Introduction aux statistiques en sciences du langage

Sous la direction de
Clara Solier, Lucille Soulier
et Nour Ezzedine

Introduction aux statistiques en sciences du langage

Traitement et analyse de données avec R

DUNOD

Mise en page : Belle Page

NOUS NOUS ENGAGEONS EN FAVEUR DE L'ENVIRONNEMENT :

Nos livres sont imprimés sur des papiers certifiés pour réduire notre impact sur l'environnement.

Le format de nos ouvrages est pensé afin d'optimiser l'utilisation du papier.

Depuis plus de 30 ans, nous imprimons 70 % de nos livres en France et 25 % en Europe et nous mettons tout en œuvre pour augmenter cet engagement auprès des imprimeurs français.

Nous limitons l'utilisation du plastique sur nos ouvrages (film sur les couvertures et les livres).

11 rue Paul Bert, 92240 Malakoff
www.dunod.com
ISBN 978-2-10-085264-2

Liste des auteurs

Auteurs coordinateurs :

Clara Solier	Docteure en Sciences du Langage, chercheure post-doctorale, Basque Center on Cognition, Brain and Language (BCBL) et Laboratoire de NeuroPsychoLinguistique (LNPL, UR 4156), Université Toulouse Jean Jaurès.
Lucille Soulier	Maîtresse de conférences en psychologie, LIRDEF, Université de Montpellier, Université Paul Valéry Montpellier 3, Montpellier.
Nour Ezzedine	Orthophoniste et Doctorante en Sciences du Langage, NeuroPsychoLinguistique (LNPL, UR 4156), Université Toulouse Jean Jaurès, ToNIC, Toulouse NeuroImaging Center, Université de Toulouse, Inserm, UPS.

Auteurs contributeurs :

Brigitte Bigi	Chargée de recherche, Laboratoire Parole et Langage, UMR 7309 CNRS, Aix-Marseille Université.
Jean-François Camps	Maître de conférences, Université de Toulouse Jean Jaurès, Laboratoire de NeuroPsychoLinguistique (LNPL, UR 4156).
Leonardo Contreras Roa	Maître de conférences, Université de Picardie Jules Verne, CORPUS – Conflits, représentations et dialogues dans l'univers anglo-saxon (EA 4295).
Jérémy Danna	Chargé de recherche, Laboratoire Cognition, Langues, Langage, Ergonomie (CLLE), UMR 5263, Université Toulouse Jean Jaurès.
Elie Fabiani	Docteure en Neurosciences Cognitives, Laboratoire de Neurosciences Cognitives, UMR 7291, CNRS, Aix-Marseille Université.

Kamel Gana	Professeur des universités, Université de Bordeaux, PHARes – MéRISP Bordeaux Population Health (BPH), Centre INSERM, UMR 1219.
Alain Ghio	Ingénieur de recherche, Laboratoire Parole et Langage, UMR 7309 CNRS, Aix-Marseille Université.
Jean-Claude Gilhodes	Laboratoire de Neurosciences Cognitives, UMR 7291, CNRS, Aix-Marseille Université.
Manuel Gimenes	Maître de conférences, Université de Poitiers, Centre de Recherches sur la Cognition et l'Apprentissage, UMR 7295, CERCA.
Marieke Longcamp	Professeure des Universités, Aix-Marseille Université, Laboratoire Parole et Langage, UMR 7309, CNRS.
Thierry Olive	Directeur de Recherche au CNRS, Centre de Recherches sur la Cognition et l'Apprentissage, UMR 7295, CNRS et Université de Poitiers.
Cyril Perret	Maître de conférences, Université de Poitiers, Centre de Recherches sur la Cognition et l'Apprentissage, UMR 7295, CERCA.
Jean Pylouster	Ingénieur d'études informaticien au CNRS, Centre de Recherches sur la Cognition et l'Apprentissage, UMR 7295, CNRS et Université de Poitiers.
Christiane Soum-Favaro	Maîtresse de conférences, Université de Toulouse Jean Jaurès, Laboratoire de NeuroPsychoLinguistique (LNPL, UR 4156).
Nicolas Stefaniak	Maître de conférences, Université de Reims, Laboratoire C2S, Cognition Santé Société.
Joseph Tisseyre	Maître de conférences, Université Toulouse III – Paul Sabatier, ToNIC, Toulouse NeuroImaging Center, Université de Toulouse, Inserm, UPS.
Jean-Luc Velay	Chargé de Recherche, Laboratoire de Neurosciences Cognitives, UMR 7291, CNRS, Aix-Marseille Université.

Table des matières

Avant-Propos

1. Pourquoi ce livre ?

Les sciences du langage se situent à la croisée de plusieurs disciplines : linguistique, psychologie, orthophonie, neurosciences, etc. Elles nous invitent à étudier le langage et les langues à travers leur fonctionnement, leur évolution, leur acquisition et leurs dysfonctionnements, en considérant leurs dimensions linguistiques, psycholinguistiques et neuropsycholinguistiques. La grande diversité de leurs approches et des domaines étudiés implique une grande variabilité de données issues de l'oral et de l'écrit.

Cet ouvrage est le premier ouvrage francophone sur le traitement et les analyses statistiques spécifiquement appliquées aux sciences du langage. Il s'adresse aux nombreux universitaires et professionnels travaillant autour de la question du langage et se veut une ressource pédagogique qui manquait aux étudiants et aux enseignants. À travers nos différentes expériences d'enseignement dans différentes filières, nous avons pu le constater : **les enseignements dispensés ne suffisent pas toujours pour répondre aux besoins des étudiants** et ces derniers rencontrent des difficultés importantes lorsqu'ils doivent appliquer les connaissances acquises en statistiques à un sujet particulier et de façon autonome, par exemple lors de la réalisation de leur mémoire ou de leur thèse. Or, le traitement et l'analyse des données constituent des étapes clefs dans la réalisation de toute recherche scientifique. **Des compétences dans ce domaine s'avèrent indispensables pour la poursuite d'un cursus universitaire** jusqu'au niveau master ou doctorat, mais aussi pour répondre aux exigences actuelles de publication et de valorisation des résultats de recherche nécessaires à la poursuite d'une carrière professionnelle de chercheur ou d'enseignant-chercheur. C'est pourquoi nous proposons un ouvrage qui présente les bases théoriques nécessaires pour choisir et comprendre les analyses statistiques appliquées aux sciences du langage.

Convaincues que l'apprentissage nécessite un engagement actif de la part de l'apprenant, cet ouvrage comprend également un volet pratique. Pour que le lecteur puisse réaliser ses analyses statistiques de **manière autonome**, cet ouvrage fournit les bases nécessaires pour prendre en main le logiciel

de traitement statistique R qui, en plus d'être libre de droits, s'avère être un des logiciels les plus utilisés dans le monde académique.

Pour toutes ces raisons, cet ouvrage introductif est particulièrement adapté à toutes les personnes débutantes en statistiques travaillant dans le domaine du langage et leur permettra de devenir autonomes dans la recherche de solutions et dans leurs futurs apprentissages. Bien que ce livre ait été conçu pour être accessible aux débutants, **il constitue un point de départ dans la formation en statistiques et n'a pas pour vocation de rendre le lecteur expert en statistiques.** En effet, la maîtrise des statistiques nécessite une formation continue et une pratique régulière et assidue, alimentée par une diversité de ressources.

Ce manuel vous permettra de :

- comprendre les fondements méthodologiques liés au recueil et au traitement des données recueillies en sciences du langage ;
- connaître les bases théoriques des analyses statistiques couramment utilisées dans l'étude du langage ;
- prendre en main le logiciel R ;
- manipuler des jeux de données spécifiques aux sciences du langage ;
- reproduire les analyses présentées grâce à des scripts R commentés.

2. Organisation de ce livre

Cet ouvrage est divisé en deux grandes parties. Dans la première, l'objectif est de donner un aperçu de la diversité des données qui peuvent être recueillies en sciences du langage et de leur traitement.

Les chapitres 1 et 2 sont consacrés aux données du langage oral. Ils présentent le matériel et les techniques d'enregistrement les plus courants utilisés dans des conditions variées de recueil de données orales ainsi qu'une vue d'ensemble des différentes méthodes et des différents outils disponibles pour la transcription, l'annotation et l'analyse des données de langage oral.

Les chapitres 3 et 4 sont consacrés au langage écrit. Le chapitre 3 présente un panorama des principales méthodes et techniques qui peuvent être utilisées pour étudier les processus cognitifs impliqués dans la production du

langage écrit, du tracé de lettres à la production de textes, ainsi que pour analyser la dynamique du texte « en train de se faire ». Le chapitre 4 expose les problématiques relatives à l'utilisation des tablettes et décrit la manière d'acquérir et d'analyser l'ensemble des variables de l'écriture manuscrite.

Dans la deuxième partie de cet ouvrage sont présentées différentes méthodes d'analyse statistique illustrées avec des jeux de données issus d'études ou d'exemples en sciences du langage.

Le chapitre 5 porte sur la prise en main du logiciel R : apprendre à naviguer dans l'environnement de R au travers de son interface RStudio, manipuler les fonctions et préparer les données de manière efficace. Si vous êtes novice en statistiques et que vous découvrez R, cette partie est indispensable.

Le chapitre 6 est dédié aux principes de base en statistiques inférentielles. Il présente la logique sur laquelle reposent les tests utilisés en statistiques inférentielles tout en définissant les notions clés associées ainsi que les limites liées à leur utilisation et à l'interprétation parfois abusive de leurs résultats.

Le chapitre 7 est consacré à la présentation des tests inférentiels bivariés classiquement utilisés dans les recherches expérimentales en sciences du langage.

Les chapitres 8 et 9 présentent les bases théoriques et les spécificités des modèles mixtes de régression linéaire et logistique. Ces types de modèles sont particulièrement adaptés pour traiter des données en sciences du langage. Ils permettent d'analyser des mesures répétées et des variables continues et catégorielles, mais aussi de prendre en compte la variabilité liée aux items et aux sujets.

Le chapitre 10 présente la méthode de classification ascendante hiérarchique qui permet de créer un groupe d'individus *(cluster)* qui rassemble des individus semblables et qui se distinguent le plus possible des individus d'autres groupes.

Enfin, le chapitre 11 décrit en détail les protocoles à cas unique. Il met en exergue les principes de base ainsi que les caractéristiques essentielles des protocoles expérimentaux à cas unique (PECU), en présente les différents *designs* et en discute les fondements méthodologiques en comparaison avec d'autres traditions de recherche.

3. Utilisation de ce livre

Pour aider le lecteur et faciliter sa compréhension, tous les chapitres de cet ouvrage contiennent :

- Une rubrique de synthèse reprenant les **points essentiels à retenir**.
- **Des ressources pour aller plus loin** permettant au lecteur d'approfondir ses connaissances.

Cet ouvrage propose également des contenus numériques à consulter en ligne :

- Des contenus complémentaires aux chapitres.
- Des jeux de données.
- Des scripts R commentés.

L'ensemble du contenu numérique mis à disposition est accessible en suivant ce lien : https://www.dunod.com/ean/9782100852642.

Pour aller plus loin et mettre toutes les chances de votre côté, des compléments sont disponibles sur le site www.dunod.com.

Connectez-vous à la page de l'ouvrage (grâce aux menus déroulants, ou en saisissant le titre, l'auteur ou l'ISBN dans le champ de recherche de la page d'accueil). Sur la page de l'ouvrage, sous la couverture, cliquez sur le logo « LES + EN LIGNE ».

4. Remerciements

D'un projet à un ouvrage… Nous souhaitons remercier plusieurs personnes pour leur soutien et leur accompagnement dans ce projet : Cyril Perret qui nous a encouragées à porter ce projet, Romain Mendez qui a participé au démarrage du projet et les personnes qui ont accepté de relire une partie de ce travail.

Nous remercions particulièrement nos chers contributeurs et chères contributrices qui ont permis de concrétiser ce projet. Merci pour leur expertise, leur disponibilité, leur confiance et le temps qu'ils ont accordé à cet ouvrage.

Partie 1

Traiter ses données

Sommaire

Chapitre 1

Étudier le langage oral : les différents types de données recueillies

Alain Ghio et Brigitte Bigi

Sommaire

Résumé

Étudier le langage oral présente un défi de taille comparé au langage écrit car son expression n'implique pas de traces discrètes écrites directement exploitables. La façon dont cette trace va être créée doit être adaptée aux objectifs de l'étude envisagée. La première étape, de grande importance, consiste à utiliser un matériel et des conditions d'enregistrement adaptés aux analyses prévues. La diversité de données orales peut aller de la simple prononciation de syllabes isolées (dans une situation hypercontrôlée) à l'enregistrement de plusieurs locuteurs au cours d'une conversation familière. Entre ces deux conditions, une multitude de situations est possible. De plus, les données de parole n'ont parfois que peu d'intérêt si elles ne sont pas mises en lien avec des contextes particuliers qu'il faut pouvoir identifier lors de la phase d'analyse ou avec des locuteurs précis dont les caractéristiques doivent être sauvegardées. Si une grande partie des corpus oraux sont constitués de lecture (avec un contenu connu), une part croissante des travaux sur l'oral spontané nécessitent une première étape de transcription. De plus, le seul signal acoustique ne permet pas toujours de décrire totalement les séquences langagières produites car la parole est le résultat complexe de mécanismes respiratoires, phonatoires et articulatoires. Il est donc parfois nécessaire de recueillir des données physiologiques en complément du signal de parole. Ce chapitre présentera le matériel et les techniques d'enregistrement les plus courantes utilisés dans des conditions variées de recueil de données orales. Ces aspects seront décrits puis illustrés avec différents corpus.

Introduction

L'étude du langage oral nécessite la plupart du temps de recueillir des données : l'approche est empirique sauf si l'on se situe dans un cadre complètement théorique, ce qui n'est pas l'objet de cet ouvrage. Cette démarche observationnelle peut être inductive et consister, par exemple, à collecter des corpus « ouverts… sans délimiter à l'avance un objet de recherche prédéterminé » (Baude, 2006, p. 28). Dans ce cadre, le chercheur fait le pari de mettre en évidence *a posteriori* des phénomènes linguistiquement pertinents, impossibles à provoquer artificiellement ou imprévisibles *a priori*. Cette approche est souvent liée à un contexte « naturel » d'élocution, dont les résultats n'auraient

aucun sens en dehors de ce contexte spontané ou semi-spontané. Une autre méthode empirique est l'approche hypothético-déductive, qui part d'une ou de plusieurs hypothèses de travail, qui se poursuit par l'acquisition de données « sollicitées » (Baude, 2006, p. 26), et dont l'analyse a pour but l'explication des hypothèses de départ. On associe souvent l'approche inductive à la linguistique de terrain, de même qu'on fait un parallélisme fréquent entre la linguistique expérimentale et l'approche hypothético-déductive. Cette vision binaire doit être nuancée par toute une palette de situations variées. Ainsi, on pourrait penser que la phonétique clinique s'inscrit totalement dans un cadre expérimental hypothético-déductif alors qu'elle s'approche d'un travail de terrain : conditions hospitalières s'éloignant du confort du laboratoire, évaluations cliniques à la volée à l'image de pratiques de la sociolinguistique, recours à des questionnaires, population très hétérogène proche des enquêtes de terrain. On peut même rapprocher l'attitude du clinicien de celle du sociolinguiste dans son rapport au locuteur car « sur certains terrains, la constitution du corpus est le résultat d'une relation de confiance qui s'est établie entre le chercheur en tant que personne… et la communauté ou certaines personnes de la communauté, dans des contextes difficiles » (Baude, 2006, p. 35). D'autre part, les hypothèses généralement formulées au commencement de l'approche déductive ne sortent pas du néant et sont souvent le fruit d'un raisonnement inductif à partir d'observations répétées. Le va-et-vient entre approche inductive et vérification déductive est habituellement un gage de réussite. De même que le mélange entre situation de laboratoire et travail de terrain est souvent générateur d'observations pertinentes et originales.

Pour obtenir de nombreux détails sur la constitution de corpus oraux, nous invitons le lecteur à consulter le guide des bonnes pratiques de Baude (2006). Quelle que soit l'approche choisie, il existe un certain nombre de préconisations à respecter ou tout au moins desquelles il faut s'approcher.

1. Les préconisations préalables

Un recueil de données, comme bon nombre d'autres réalisations, est régi par la règle « C.Q.F.D. », pour *Coûts, Qualité, Fonctionnalités, Délais*. Cette vision montre les tensions qui peuvent exister entre les principaux critères. Dans un monde idéal, il serait possible d'obtenir des données « chics et pas chères », c'est-à-dire avec « F » élevé et « C » faible, et de le faire « vite et

bien », c'est-à-dire « D » faible et « Q » élevé. Malheureusement, cet espoir est bien souvent déçu du fait de contraintes de temps, de difficultés logistiques, de conditions d'enregistrement délicates, de participants en nombre insuffisant, de consignes mal appliquées… De plus, tout recueil de données devrait être effectué en intégrant les principes *FAIR – Findable, Accessible, Interoperable, Reusable*, qui « décrivent comment les données doivent être organisées pour être plus facilement accessibles, comprises, échangeables et réutilisables[1] ».

Dans un monde idéal, au préalable du recueil des données, les chercheurs du projet auront défini le cahier des charges et décrit le contexte empirique. Ces deux documents détermineront les contraintes scientifiques qui amèneront à définir les conditions du recueil.

Le cahier des charges est un document formalisant les besoins, les objectifs, les contraintes, les fonctionnalités attendues, les délais et le budget prévisionnel si nécessaire. C'est généralement le responsable de l'étude qui est en charge de la rédaction d'un cahier des charges. Concrètement, il sert de base pour la planification et le pilotage du recueil des données de parole. Ce cahier des charges est fonctionnel : il vise à dessiner les contours du recueil des données, il définit les besoins auxquels les futures données devront répondre, en termes de fonctionnalités. Ainsi, par exemple, il s'agira de savoir si les enregistrements de parole auront pour but uniquement d'être transcrits pour étudier le contenu verbal ou si des détails acoustiques et phonétiques seront analysés, ce qui nécessite une qualité du signal bien plus exigeante que dans le premier cas où l'enregistrement doit être simplement audible pour être transcrit. Toutefois, le principe du « qui peut le plus peut le moins » doit être le plus souvent adopté car le critère de réutilisabilité des données, notamment dans un cadre scientifique différent du cadre original, dépendra de la qualité des données.

La méthode empirique utilisée (le protocole dans une approche expérimentale) met l'accent sur la partie technique du recueil. Il s'appuie sur les besoins fonctionnels en les traduisant en exigences techniques et fait émerger les différentes contraintes. Pour le recueil de données de parole, on doit quasi systématiquement définir des contraintes liées à la qualité audio. Ces contraintes amèneront, ultérieurement, à décider des conditions d'enregistrement (chambre sourde ou non, etc.), et/ou du choix du microphone, par exemple.

1. https://www.ccsd.cnrs.fr/principes-fair/

2. Les diverses données de parole

Les données de parole peuvent revêtir de nombreux aspects (figure 1.1). Le **signal acoustique** issu d'un enregistrement sonore reste central et incontournable. Brut, il peut être directement utilisable pour être **transcrit**, opération qui consiste à passer de la forme orale à une forme écrite du discours. Ce signal peut être soumis à différentes procédures d'**analyse acoustique** *via* des logiciels qui vont fournir des données spectrales, fréquentielles, énergétiques, temporelles issues du signal de parole. Ce signal peut aussi être utilisé comme stimulus linguistique à des expériences de perception dans lesquelles des auditeurs vont fournir des **réponses perceptives**. Si ces réponses sont explicites (l'auditeur fait un choix, apporte un jugement, une transcription…), ces données sont comportementales. Si l'auditeur est équipé d'un casque d'électroencéphalographie, il est possible d'obtenir des réponses évoquées par le stimulus.

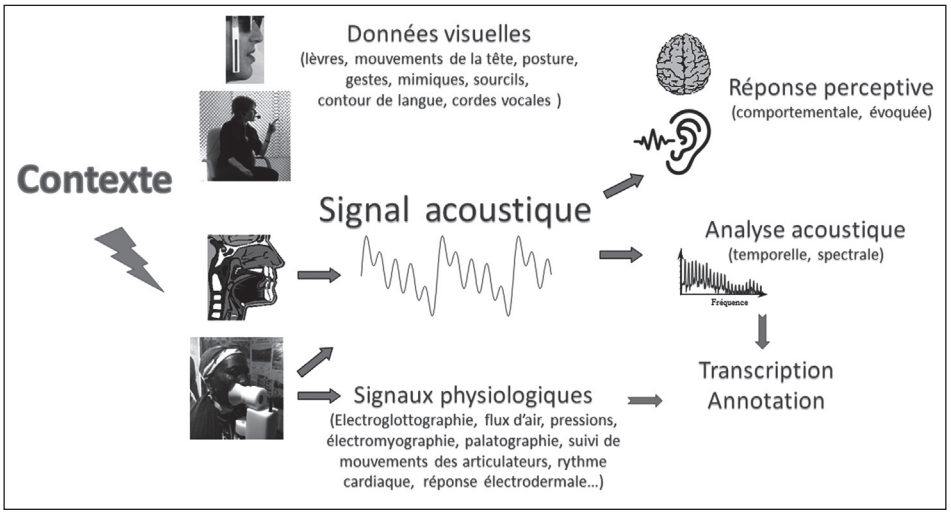

Figure 1.1 – Représentation synthétique des différentes données de parole

En complément du signal acoustique, il est possible de recueillir des **données visuelles** comme le mouvement des lèvres, de la tête, des sourcils ainsi que les mimiques faciales, le comportement postural, les gestes des bras, des mains… En contexte hospitalier, on peut obtenir des images du conduit vocal (IRM anatomique dynamique) ou du mouvement des cordes vocales (vidéo-endoscopie). Pour être pleinement exploitables, il est impératif que ces données visuelles soient parfaitement synchronisées avec le

signal acoustique de manière à mettre correctement en lien les phénomènes observés visuellement avec le signal de parole. De même, il est possible de recueillir en plus du signal sonore des **données physiologiques** comme le signal d'accolement des cordes vocales par électroglottographie (EGG), les flux d'air qui passent par la cavité orale ou nasale (aérophonométrie), les pressions intraorale ou sous-glottique, l'activité musculaire par électromyographie (EMG), le contact entre la langue et le palais (palatographie), le suivi de mouvement des organes articulateurs (électro-magnéto-articulographie), le rythme cardiaque, la réponse électrodermale… Là encore, la synchronie des signaux doit être parfaite pour correctement interpréter les données physiologiques et le signal de parole.

La plupart du temps, les données complémentaires visuelles ou physiologiques sont elles-mêmes annotées non pas dans le contenu verbal mais dans leurs spécificités. Par exemple, les gestes sont repérés et mis en relation avec la parole, de même pour des signaux EGG, EMG, aérophonométriques… On parle alors d'analyse multiparamétrique dans la mesure où divers signaux issus de la parole sont utilisés simultanément dans l'analyse. Pareillement, l'analyse acoustique va elle-même fournir différents signaux secondaires qui seront observés de façon multiparamétrique (figure 1.2) : courbe d'intensité, courbe de fréquence fondamentale (F_0), représentations spectrales (figure 1.4), etc.

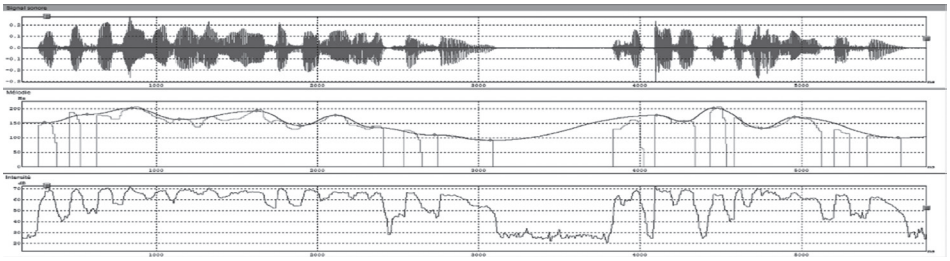

Figure 1.2 – Locuteur masculin ayant prononcé « Monsieur Seguin n'avait jamais eu de bonheur avec ses chèvres # il les perdait toutes de la même façon »

Note : en haut le signal sonore, au milieu la fréquence fondamentale brute (courbe claire interrompue) et modélisée (courbe continue foncée reliée par des points), en bas l'intensité en dB (données obtenues avec le logiciel Phonedit-SESANE, LPL, Aix-en-Provence).

3. La prise de son

La prise de son est l'opération fondamentale pour l'étude de la parole. Le choix et le positionnement du microphone sont décisifs. Selon les objectifs du travail, il sera préférable d'utiliser un microphone **omnidirectionnel** qui captera le son de tous les côtés (figure 1.3). Cela sera intéressant si plusieurs locuteurs sont concernés et si l'on ne dispose que d'un seul microphone. Cela pourra aussi être nécessaire si l'environnement sonore ambiant fait partie intégrante du corpus. À l'inverse, l'usage de microphones **directionnels** (hyper-super cardioïde ; figure 1.3) sera privilégié si l'observateur souhaite concentrer son enregistrement sur une source unique. L'usage de deux microphones serre-tête directionnels est une bonne solution pour enregistrer un dialogue dans lequel chaque locuteur sera enregistré sur une piste séparée (**enregistrement stéréo**). En revanche, l'enregistrement de plus de deux locuteurs sur des pistes séparées nécessite l'usage d'une **console multipistes** et de logiciels spécialisés qui sortent de l'ordinaire. Dans tous les cas, il est très difficile d'isoler complètement la parole d'un locuteur dans le cas d'une conversation à plusieurs car même si le microphone est directionnel et dirigé vers une cible, la parole des interlocuteurs sera malgré tout un peu captée mais à des niveaux très faibles.

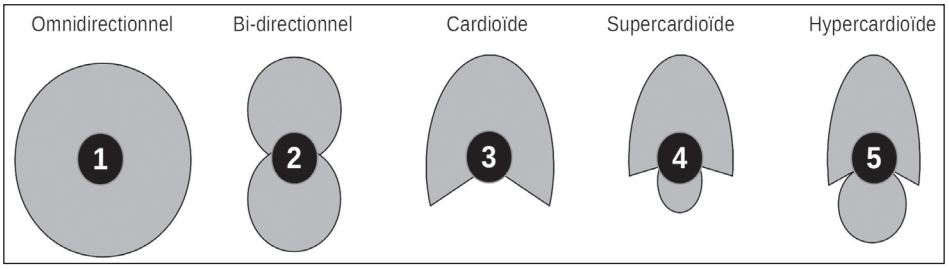

Figure 1.3 – Diagramme polaire des microphones

Le ou les microphones doivent être connectés soit à un enregistreur autonome (ex : Zoom H4n), soit à une **carte son externe** (ex : Focusrite scarlett) reliée à un ordinateur *via* l'USB. Pour éviter la captation de bruit, il est nécessaire que le **microphone** soit **le plus près possible de la source**. Les microphones serre-tête sont donc intéressants pour cela. Il faudra aussi veiller au format de numérisation du signal. Un format **non compressé PCM sur 16 bits** est un bon format pour la parole. Un **échantillonnage minimal de 16 kHz** est nécessaire (8 kHz de bande passante). La plupart du temps, les systèmes proposent un échantillonnage à 44,1 kHz ou 48 kHz.

Cela permet d'obtenir une bande passante très large, utile pour la musique mais pas forcément pour la parole. Une bande passante de 10 à 12 kHz est suffisante, ce qui implique un échantillonnage entre 20 et 25 kHz. Mais qui peut le plus peut le moins. Il faut aussi rappeler que les microphones les plus sensibles et ceux de bonne qualité sont le plus souvent issus d'une technologie électrostatique qui nécessite une **alimentation « phantom** ». Cette fonctionnalité doit être proposée par l'enregistreur utilisé et activée.

Des **bruits électromagnétiques** peuvent apparaître sur les enregistrements. Ces bruits ne sont pas captés par le microphone mais sont captés ou générés par l'électronique ou l'informatique utilisées. Dans ces cas-là, il faut arriver à isoler les sources de bruits électromagnétiques et, parfois, l'éloignement d'un appareil électrique voisin, d'un éclairage, la déconnexion du secteur des outils d'enregistrement et un fonctionnement sur batterie élimine le problème. Parfois, c'est le branchement de l'installation à la terre qui est la solution. Ces questions relèvent souvent de l'essai-erreur et il est difficile de donner des préconisations qui tiennent lieu de formule systématique.

L'utilisation d'un smartphone est à présent une solution possible grâce aux progrès des technologies audio développées dans ce cadre (Petrizzo *et al.*, 2020). En revanche, les enregistrements au smartphone doivent rester, jusqu'à nouvel ordre, des documents illustratifs ou des documents de terrain à transcrire uniquement ; il faut rester extrêmement prudent sur l'utilisation d'enregistrements sur smartphone dans une analyse phonétique fine.

Dans tous les cas, il est important avant de se lancer dans des enregistrements en série de **vérifier la qualité des données** sous la forme d'un contrôle qualité (voir la section 11 de ce chapitre).

4. Les données temporelles

Les données temporelles sont la plupart du temps des **durées de segments linguistiques** de natures et de tailles variées. Cela peut être le repérage temporel de tours de parole, d'unités prosodiques (Astésano *et al.*, 2016), d'unités lexicales, de syllabes, de phonèmes, de segments sous-phonémiques. La façon de segmenter le signal se fonde la plupart du temps sur des informations oscillographiques (signal de parole) et spectrographiques. L'obtention

des données temporelles est une sortie sous forme formatée des marques d'annotation avec leurs positions temporelles. La variété des mesures temporelles ne permet pas de les présenter toutes. Mais nous pouvons aborder quelques notions importantes.

Dans la segmentation temporelle du *continuum* acoustique, il est souvent nécessaire de faire la distinction entre segments de parole et silences. Des détecteurs automatiques permettent de faire cela, la plupart de temps sur la base de l'intensité du signal (voir par exemple Bigi *et al.,* 2022). Toutefois, il est important de pouvoir faire la distinction entre les pauses non remplies du discours et les tenues d'occlusives qui sont toutes les deux des parties silencieuses. Le critère généralement utilisé pour cela est un critère de durée. Grosjean et Deschamps (1975) considèrent une **pause comme silencieuse à partir de 250 ms**. Certains travaux dont ceux de Duez (1991) montrent que cette durée peut être raccourcie en fonction du débit du locuteur et peut être calculée à partir de la durée moyenne des occlusives intervocaliques.

En ce qui concerne la vitesse de parole, il peut être utile de distinguer le **débit de parole,** qui consiste à mesurer le nombre de phonèmes ou de syllabes produits sur la durée de l'énoncé, du **débit articulatoire,** qui est calculé par le nombre de phonèmes ou de syllabes produits sur la durée de l'énoncé, pauses exclues. Dans certains domaines de la linguistique, dont notamment les travaux sur l'interaction multimodale, il est d'usage de calculer un « débit gestuel » *(gesture rate),* qui consiste à compter le nombre de gestes effectués par rapport au nombre de mots produits dans le même énoncé. La notion de « normalisation » temporelle est souvent une précaution à prendre de façon à comparer ce qui est comparable.

5. Les données de fréquence fondamentale (F_0)

La fréquence fondamentale (F_0) de la parole est un paramètre important. Elle se mesure en Hertz et correspond à la vitesse de vibration des cordes vocales. Plus elle est élevée, plus la parole est perçue aiguë. Cette fréquence fondamentale varie au cours du temps et passe par des valeurs basses ou hautes qui constituent la mélodie de la parole (voir figure 1.2). Ces variations temporelles de F_0 sont essentielles à la structure prosodique et occupent des fonctions syntaxiques, contrastives, expressives, etc.

5.1 La détection de la F_0

L'étape préliminaire à la mesure de la F_0 consiste à repérer le voisement dans le *continuum* acoustique. Cette détermination des segments voisés/non voisés n'est pas triviale. La plupart des algorithmes de détection explicite de voisé/non-voisé sont fondés sur un ensemble de paramètres physiques tels que l'énergie, le taux de passage par zéro, le 1er coefficient LPC *(Linear Predictive Coding),* le facteur de balance grave/aigu, etc. dont les valeurs sont comparées à des seuils plus ou moins dynamiques. Des mécanismes de suivi temporel permettent ensuite de statuer sur le voisement ou pas des différents segments de parole.

Sur les parties détectées comme voisées, il est alors possible de mesurer la F_0. Il existe différentes techniques (Hess, 1983) : techniques temporelles d'autocorrélation, AMDF *(Average Magnitude Difference Function),* techniques spectrales d'analyse harmonique (méthode du peigne de Martin, 1982), analyse cepstrale (Noll, 1967). La plupart de ces méthodes sont fondées sur un calcul par trame, c'est-à-dire que le signal de parole est découpé en morceaux, et une valeur de F_0 est fournie sur chaque section de durée fixe comme, par exemple, toutes les dix millisecondes. Ce principe est suffisant pour ensuite obtenir des courbes mélodiques rendant compte des montées ou descentes de la F_0. En revanche, un tel procédé est insuffisant pour évaluer la stabilité à court terme du vibrateur laryngé. Effectivement, pour cela, il est nécessaire d'utiliser un détecteur instantané de cycle fondé, par exemple, sur du passage par zéro ou de la détection de pic. L'objectif est alors de « marquer » très précisément le début et la fin de chaque cycle vibratoire de façon à en calculer les variations instantanées. Inversement, il peut être utile de s'intéresser, non pas aux variations locales de fréquence, mais au contraire aux macro-variations mélodiques. On peut alors obtenir une courbe mélodique modélisée qui rend compte des aspects intonatifs de la parole comme avec l'algorithme MOMEL de Hirst *et al.* (2000), représenté par la courbe du milieu sur la figure 1.2.

5.2 Les problèmes de détection de la F_0

La plupart des problèmes de détection de F_0 portent sur deux catégories : le voisement et les sauts d'octave. La décision du voisé/non-voisé est un compromis délicat : soit le détecteur élimine trop de parties voisées en les considérant comme non voisées et, par conséquent, la F_0 n'est pas calculée sur ces parties intéressantes ; soit le détecteur favorise trop le voisement, et

des parties non voisées sont soumises à un calcul de F_0 qui se solde par des valeurs aberrantes car portant sur du bruit apériodique. Il peut être intéressant d'introduire une contre-réaction sur la décision voisé/non-voisé à la suite du calcul de F_0 en cas de valeurs aberrantes et de déclarer non voisées des suites anarchiques de F_0 dues, par exemple, à un calcul sur du bruit. Toutefois, cette solution peut s'avérer malvenue, notamment dans le cas de la parole pathologique. Dans ce cas précis, il est préférable d'introduire un réglage semi-manuel laissé à l'expertise de l'observateur.

Le deuxième problème fréquent dans la détection de F_0 est le saut d'octave. Dans les techniques temporelles, si une périodicité est observée pour un temps Δt, on considérera qu'il s'agit de la période fondamentale T_0, inverse de la F_0 (car $T = 1 / F$). Mais $2T_0$, $3T_0$, $4T_0$ constituent aussi des candidats à la périodicité. Si l'algorithme, pour diverses raisons, sélectionne comme résultat $T = 2\ T_0$, cela entraîne une erreur de calcul de fréquence fondamentale égale à $F_0 / 2$. La fréquence détectée est la moitié de la fréquence réelle. Inversement, dans les techniques spectrales de type analyse harmonique (méthode du peigne), si un espacement Δf constant est observé entre deux raies spectrales, on considérera qu'il s'agit alors de la fréquence fondamentale F_0, cette valeur étant l'écart systématique entre les harmoniques. Mais $2\ F_0$, constitue aussi un candidat à l'intervalle spectral. Si l'algorithme, pour diverses raisons, sélectionne comme résultat $F = 2\ F_0$, cela entraîne une erreur de calcul de fréquence fondamentale. La fréquence détectée est le double de la fréquence réelle. On comprend pourquoi il peut être intéressant de combiner deux techniques de détection complémentaires comme dans la proposition de Husson *et al.* (1998). L'ajout d'une troisième technique fondée sur l'analyse cepstrale (Noll, 1967) permet d'ajouter un élément de comparaison et d'améliorer la détection (Espesser, 1996).

5.3 Garder un œil critique et utiliser des valeurs adaptées aux locuteurs

La détection de F_0 n'est pas un processus sans faille comme peut l'être le calcul de l'intensité ou du spectrogramme. Il existe des seuils, des limites, des choix qui ont une incidence sur le résultat. Aussi, face à une courbe de F_0, il faut toujours garder un œil critique. Tout saut brutal ou toute montée/descente tronquée doit être vu avec circonspection. Une telle rupture peut être liée à des paramètres de détection mal adaptés, comme une F_0 maximale trop basse. Une bonne façon de s'affranchir de ces données aberrantes est d'utiliser une