pour énorme avantage de renfermer tout un ensemble d'éléments conçus pour la vérification d'hypothèses, le tout dans un espace réduit, autorisant de ce fait un étiquetage linguistique très complet.

Après avoir décidé entre discours lu ou spontané, se pose alors la question de son type, et corrélativement, du naturel de l'élocution. De préférence aux phrases, nous avons opté pour un texte qui selon Searle (1983), réalise une "unité minimale de communication".

En effet à la différence d'une suite de phrases qui ne constitue pas un espace linguistique

clos, le texte représente une entité avec une argumentation interne, une structuration cohérente, dans laquelle s'incarne naturellement un sens. Par suite la communication de ce sens à autrui est perçue davantage comme allant de soi. A l'inverse, les phrases hors contexte hors situation risquent ou bien d'être lues mécaniquement lorsque le locuteur ne s'investit pas dans sa lecture, ou bien d'être lues emphatiquement lorsque ce dernier pallie cette difficulté en exagérant son investissement. Notre étude étant de nature lexicale, donc locale, le choix s'est d'emblée orienté vers un texte de composition, solution qui nous a semblé offrir un bon compromis entre les conditions de nature¹ optimales du discours spontané et les conditions minimales qu'offrent les phrases sans lien sémantique ni justification naturelle.

Lorsque l'on étudie la prosodie, deux types d'analyses ont cours : soit on s'attache à quelques processus ciblés sur des corpus importants (de l'ordre de plusieurs heures), soit on projette d'analyser le discours en profondeur, et dans ce cas, les corpus sont beaucoup plus restreints. Pour notre part nous avons opté selon les études, pour l'une et l'autre solution. L'étude principale de cet ouvrage porte sur un corpus restreint de lecture, mais le dernier chapitre intègre des analyses en discours spontané.

Concernant le corpus de lecture, nous avons pu étudier non seulement les modifications des paramètres prosodiques induites par les 12 locuteurs, par les consignes de lecture différentes, mais aussi évaluer de nombreux indices prosodiques et modèles, et explorer ainsi les stratégies des locuteurs. De ce fait, nous avons pu poser dans un cadre bien circonscrit, un problème précis : celui des relations entre plan syntaxique, plan sémantique et plan prosodique dans les stratégies de lecture développées sous certaines conditions pragmatiques. Comme on le comprend, notre étude ne concerne pas le texte pour lui-même, mais les interprétations qui en sont faites dans les 36 énoncés.

Pour ce faire, nous avons donc utilisé le contenu d'un article paru dans la revue de vulgarisation scientifique *Sciences et Vie* (1979) et recomposé le texte en fonction de nos objectifs. Le choix de ce type de texte nous fournissait des éléments intéressants : l'authenticité d'un fait qui légitimait un faire-croire (donc assurait une prosodie plus naturelle), le caractère nouveau et inattendu de l'information (découverte de vers géants marins), leurs conditions de vie surprenantes, et de surcroît l'utilisation d'un vocabulaire diversifié, allant de mots courants aux plus diversifiés. Voici le texte de l'énoncé :

D'éminents biologistes et d'éminents zoologistes américains ont créé pour des vers géants un nouveau phylum dans l'actuelle classification des nombreuses espèces vivantes. Ces longs vers prospèrent sur le plancher marin des zones sous-marines profondes. Des sources thermales chaudes y maintiennent une température moyenne élevée. Caelen-Haumont (1991).

2.2. ANALYSE DU TEXTE

Ce texte est composé d'une cinquantaine de mots¹ dont 30 lexicaux. Il est par ailleurs constitué de 3 phrases que nous avons voulues de longueur différente : la première est la plus longue et avec ses 15 unités lexicales, elle représente la moitié du texte; les phrases 2 et 3 sont sensiblement égales, puisqu'elles comportent respectivement 8 et 7 mots lexicaux.

¹ Indépendamment de l'unité de sens qu'il constitue, nous considérons le mot comme unité comprise entre blancs, éléments de graphie (apostrophe, guillemets) ou de ponctuation. Les mots grammaticaux correspondent dans le texte aux prépositions, déterminants, conjonctions de coordination, auxiliaires de verbe, adverbes de lieu ("y"). Les mots lexicaux sont des les adjectifs, les participes passés, les verbes, les substantifs.

Autre disparité, celle des constituants minimaux, ou constituants syntaxiques dominant directement le niveau superficiel : les phrases 1 à 3 en contiennent respectivement 7, 4 et 3. Ceci a pour effet de varier le niveau de hauteur de ces constituants dans la hiérarchie syntaxique, puisque par exemple se donnent rendez-vous des constituants majeurs comme des syntagmes nominaux à fonction sujet (ou SN1, ex : *d'éminents biologistes*), et d'autres constituants de niveau inférieur comme des groupes prépositionnels (GP, ex : *des nombreuses espèces vivantes*) issus de syntagmes eux-mêmes prépositionnels (SP).

un adjectif	PH1	PH2	PH3	deux adjectifs	PH1	PH2	PH3
antérieur	3	1	/	antérieur et postérieur	2	/	/
postérieur	1	1	/	postérieurs	/	1	2
Total	4	2	/	Total	2	1	2

Tableau n° 2-1 : Distribution des adjectifs par rapport au substantif. Les nombres dans les colonnes précisent le nombre d'adjectifs par phrase selon leur position par rapport au nom (par exemple, antérieur).

mots lexicaux	PH1	PH2	PH3	mots lexicaux	PH1	PH2	PH3
6 monosyllabes	1	3	2	3 quadrisyllabes	2	/	1
14 dissyllabes	7	4	3	1 pentasyllabe	1	/	/
6 trisyllabes	4	1	1				

Tableau n° 2-2 : Longueur des mots lexicaux. Les nombres dans les colonnes précisent le nombre de mots lexicaux par phrase selon qu'ils sont monosyllabiques ou plurisyllabiques.

La fonction syntaxique des groupes étant différente, il nous a semblé intéressant d'accumuler un même type, à savoir les groupes nominaux (prépositionnels ou non) dont la composition interne en français varie avec un degré de liberté assez important. Nous avons donc joué systématiquement sur le nombre et la position des adjectifs (cf tableau 2-1 ci-dessus), et également sur le nombre de syllabes des mots lexicaux (cf tableau 2-2 ci-dessus).

2.2.1. LES GROUPES MINIMAUX

Sur le plan prosodique les constituants minimaux (nominaux et verbaux) peuvent se regrouper et constituer des groupes un peu plus longs (dans ce cas leur statut devient pseudo-syntaxique). Dans notre corpus, les verbes, du fait de leur petit nombre de syllabes, constituent des unités pouvant se rattacher, selon les stratégies, au constituant de droite ou de gauche. Ce phénomène de libre association du verbe avec le SN sujet ou objet selon les locuteurs, est un problème souvent abordé par les auteurs en analyse ou synthèse. Citons à titre d'exemple, pour l'anglais Suci (1967), Hillinger, James, Zell et Prato (1976), Cooper & Paccia-Cooper (1980), Grosjean et Dommergues (1983), mais aussi pour le français, Dell (1984), Martin (1987).

On sait que plusieurs facteurs interviennent. Outre le débit, la recherche des moyens expressifs, on sait que lorsque le facteur de la dépendance syntactico-sémantique la plus forte n'est pas remis en cause par le principe eurythmique (eurythmie ou principe d'équilibre syllabique), le regroupement s'opère de manière syntaxique :

- (1) a PH3 : SN // GV + GN2 : "des sources thermales chaudes // y maintiennent une température moyenne élevée",

b PH1 : SN // GV + GP datif : "d'éminents biologistes et d'éminents zoologistes américains // ont créé pour des vers géants".

Pour cette dernière phrase, le regroupement GV + GP est cependant d'autant mieux assuré que le SN initial constitué de 2 GN coordonnés est long (18 syllabes). Mais dans le cas contraire ou lorsque les dépendances à droite et à gauche sont du même ordre, alors le rattachement qui semble naturel se fonde sur le critère de l'équilibre syllabique :

(2) c PH2 : SN + GV // GP : "ces longs vers prospèrent // sur le plancher marin".

Dans ce cas, les groupes minimaux ne sont plus des groupes syntaxiques, mais des groupes "naturels" ou pseudo-syntaxiques.

En fonction de ces divers regroupements, les 3 phrases de notre texte comptent donc 11 groupes minimaux (11 groupes par 12 locuteurs, soit 132), et l'étude, nous le rappelons, ne concerne que les unités lexicales de ces groupes.

Par ailleurs, comme on le sait, dans l'étude de la relation entre modèles et indices, ces groupes constituent notre espace d'observation des stratégies. La méthode consiste alors à sélectionner les valeurs numériques des 7 indices les plus proches des valeurs prédictives des modèles. Il s'agit en fait de comptabiliser le nombre de mots lexicaux qui "réalisent" au mieux les prédictions d'un modèle (soit le modèle linguistique le mieux représenté numériquement, parmi les six). Ce comptage est opéré dans le cadre du syntagme minimal, puis du suivant et de leur ensemble, jusqu'à la macro-structure "syntaxique" ou la phrase, avec la contrainte de minimiser le nombre de modèles, le nombre d'indices et d'unifier les choix dans l'ensemble de l'énoncé.

Comme la phrase ne constitue plus le cadre de référence, les calculs de pourcentages porteront sur le nombre total de groupes dans les 3 phrases, syntaxiques ou pseudo-syntaxiques. Après expérimentation, comme précisé dans l'introduction, le cadre d'analyse qui nous a semblé le mieux adapté aux réalisations des locuteurs, était celui du groupe, syntaxique lorsque le nombre des syllabes était suffisant, ou pseudo-syntaxique dans le cas de regroupements "naturels" entre un groupe verbal comportant un petit nombre de syllabes et un groupe nominal voisin court.

En ce domaine, nos observations (Caelen-Haumont, 1989 a) convergent avec ce qui est généralement admis par les auteurs, à savoir que le groupe pseudo-syntaxique ne peut exister à moins de 5 syllabes, ce qui, dans notre texte, est le cas des groupes verbaux.

Tous ces groupes syntaxiques ou pseudo-syntaxiques constituent de toute façon des "groupes prosodiques minimaux" ou en abrégé "groupes minimaux" (GM). De manière à uniformiser la méthode d'analyse et de calcul, nous avons retenu sur la base des réalisations les plus courantes chez les locuteurs, trois GM pseudo-syntaxiques soit en phrase 1, *ont créé pour des vers géants*, en phrase 2, *ces longs vers prospèrent*, et en phrase 3, *y maintiennent une température moyenne élevée*.

Comme on le sait, les groupes prosodiques se constituent d'autant plus facilement qu'ils sont introduits par des déterminants, prépositions ou mots-outils, ceux-ci se comportant comme des articulateurs linguistiques et prosodiques.

Nous avons décrit ci-dessus le type de structuration syntaxique et lexicale élémentaire de notre corpus qui au sein d'une disparité des phrases, et d'une variabilité de la position de l'adjectif par rapport au nom, offre une accumulation de groupes nominaux. Dans le chapitre

suisant nous traiterons plus à fond les aspects syntaxiques, lexicaux et analyserons la composition sémantique du texte.

2.2.2. LES RYTHMES SOUS-JACENTS

Avant d'entreprendre l'analyse de ces différents aspects, il est nécessaire de présenter les structures sous-jacentes susceptibles de rythmer l'énonciation des locuteurs. En fait, il semble que ces rythmes d'énonciation, correspondant sans doute aux rythmes d'encodage des unités linguistiques, reposent sur deux modes de structuration principaux. Ces deux modes prennent effet sur les articulations syntaxiques et sémantiques majeures. Selon les phrases, ou le niveau des constituants, ces diverses articulations peuvent être confondues.

En fait il semble qu'interviennent deux processus d'encodage des groupes : un processus d'équilibre syllabique (ou eurythmique), et un processus plus syntagmatique. Ce dernier processus prendrait appui sur le groupe minimal (de nature syntaxique ou pseudo-syntaxique c'est-à-dire prosodique), ce dernier fonctionnant comme unité de base. Selon la structuration syntaxique et sémantique de la phrase, des regroupements seraient sollicités, formant selon le cas des motifs binaires ou ternaires (et unaires lorsque le nombre de groupes minimaux dans la phrase est impair). Aucun schéma particulier ne correspondrait a priori à une structuration syntaxique ou sémantique, l'une et l'autre étant soumise à une grande variabilité de réalisation dans les textes.

En ce qui concerne la lecture, il semblerait que le processus d'équilibre syllabique interviendrait à plus court terme dans la constitution des groupes minimaux alors que le processus linguistique agirait à plus long terme en impulsant un motif récurrent d'encodage de ces groupes minimaux, binaire ou ternaire. Bien entendu lorsque la structure du texte autorise la combinaison de ces deux processus, la cadence peut dans les réalisations des locuteurs, se révéler de manière plus évidente.

Ainsi les réalisations des locuteurs peuvent être encodés selon trois arguments, syntaxique, sémantique et rythmique (cadence). En fonction du texte de l'expérimentation, les regroupements possibles sont les suivants :

- (3) a Plan syntaxique
 - PH1 Groupes 1 et 2 : 2 GN1 coordonnés (18 syllabes)
Groupes 3 et 4 : GV + datif + GN2 (13 syllabes)
Groupes 5 et 6 : GP à deux constituants (16 syllabes)
 - PH2 Pas de regroupement conforme à la syntaxe
 - PH3 Groupe 11 : SV en 2 constituants (13 syllabes)
- b Plan sémantique
 - PH1 Groupes 1, 2 et 3 : thème majeur (26 syllabes)
Groupes 4, 5 et 6 : rhème majeur (21 syllabes)
 - PH2 Groupes 8 et 9 : rhème majeur (13 syllabes)
- c Cadence
 - PH1 Pas distinct des plans syntaxique ou sémantique
 - PH2 Groupes 7 et 8 : SN1 + GV + GP (11 syllabes)
Groupes 7, 8 et 9 : ensemble de la phrase (18 syllabes)
 - PH3 Groupes 10 et 11 : ensemble de la phrase (18 syllabes).

Le texte de PH1, comme on le voit ci-dessus, possède un groupement syntaxique binaire (2 groupes), un groupement sémantique énonciatif ternaire (3 groupes), et l'un et l'autre répondent aux principes de l'équilibre syllabique.

En PH2, quand le mécanisme du regroupement est opéré dans le cadre des groupes 7 et 8, aucun argument syntaxique ou sémantique ne pouvant être retenu, c'est alors vraisemblablement le principe de cadence qui joue (2 groupes minimaux), impulsée par le rythme et le motif des regroupements syntaxiques de la phrase 1 (3 fois 2 groupes). Opéré dans l'ensemble de la phrase (3 groupes), il semble que ce mécanisme soit enclenché à la fois en vertu de lois de symétrie eurhythmique (équilibre syllabique), et en vertu de lois de cadence reposant sur la reprise du motif sémantique de la phrase 1 (3 groupes). Ce sont ces mêmes principes d'eurhythmie et de cadence, mais cette fois binaires et syntaxiques, qui se retrouvent en phrase 3 lorsque les locuteurs ne dissocient pas les groupes 10 et 11.

2.3. ENREGISTREMENTS

Les enregistrements se sont déroulés dans une pièce peu bruyante (située dans bâtiment administratif donnant sur une cour intérieure), sur un magnétophone Radiola N 4420. Dans cette pièce était installé un ordinateur, "auditeur" fictif mis en scène par la troisième consigne. Le lecteur était seul en présence de l'expérimentateur.

Une séance d'enregistrement se déroulait de la manière suivante. Le locuteur est assis à une table face à l'expérimentateur. Sur celle-ci un micro est placé de face à 25 cm des lèvres du locuteur. Le texte dactylographié est présenté sur une feuille blanche. Celui-ci prend connaissance du texte, lit ou relit à loisir le texte et si besoin est, une explicitation des mots spécialisés lui est donnée. En effet il est capital de préciser que l'expérience ne porte pas sur la compréhension du texte, mais sur la *production* : il s'agit d'étudier quelles stratégies utilise le locuteur pour communiquer un contenu textuel à un tiers.

Pour réaliser les trois consignes, trois feuilles sont présentées successivement aux locuteurs, sur lesquelles sont dactylographiés à chaque fois le même texte et quelques lignes plus bas, le type de consigne demandé (présenté sans numérotation). Le locuteur ne sait jamais donc qu'une autre consigne suit celle qu'il vient d'achever. Les consignes sont les suivantes :

- consigne 1 : lecture naturelle et intelligible
- consigne 2 : lecture très intelligible
- consigne 3 : lecture très très intelligible pour un ordinateur.

2.4. LES LOCUTEURS

Sur 13 locuteurs ayant enregistré, 12 locuteurs ont été retenus. Le treizième a été écarté car la première consigne n'a pas été respectée : la lecture produite n'étant ni naturelle, ni ce qu'il est convenu d'appeler "intelligente". Ces lecteurs (6 locuteurs, 6 locutrices) sont tous personnels d'établissements scolaires et ont une formation universitaire (sauf F2). L'ensemble des locuteurs fait donc partie d'un groupe socioculturel homogène :

Locutrices		Locuteurs	
F1 : secrétaire	F4 : enseignante	M1 : technicien	M4 : enseignant
F2 : secrétaire	F5 : étudiante en thèse	M2 : enseignant	M5 : enseignant
F3 : enseignante	F6 : intendante	M3 : enseignant	M6 : enseignant

3. ETIQUETAGE

La phase de segmentation et d'étiquetage est une phase particulièrement délicate car dans la parole les unités phonétiques ne se juxtaposent pas, mais se chevauchent et subissent en plus des influences réciproques qui rendent leurs frontières approximatives. L'étiquetage étant une affaire d'appréciation, il faut donc se donner des règles pour tenter de contraindre cette subjectivité, règles qui favorisent des décisions homogènes d'un locuteur à un autre, d'une séance à une autre. Pour tendre vers une plus grande uniformité, il est donc nécessaire de revenir à plusieurs reprises sur les données de segmentation et d'étiquetage.

Dans le cadre de notre expérimentation, cette opération était appuyée par une présegmentation automatique (Caelen, 1979 ; Caelen et al., 1986). Sur le critère d'une continuité acoustique, celle-ci proposait un premier découpage qui pouvait être validé ou non par la tâche d'expertise. La figure 2-3 ci-dessus présente un extrait de parole issu de la base de données, et étiqueté. De gauche à droite, on lit les informations suivantes :

- a le numéro de la trame de l'échantillon, calculé toutes les 10 ms,
- b les valeurs de l'énergie en dB (valeurs audessus d'un seuil de bruit),
- c les frontières des phases acoustiques stables, posées automatiquement (avec des chevrons précisant le degré d'instabilité), ou manuellement (avec un dièse)
- d la courbe de l'énergie,
- e le sonagramme squelettisé,
- f la courbe de F0
- g l'indice de continuité vs. discontinuité spectrale
- h les 6 champs d'étiquettes posées manuellement.

Pour les 3 consignes, environ 40 000 étiquettes ont été posées. Se répartissant dans 6 champs, elles notent une ou plusieurs informations : pondération sémantique, pondération syntaxique, nature de l'unité phonétique et phases (établissement, tenue, coda), frontières du "contour" mélodique, structuration du groupe accentuel, structuration morphosyntaxique (syllabe lexicale / grammaticale coïncidant avec début de mot, syllabe interne, fin de mot, début de phrase, début de groupe syntaxique; syllabe monosyllabique vs. plurisyllabique).

Cette recherche de cohérence n'est pas sans raison car sur la phase d'étiquetage repose la légitimité de la recherche entreprise : si par exemple les frontières de mots sont déplacées, alors que les valeurs à ces bornes sont généralement des extrema, toute l'architecture des modèles linguistiques prédictifs et des indices prosodiques peut s'effondrer : les écarts de F0 peuvent devenir incohérents, de même que les variations de durée et d'énergie.

