

# ÉLIS

[Échanges Linguistiques en Sorbonne]

Volume 6 (2019)

La revue *Échanges Linguistiques en Sorbonne* est éditée par un ensemble de linguistes appartenant au Centre de Linguistique en Sorbonne (EA 7332).

<https://celiso.paris-sorbonne.fr/>

### ***Comité de rédaction***

Romain DELHEM, directeur de la publication

Mathilde GAILLARD, co-rédactrice en chef

Naomi TRUAN, co-rédactrice en chef

### ***Comité scientifique***

Antoine AUFRAY

Adrien BARBARESI

Justyna BERNAT

Katrina Ann BRANNON

Charles BRASART

Élise BONNIEC

Cécile DELETTRES

Romain DELHEM

Manar EL KAK

Célia HOFFSTETTER

Ernest HOUNHOUAYENOU-TOFFA

Aude LABETOULLE

Caroline MARTY

Pierre-Yves MODICOM

Kim OGER

Sébastien SAUDREAU

Eva SCHAEFFER-LACROIX

Antoine THOLLY

Naomi TRUAN

Hélène VINCKEL-ROISIN

Verena WEILAND

*Échanges Linguistiques en Sorbonne* est mis à disposition selon les termes de la Licence Creative Commons BY-NC (Attribution-Pas d'utilisation commerciale) 4.0.

ISSN 2425-1526

La revue *Échanges Linguistiques en Sorbonne*, portée par les jeunes chercheur·e·s en linguistique de Sorbonne Université, a le plaisir de vous présenter son nouveau numéro, le sixième volume depuis sa création. Cette année, les articles portent pour la plupart sur la langue française.

Les deux premiers articles sont des analyses diachroniques de l'anglais et du français. **Christelle Exare** propose une revue de la littérature sur les origines et la variabilité du phonème /h/ de l'anglais en diachronie et en synchronie. Elle se penche notamment sur les facteurs d'ordre extralinguistique et linguistique qui peuvent en expliquer l'évolution. **Laurie Raymond** mène, quant à elle, une réflexion autour du réseau hypéronymique relatif au vocable *insulte* de la naissance du français à nos jours. Elle interroge les définitions parfois lacunaires des dictionnaires et offre un schéma d'utilisation des termes précis.

**François Delafontaine** s'intéresse, pour sa part, à la frontière entre macro-syntaxe et discours ou interaction, et se demande ce que constitue l'unité maximale ou minimale dans chacune des deux approches. Enfin, dans une perspective sociolinguistique, **Fatma Ben Barka Messaoudi** se concentre sur un corpus de français oral pour déterminer l'évolution de l'usage du subjonctif à Orléans en microdiachronie. Elle suggère par ailleurs des outils méthodologiques pour analyser de tels corpus.

Ce volume sera le dernier pour notre comité de rédaction. Dès l'année prochaine, nous passerons la main à une équipe renouvelée : Caroline Marty prendra la direction de la revue, tandis que Manon Philippe, Olivia Reneaud-Jensen et Marie Turlais deviendront co-rédactrices en chef. Nous leur souhaitons bon vent !

Pour notre part, nous resterons bien entendu au comité scientifique de la revue. Nous remercions les auteur·e·s et les membres du comité scientifique pour leur travail remarquable et nous souhaitons à nos lecteurs et lectrices une très agréable lecture !

Le comité de rédaction  
Romain Delhem, Mathilde Gaillard et Naomi Truan

## SOMMAIRE

Christelle EXARE : <i>Variabilité du phonème /h/ de l'anglais en diachronie et en synchronie</i> .....	1
Laurie RAYMOND : <i>Des mots pour dire l'insulte (de la naissance du français à nos jours)</i> .....	23
François DELAFONTAINE : <i>Éléments de comparaison pour une correspondance entre unités macro-syntaxiques et interactionnelles</i> .....	44
Fatma BEN BARKA MESSAOUDI : <i>Constitution et annotation d'un corpus micro-diachronique de français oral en vue de l'étude du subjonctif</i> .....	65

# Variabilité du phonème /h/ de l'anglais en diachronie et en synchronie

Christelle Exare

INSPÉ de l'Académie de Créteil — Université Paris-Est Créteil

Laboratoire de Phonétique et Phonologie (UMR 7018)

[christelle.exare@gmail.com](mailto:christelle.exare@gmail.com)

## Résumé en français

Cet article est une revue de la littérature sur les origines et sur la variabilité du phonème /h/ de l'anglais, en diachronie et en synchronie. La mutation de /x/ (issu du proto-indo-européen \*k) vers /h/ a lieu lors de la transition du vieil anglais vers le moyen anglais. Les changements phonétiques aboutissant à la distribution lacunaire de /h/ sont décrits. Trois facteurs permettent d'expliquer l'instabilité de /h/ en anglais médiéval : la faiblesse intrinsèque de /h/, des biais méthodologiques et des paramètres linguistiques. En synchronie (au XX<sup>e</sup> siècle), la variation du /h/ anglais dépend de facteurs qui sont d'une part, extralinguistiques instables (le lieu d'origine, l'appartenance sociale et le style de parole d'un locuteur) et d'autre part, linguistiques assez stables (lexicaux, grammaticaux et phonétiques). L'évolution de /h/ en anglais est le résultat d'une tension entre deux forces qui s'opposent : sa lénition progressive liée à sa faiblesse perceptive, et des occurrences répétitives d'hypercorrection liées à son rôle de marqueur de hiatus.

**Mots-clés :** phonétique, variabilité, phonème /h/, fricative glottale, anglais

## Abstract in English

This article is a review of the literature on the origins and variability of the phoneme /h/ in English, diachronically as well as synchronically. The mutation of /x/ (deriving from Proto-Indo-European \*k) to /h/ took place during the transition from Old English to Middle English. The phonetic changes leading to the defective distribution of /h/ are described. Three factors can explain the instability of /h/ in medieval English: the intrinsic weakness of /h/, methodological bias, and linguistic parameters. In synchrony (i.e. in the 20<sup>th</sup> century), the variation of English /h/ depends on factors that are, first, extralinguistic and unstable (a speaker's origin, social class and speaking style), and second, linguistic and quite stable (i.e. lexical, grammatical and phonetic). The evolution of /h/ in English is the result of a tension between two conflicting forces: its gradual lenition, connected to its perceptual weakness, and some repetitive occurrences of hypercorrection, connected to its role as a hiatus breaker.

**Key words :** phonetics, variability, /h/ phoneme, glottal fricative, English

*Sound change is drawn from a pool of synchronic variation.*

Ohala (1989)

La prononciation de /h/<sup>1</sup> en anglais natif est un marqueur sociolinguistique qui évolue au cours du temps, et dont la labilité est persistante en synchronie. Trask (2003 : 106) écrit que, de nos jours, « pour la plupart des locuteurs anglais et gallois, le <h> dans *hair* et *head* est tout aussi mort que ceux de *light* et *loud* ». La chute de /h/ (*H dropping*, ou *aich dropping*) reste très stigmatisée et contraste avec une tendance à l’hypercorrection (c’est-à-dire à l’insertion de [h] illicite). Cet article propose une synthèse de la littérature portant sur la variabilité du phonème /h/ de l’anglais en diachronie et en synchronie. Une première section décrit l’émergence du phonème /h/ de l’anglais en Grande-Bretagne à partir du vieil anglais. La labilité de /h/ en diachronie peut être expliquée par trois facteurs : le caractère faible de [h] au plan acoustique, les biais méthodologiques issus de l’analyse des manuscrits médiévaux, et des paramètres linguistiques. Une deuxième section montre que la variabilité de /h/ en synchronie dépend de paramètres extralinguistiques et linguistiques.

## 1. Variation de /h/ en diachronie

L’anglais est une langue indo-européenne dont l’héritage germanique s’enrichit d’emprunts au latin, au grec, au flamand, au bas-allemand, au français, ainsi qu’aux langues celtiques et scandinaves (Crépin 1972 ; Bacquet 1974). D’après la Loi de Grimm, la forme reconstituée /k/ de l’indo-européen mute en /x/ en germanique tandis qu’elle est préservée dans les langues latines. Par exemple, \**kerd* de l’indo-européen devient *cor*, *cordis* en latin (puis « cœur » en français), mais *heorte* en vieil anglais et *heart* en anglais moderne (Larreya & Watbled 2004 : 85). L’étymologie des mots contenant <h> en anglais peut ainsi expliquer la variation dans les réalisations du phonème /h/ au fil du temps.

---

<sup>1</sup> Les graphèmes sont indiqués entre chevrons (*p. ex.* le graphème <h>). Les phonèmes sont indiqués entre barres obliques (*p. ex.* le phonème /h/). Les sons tels qu’ils sont réalisés dans la parole sont indiqués entre crochets (*p. ex.* le phone ou le son [h]). On dit qu’un phonème se réalise en un phone ou deux (*p. ex.* /h/ se réalise en [h] ou [ʔ]). On réserve le terme « épenthèse » au domaine phonologique et les termes « intrusion » ou « insertion » au domaine phonétique. Un effacement ou une élision de /h/ initial licite relève de la phonologie, car le phonème fait bien partie de la forme sous-jacente du mot, même s’il est absent de la forme de surface. En revanche, une insertion de [h] intrusif relève de la phonétique car l’aspiration initiale illicite n’est pas contenue dans la forme sous-jacente du lexème. Dans la partie portant sur le vieil anglais et le moyen anglais, les sources étant essentiellement manuscrites, /h/ et <h> sont censés correspondre, même si Milroy (2009 : 8) indique, à la suite de Scragg (1970), que le graphème <h> prévoicalique allitère parfois avec des voyelles.

## 1.1. Origines et évolution de /h/ dans l'histoire de la langue anglaise

### 1.1.1. /h/ en vieil anglais

Le vieil anglais est une langue parlée en Angleterre pendant toute la période de la suprématie anglo-saxonne, soit à partir du V<sup>e</sup> siècle après J-C, qui correspond au départ des Romains, jusqu'au XI<sup>e</sup> siècle avec la conquête normande.<sup>2</sup>

Le système fricatif du vieil anglais compte trois phonèmes sourds : /f/, /θ/ et /x/. Ils ont des allophones en distribution complémentaire. Les sons sourds apparaissent en position initiale et finale (p. ex. le mot <full>, *full* se prononce [full]), mais les allophones sonores font surface en position intervocalique (p. ex. <drīfan>, *drive* se prononce [dri:vən]).

Le graphème isolé <h> en vieil anglais correspond au phonème fricatif vélaire /x/. En position initiale de mot, ce phonème est réalisé comme la fricative vélaire sourde de l'allemand [x]. À la fin de la période, au moment de la conquête anglo-normande, cependant, la réalisation en [h] est attestée à l'initiale, comme une variante allophonique du phonème /x/, (p. ex. <hēah>) (Hogg 2002 : 9). En position intervocalique, la vélaire sourde /x/, comme les autres fricatives sourdes, a une variante sonore : [ɣ]<sup>3</sup>. En position finale, /x/ est toujours réalisé [x] (<sorh> *sorrow*).

<h> apparaît aussi dans des digraphes : <hl>, <hr>, <hn> et <hw>. Le phonème /x/ se retrouve dans des agrégats présentant des liquides (/hl/, /hr/), des nasales (/hn/) ou des approximantes (/hw/), c'est-à-dire des sons voisés.

Vieil anglais	Anglais contemporain
<hw> <i>hwæt</i>	<wh> <i>what</i>
<hr, hl, hn> <i>hring, hlūd, hnægan</i>	<r, l, n> <i>ring, loud, neigh</i>

Tableau 1. Correspondance des digraphes en vieil anglais et en anglais contemporain. Exemples tirés de Hogg (2002 : 10)

### 1.1.2. Évolution de /h/ du vieil anglais vers le moyen anglais

Le moyen anglais est parlé en Angleterre du XII<sup>e</sup> siècle au XVI<sup>e</sup> siècle. Le caractère germanique du vieil anglais décline au profit de l'influence des langues à son contact : le français, le latin et, dans une certaine mesure, le grec. Les flexions sont en voie de disparition et le vocabulaire se romanise, particulièrement grâce aux emprunts d'adjectifs au latin et au français, comme *honest*, *abundant*, *active*, *horrible*, *hasty*, *original*, *savage*, *usual* etc. (Crépin : 9 et 94). L'inventaire consonantique du moyen anglais est identique à celui de l'anglais moderne, à quelques exceptions près.<sup>4</sup>

<sup>2</sup> Cette section se fonde essentiellement sur Hogg (2002 : 4-12).

<sup>3</sup> /ɣ/ est aussi un phonème du vieil anglais apparaissant à l'initiale (p. ex. <guma>, *man*) et en position intervocalique (p. ex. <dagas>, *days*).

<sup>4</sup> Cette section se fonde essentiellement sur Lass & Laing (2010) et Horobin & Smith (2002).

Dans leur chapitre sur le moyen anglais, Millward et Hayes (2011 : 153) indiquent que /h/ est retrouvé à l'initiale des syllabes (p. ex. <high> est prononcé [hiç]) et en position intervocalique. Il disparaît souvent en position inaccentuée (p. ex. <hit> en vieil anglais devient <it> en moyen anglais).

Au plan graphémique, <h> disparaît dans les digraphes <hn->, <hl->, <hr->, tandis que <hw-> est préservé sous des formes orthographiques variées. Les lettres du digraphe peuvent s'inverser. Les modifications qui affectent /h/ dans les agrégats qu'il forme avec les sonantes, lors du passage du moyen anglais à l'anglais moderne, sont appelées *Glide Cluster Reduction*.

Vieil anglais		Moyen anglais		Anglais contemporain	
Graphème	Phonème	Graphème	Phonème	Graphème	Phonème
<hl->	/xl/	<l>	/l/	<l>	/l/
<hw->	/xw/	<hw->	/hw/	<wh->	/w/ : RP <sup>5</sup> <i>which, what</i> /h/ : RP <i>who</i> /ʍ/ : <i>when, whine</i> (anglais d'Écosse) /hw/ <sup>6</sup> : <i>what, which</i> (certains locuteurs RP)

Tableau 2. Glide cluster reduction, d'après Gimson (2001 : 214) et Wells (1982 : 228)

En position médiane et finale, <h> est remplacé par <gh> et <ȝ>. Par exemple, le mot *night*, qui est <niht> en vieil anglais, devient <niȝt> en moyen anglais. DE plus, <h> apparaît comme diacritique dans les groupes de l'anglais contemporain <ch>, <wh>, <sh> et <th>.

Au plan phonémique, /x/ est vocalisé en position finale. Autrement dit, il mute en voyelle, comme avec la diphtongue de *saw* qui s'écrit <saugh>. Préservé en position initiale, /h/ se phonémise. Il chute dans les agrégats initiaux /hn/, /hl/, /hr/, mais /hw/ est maintenu dans plusieurs dialectes, sous des formes orthographiques variées. Par exemple, le mot *what* est <hwæt> en vieil anglais. Il s'écrit <quhat> dans le Nord et en scots, mais <wat> dans le Sud.

### 1.1.3. Questions posées par l'analyse diachronique

Horobin & Smith (2002 : 43-44) expliquent comment s'effectue la reconstruction des états de langue antérieurs.

<sup>5</sup> RP est l'acronyme de *Received Pronunciation*. Autrement désigné par l'expression « anglais de la Reine », l'accent RP est à la fois la référence (utilisée par les dictionnaires) et la variété de prestige (parlée par les lettrés). L'anglais RP désigne une norme linguistique idéale qui, dans la réalité, tend à décliner au profit d'usages dialectaux pluriels. Dans la dernière édition de Gimson par Cruttenden (8<sup>e</sup> édition de 2014), « RP » a été remplacé par « GB » (*General British*).

<sup>6</sup> /hw/ est un agrégat consonantique formé de la fricative glottale sourde /h/ et de l'approximante labiovélaire sonore /w/, tandis que /ʍ/ est une fricative labiovélaire sourde. La différence phonétique entre les deux sons réside, au plan physiologique, dans l'ouverture glottale nécessaire pour /h/, qui peut être corrélée, au plan acoustique, à l'affaiblissement du premier formant.

Premièrement, sur la base de l'axiome selon lequel des langues ou dialectes apparentés partagent le même héritage, la reconstruction comparative permet de mettre des lexèmes en regard grâce à l'analyse de cognats. Par exemple, pour dire les mots « père » ou « poisson », la consonne initiale est une fricative dans les langues germaniques (*father*, *fish*), tandis que c'est une occlusive dans les langues indo-européennes dérivées du latin (*pater*, *piscis*).

Deuxièmement, la reconstruction interne, en se fondant sur des éléments résiduels d'une langue, donne des indications sur des usages passés d'un lexème. Par exemple, aujourd'hui, le mot *house* est prononcé avec [s] quand c'est un nom et [z] quand c'est un verbe. Or, on sait, d'une part, que le verbe du vieil anglais (*hūsian* à l'infinitif) est flexionnel, et d'autre part, que les occurrences de [z] dans les mots d'anglais contemporain dérivant du vieil anglais émergent à l'intervocalique avec [s] présent ailleurs. On en déduit la probable allophonie de [z] et [s] en vieil anglais.

Troisièmement, l'analyse métrique et rythmique des poèmes en vers donne des indications sur la prononciation des mots. Par exemple, Chaucer écrit sa poésie en pentamètres iambiques. La forme *gode* pour exprimer le mot *good*, et l'étude de la versification des contextes dans lesquels le mot apparaît, suggèrent que la marque orthographique *-e* est bien prononcée dans certains environnements grammaticaux au Moyen Âge. Dans les manuscrits écrits en vieil anglais, les occurrences de <h> devant une voyelle allitérant avec une voyelle sont généralement corrigés par la suite, car on suppose des erreurs scripturales. Ceci vaut pour les intrusions et les élisions (Scragg 1970 : 173).

Quatrièmement, l'analyse orthographique permet de comparer des langues dont le système d'écriture est phonographique. Il fait correspondre un graphème ou plusieurs graphèmes à un phonème. Par exemple, en vieil anglais, <f> représente une fricative labiodentale, sourde ou sonore, car [f] et [v] sont deux allophones du même phonème (p. ex. *yfel* pour *evil*, *fisc* pour *fish*). La distinction entre /f/ et /v/ émerge en moyen anglais, avec l'influence des emprunts au français (exp : *vine*, *fine*).

L'analyse diachronique pose plusieurs questions irrésolues ou partiellement résolues.

- Premièrement, les digraphes <hl, hr, hn > étaient-ils prononcés comme un son (une liquide ou une nasale dévoisée [l̥], [r̥] et [ŋ̥]) ou comme deux sons (Gimson 2001 : 192 ; Lass & Laing 2010 : 361) ?
- Deuxièmement, pourquoi /h/ ne peut-il former des agrégats qu'avec des voyelles ou /n/, /l/, /r/ et /w/ (Minkova 2003 : 340) ? Pourquoi ne s'associe-t-il pas à /m/ comme en grec ancien (Lejeune 1965 : 252) ?
- Troisièmement, si <h> du vieil anglais représente la fricative vélaire /x/ (Horobin & Smith 2002 : 54), comment connaître la nature exacte des réalisations de ce phonème ? Gimson (2001 : 192) donne trois phones: [h], [x] et [ç], auxquels on peut ajouter [ʔ] et [ɦ] (Häcker 2002 : 113 et 118), et même Ø pour les cas d'élision totale sans glottalisation de la voyelle initiale.
- Quatrièmement, comment expliquer qu'en anglais, /h/ soit la seule consonne parmi les fricatives à ne pas voir son allophone voisé se phonémiser lors de la

transition du vieil anglais au moyen anglais, alors même que, pour Hogg (2002 : 9) un allophone [ɣ] de /x/ est aussi un phonème ?

Nous soulevons deux hypothèses. Le contraste phonémique ne s'est peut-être pas réalisé par manque de productivité. En effet, on ne trouve pas de nombreux lexèmes ou syntagmes ayant une prononciation semblable, comme *a head ~ ahead*. L'alternance ne se fait pas entre /h/ et /h/, mais entre /h/ et Ø en anglais contemporain. Une autre hypothèse est qu'en moyen anglais, avec l'influence moindre des langues germaniques, les emprunts de l'anglais aux langues d'origine gréco-latines sont peu favorables à une phonémisation des fricatives glottales absentes de leur inventaire.

Minkova (2003 : 340) indique que la seule certitude concernant ces mutations est que /x/ provient de l'indo-européen par spirantisation de la plosive \*k. Le vieil anglais voit se produire la lénition progressive de /x/ en /h/, c'est-à-dire l'affaiblissement de la vélaire dû à la perte du lieu d'articulation buccal, autrement appelée débuccalisation. La lénition de /h/ se poursuit ensuite. Contextuelle et stylistique, elle semble liée à la faiblesse perceptive intrinsèque au phonème.

En résumé, /h/ est issu de la spirantisation de \*k de l'indo-européen vers /x/. Il subit une phonémisation lente lors du passage du vieil anglais au moyen anglais, mais il est le seul phonème fricatif qui n'a pas sa contrepartie voisée en anglais moderne. Il compose des agrégats initiaux de mot avec des nasales, des liquides et des approximantes en vieil anglais, pour chuter progressivement dans ces contextes lors de la transition vers le moyen anglais. En moyen anglais, alors qu'il est robuste en position initiale, il se vocalise en position finale de mot et il subit une lénition, parfois totale, à l'intervocalique.

## 1.2. Trois facteurs pouvant expliquer la labilité de /h/ en diachronie

Dans l'histoire de l'anglais, /h/ est décrit comme instable, labile, propre à l'effacement (Lutz 1994). La variabilité des omissions et des additions de <h> en diachronie pourrait être due à la faiblesse intrinsèque de /h/, à des biais méthodologiques et à des paramètres linguistiques.

### 1.2.1. Caractère « faible » de /h/

L'hypothèse historique la moins controversée est que /h/ est labile au cours du temps à cause de son caractère « faible » (Horobin & Smith 2002). Lass & Laing (2010) expliquent que /h/ est faible pour deux raisons. Premièrement, c'est la consonne la plus susceptible d'être élidée. Deuxièmement, il est rare que l'on reconstruise \*h comme une forme appartenant à une protolange.<sup>7</sup> En effet, selon les auteurs, /h/ en position initiale de syllabe est généralement une forme historiquement dérivée d'une autre.

---

<sup>7</sup> Dans l'état actuel des reconstructions, il semble que cette affirmation de Lass & Laing (2010 : 358) soit inexacte. La protoforme \*h existait au moins chez l'ancêtre des langues afroasiatiques et sémitiques, du géorgien, des langues turques, des langues austro-asiatiques, des langues hmong-mien ou des langues mayas (communication personnelle d'un relecteur anonyme).

Hogg (2002 : 62) décrit la lénition progressive de la fricative vélaire sourde [x], en contexte intervocalique en particulier. Il cite l'exemple de *sēon* (*see*), qui dériverait de *sīhan*. Autrement dit, le graphème <h>, qui correspond au phonème /x/ au début du Moyen-Âge, se réalise [x] en position initiale, position forte par définition. Ensuite, il s'affaiblit pour se réaliser [h], puis [ɦ] à l'intervocalique par propagation du voisement. Il finit par chuter entre deux voyelles qui, en position de hiatus, peuvent fusionner en une diphtongue.

Au plan physiologique, la lénition consiste dans l'affaiblissement de la constriction vélaire, avec un relâchement des organes articulatoires qui permet un bruit de friction, lequel est en réalité la turbulence de l'air qui est expiré dans le tractus vocal sans être interrompu par une constriction.

Au plan acoustique, la faiblesse du phone [h] dérive d'une part de l'absence du premier formant F1, principal responsable de l'intensité. Un formant est une zone d'harmoniques renforcés par une concentration d'énergie. Sa mesure s'exprime en Hertz. À un formant correspond la fréquence des résonances qui se produisent dans le tractus vocal et qui dépendent de sa configuration. L'absence de premier formant (F1) pour l'articulation de [h] est liée à l'ouverture de la glotte nécessaire pour la réalisation de ce son.

D'autre part, le fait que [h] ait les mêmes caractéristiques formantiques que la voyelle suivante, avec des transitions de F2 et F3 claires qui correspondent essentiellement aux mouvements des articulateurs supra-glottaux, gêne la création d'un contraste perceptif. En général, un bruit d'origine glottique peut se confondre avec le bruit ambiant et avec d'autres bruits d'origine supra-glottique (Vaissière 2001).

Le caractère « faible » du phonème fait varier ses traits et sa distribution. A la suite de Lutz (1994 : 175), McMahan (2000 : 244) propose, dans son chapitre sur /r/ anglais, de mettre en relation les consonnes /r/, /w/, /j/ et /h/. En effet, McMahan indique que ces phonèmes ont tous subi un affaiblissement positionnel et structurel dans l'histoire de l'anglais. Alors qu'en vieil anglais, ils pouvaient émerger en attaque et en coda, ils ont subi une attrition progressive en coda, en se vocalisant et en fusionnant avec les voyelles précédentes. Lutz (1994 : 175) donne l'exemple de *bohte* en vieil anglais, qui devient *boughte* en moyen anglais puis *bought* en anglais moderne.

Cependant, cette hypothèse expliquant la mutation de /x/ en /h/, et la labilité de /h/ dans les corpus, en tant qu'elle est basée sur la « faiblesse » de la consonne, est nuancée par d'autres chercheurs.

### 1.2.2. Biais méthodologiques

La lecture et la compréhension de manuscrits médiévaux sont susceptibles d'être affectées par des biais méthodologiques. Il est en effet peu probable que le texte soit une fidèle transcription de la parole, exacte et cohérente. Le graphème <h> en est un bon exemple.

Lass & Laing (2010 : 358) rappellent que l'anglais médiéval se passe d'orthographe fixe. Les manuscrits sont parsemés de bizarreries scripturales et de variantes diverses. Ils notent que le texte choisi comme corpus par un chercheur est déterminant par

rapport au calcul du nombre de <h> insérés et élidés. Ils mentionnent un manuscrit qui témoigne d'un système scriptural prolifique, avec vingt-cinq contextes différents pour les occurrences de <h>, pour conclure sur le manque d'hypothèse tangible expliquant le grand nombre de graphèmes <h> intrusifs. La seule hypothèse adéquate, selon les auteurs, reste le plaisir du scribe qui jouit de tous les possibles : *the celebration of potential variation* (Lass & Laing 2010 : 365).

Scragg (1970) suggère que les omissions et insertions non historiques de <h> dans les manuscrits médiévaux procèdent d'autres causes que la simple correspondance entre les lettres et les sons des dialectes anglo-saxons. Il note des possibilités d'erreurs de copie des scribes. Il mentionne des traces de dittographie impliquant la répétition fautive d'une lettre (*hiera hierra* au lieu de *hiera ierra* pour *their anger*) et d'haplographie, quand un graphème attendu deux fois n'est écrit qu'une seule fois (*he afað* au lieu de *he hafað* pour *he has*).

### 1.2.3. Paramètres linguistiques

Minkova (2003 : 365), dans son étude de la poésie médiévale, considère que la force ou la faiblesse des consonnes explique peu la lénition de /x/ en /h/ dans les agrégats initiaux de mot /xn-, xl-, xr-, xw-/. Ce changement phonétique progressif est lié, selon elle, à l'alternance entre /h/ et Ø. Les réalisations intermédiaires, qui gardent <h>, comme <hr-, hl-, hn->, ou qui omettent <h> en ne conservant que les sonantes [r], [l], [n] dévoisées, sont ainsi instables.<sup>8</sup>

Crisma (2007) pense que la chute aléatoire de /h/ (*H dropping*) existait déjà en moyen anglais. Pour elle, les variantes sans <h> alternent avec les formes avec <h>. La chute de /h/ dépend des contextes phonologiques. Elle conteste l'hypothèse de Milroy (1983) qui explique, dans son étude sur *H dropping*, que les variations peuvent être attribuées à des critères sociolinguistiques.

En particulier, Crisma (2007 : 71-72) met en relief le contexte propice à l'omission de <h> dans un corpus composé du *Middle English Dictionary* (MED) et du *Penn-Helsinki parsed corpus of Middle English*, 2<sup>nd</sup> édition (PPCME2). Elle analyse 10 072 occurrences de noms et d'adjectifs commençant par la lettre <h>. Le graphème <h> est plus souvent omis dans les emprunts aux langues latines (p. ex. *habit, hospital, honest, horrible*) que dans les mots à racine germanique (p. ex. *half, hound, hundred*). Les omissions, quand elles existent, sont plus fréquentes quand le mot précédent termine par une consonne plutôt que par une voyelle finale. En se fondant sur l'hypothèse

---

<sup>8</sup> “In summary, there is no necessary temporary relationship between consonantal strength and /hC-/ reduction, or sonority re-ranking and preservation of the cluster. Whenever reduction occurs, it is triggered by confusability of /h/ with Ø. The intermediate realizations of <hr-, hl-, hn->, the voiceless [ɾ], [l̥], and [ŋ̥], were unstable for typological reasons. In line with all previous accounts, cluster reduction as outlined here must assume a gradual change from a cluster to a single segment passing through phonetically intermediate stages, when traces of the earlier pronunciation surface allophonically in the realization of the new variant. Different regional and social varieties interpret the variants differently resulting in classic cases of merger and phonemicization of contrasts.” (Minkova 2003 : 365)

selon laquelle l'omission de <h> est plus probable quand le son n'est pas prononcé, l'auteure conclut que les données étudiées indiquent que le son [h] est préservé après une voyelle finale de mot, mais susceptible de chuter après une consonne finale de mot. Elle n'observe pas de corrélation entre élisions et insertions.

La question de la proportion de <h> insérés par rapport aux <h> élidés est traitée par Lass & Laing (2010 : 359). Les auteurs indiquent que dans certains manuscrits, les quantités sont quasi-identiques, tandis qu'ailleurs, on trouve davantage d'insertions que d'élisions.

Minkova (2003) s'appuie sur Scragg (1970) pour faire de l'instabilité de /h/ une preuve de l'existence de l'occlusive glottale sourde [ʔ] en vieil anglais. Selon elle, les incohérences et fluctuations graphiques retrouvées dans les sources ne peuvent pas être associées à des propriétés phonologiques de /h/. Elle note deux types d'anomalies dans son corpus de dix-sept textes médiévaux. D'une part, elle retrouve davantage d'insertions que d'omissions (soit une omission pour quatre insertions). Ces écarts n'ont pas de base étymologique, comme *herian* pour *erian* (*plough* : « labourer »), ou *hup* pour *up* (exemples tirés de Scragg 1970 : 170). Ils peuvent être attribués à un artefact scriptural pour marquer le hiatus ou encore pour différencier les mots d'anglais du latin, où /h/ est élidé en position prévocale. D'autre part, elle observe davantage d'omissions et d'insertions devant les syllabes accentuées que devant les syllabes inaccentuées. Elle conclut que la lettre <h> était utilisée par les scribes pour marquer une occlusive glottale (Minkova 2003 : 163).

Häcker (2004 : 115-116) compare l'insertion de <h> dans les accents de l'anglais contemporain avec les insertions dans les manuscrits médiévaux, afin de discuter les hypothèses d'hypercorrection et de contrainte langagière. Elle remarque que les insertions se font entre deux sons voisés. Les insertions non étymologiques en position médiane sont rares. Häcker (2004 : 121-122) fait le lien entre la parole de locuteurs du XX<sup>e</sup> siècle et les occurrences retrouvées dans les manuscrits du Moyen-Âge. Elle explique qu'aux deux époques, l'intrusion de <h> corrèle avec, d'une part, le niveau d'éducation du locuteur ou du scribe et d'autre part, le contexte phonologique. En effet, <h> émerge, non pas aléatoirement, mais dans des contextes bien définis, soit entre deux voyelles en hiatus, là où il occupe une fonction liante.

En conclusion, la labilité de /h/ en anglais médiéval peut être expliquée par la faiblesse intrinsèque de /h/, les erreurs des scribes et des paramètres linguistiques. Par exemple, /h/ s'efface plus facilement après une consonne, mais il est maintenu après une voyelle. Davantage d'insertions sont observées quand la voyelle initiale est accentuée. La lettre <h> est peut-être utilisée pour indiquer la présence de l'occlusive glottale [ʔ].

## 2. Paramètres de variation de /h/ anglais en synchronie

En anglais moderne, des variations dans la prononciation de /h/, sans être systématiques pour des locuteurs donnés, sont observées au Royaume-Uni et dans les autres territoires anglophones. Différents facteurs de variation déclenchent l'effacement de /h/ ou au contraire son insertion, aux plans extralinguistique et linguistique.

## 2.1. Variation au plan extralinguistique

La question est de savoir si l'apparition de [h] en contexte illicite est une variante de /h/ qui, conditionnée par un facteur social, émergerait d'une façon similaire aux variations de /r/ en anglais new-yorkais (Labov 1963). Ramisch (2010) propose d'analyser les occurrences selon un continuum *H dropping* – *H retaining*, avec des variantes entre les deux pôles, qui feraient intervenir les semi-voyelles /j/ et /w/. Cette perspective, quoiqu'intéressante, est pourtant marginale, tant la variation de /h/ est considérée dans la littérature comme un phénomène binaire.<sup>9</sup> En général, deux options sont possibles : /h/ est préservé, ou /h/ est éliidé.

Les travaux sur *H dropping* manquent de rendre compte, de façon homogène et cohérente, du niveau de préservation ou d'effacement de /h/ dans les dialectes anglais. Il est difficile de dessiner une ligne imaginaire, ou isoglosse, délimitant des aires géographiques dans lesquelles il existe une tendance à faire chuter /h/. Nous voyons comment ceci s'exprime premièrement dans l'opposition entre le Royaume-Uni et d'autres pays anglophones, et deuxièmement, au sein des parlers britanniques.

Premièrement, *H dropping* s'analyse dans un contraste entre le Royaume-Uni et les autres pays anglophones. Wells (1982 : 252) considère *H dropping* comme une « innovation britannique » qui serait une variante basilectale, c'est-à-dire peu « prestigieuse », par opposition à l'acrolecte. En effet, /h/ contraste avec Ø dans tous les parlers populaires de l'Angleterre et du sud du pays de Galles (Wells 1970 : 240). Il affirme que *H dropping* est inconnu en Amérique du Nord, ce qui, selon lui, permet de dater son émergence à une période postérieure à la colonisation de l'Amérique du Nord.<sup>10</sup> Il attribue ainsi la fréquence de *H dropping* en Australie au fait que l'Australie a été peuplée plus tard par les Britanniques, alors que la tendance à effacer /h/ était déjà bien ancrée dans les mœurs anglaises.

Outre-Atlantique, l'effacement de /h/ initial devant une voyelle est attesté dans les parlers de certains locuteurs des Caraïbes, sans que l'on sache s'il s'agit d'une innovation importée du Royaume-Uni ou d'une particularité locale (Wells 1982 : 256). Aux Antilles (à l'exception des Iles-sous-le-Vent), l'élosion de /h/ initial est attestée, ainsi que l'insertion de [h] ou même [w] dans les mots commençant par une voyelle, comme *egg* ou *ugly* (Aceto 2006 : 218).

---

<sup>9</sup> Pierrehumbert & Talkin (1991) évoquent un possible caractère scalaire des réalisations de /h/.

<sup>10</sup> Un relecteur anonyme indique que Milroy n'est pas d'accord avec cette hypothèse.

Tendance à élider /h/	Tendance à conserver /h/
Basilectes d'Angleterre et du Pays de Galles Australie Jamaïque, Guyane, Bahamas	Région du Yorkshire, Est de l'Angleterre, Wessex, Ecosse, Irlande Etats-Unis, La Barbade, Les Iles-sous-le-Vent

Tableau 3. Géographie de H dropping et H retaining d'après Wells (1970), Wells (1982), Gimson (2001), Aceto (2006) et Foulkes & Docherty (1999)

Bauer (2002 : 82) précise que les réalisations du /h/ initial du mot *herb* diffèrent selon la variété d'anglais parlée.

Pronunciation of the marked consonant in different varieties						
word	RP	General American	Canadian	Australian	New Zealand	South African
herb	h	∅	h ~ ∅	h	h	h

Tableau 4. Réalisation de /h/ dans six accents anglais, adapté d'après Bauer (2002 : 82)

Cette information concernant la prononciation de *herb* est en conflit avec Trask (1996 : 168) qui indique que l'élision de /h/, qui rend homophones des noms comme *art* et *heart*, est très rare en Amérique du Nord et fréquente en Angleterre.

Pour Jones, une réalisation lénifiée et voisée de /h/ est attestée chez la plupart des locuteurs sud-africains (Jones 1956 : 117). Selon Choon *et al.* (2012 : 2) l'insertion du phone intrusif [h] dans l'anglais du Nigeria (attestée dans moins de 1% des contextes possibles) est en général moins fréquente que l'élision (attestée dans presque 20 % des contextes possibles). En anglais du Nigéria, Gut (2012) observe que l'élision est un phénomène répandu parmi toutes les ethnies représentées dans son corpus. En revanche, l'insertion est rencontrée seulement chez les locuteurs d'un groupe ethnique : les Yorubas. La présence ou non de /h/ dans l'inventaire phonémique de la langue maternelle des Nigériens (Igbo, Yoruba, Hausa) ne semble pas être un facteur déterminant puisque ces langues comptent toutes /h/ parmi leurs consonnes.

Une trace de la lénition historique de /h/ se retrouve dans la réalisation des agrégats /hw/, pour lesquels /h/ n'est pas lénifié dans tous les accents. Des survivances de /h/ initial sont attestées, en particulier en anglais d'Écosse, variété dans laquelle *whine* s'oppose à *wine*. Un doute existe quant à l'analyse de l'attaque de *whine* comme un seul phonème /w/ ou bien comme deux phonèmes /h/ + /w/. Autrement dit, est-on face à une mutation consonantique de /hw/ vers /w/, résultant dans l'émergence d'un phonème supplémentaire en anglais d'Écosse, ou à une trace résiduelle d'un processus, ancien mais inchoatif, d'amuïssement de /h/ ? Wells (1982 : 228) distingue les processus que sont *H dropping* et *Glide Cluster Reduction* (la transition de /hw/ à /w/). Ils sont pour lui perçus différemment au plan social, *H dropping* étant négativement connoté, à l'inverse de *Glide Cluster Reduction*. La lénition dans les *clusters* est aussi attestée au Pays de Galles. Dans la région de Cardiff, /h/ est palatalisé, avec des réalisations de /hj/ en [ç] (Mees & Collins 1999 : 192). Il semble ainsi difficile de déterminer des isoglosses pour la tendance à lénifier, voire à faire chuter /h/.

Tendance à effacer /h/ devant /w/	Tendance à préserver /h/ dans des réalisations comme [hw]
<p>Angleterre</p> <p>Pays de Galles, Hémisphère sud, les Caraïbes</p> <p>Certains accents américains</p>	<p>Ecosse, Irlande, région de Northumberland</p> <p>Certains parlers anglais (RP)</p> <p>Sud des Etats-Unis</p> <p>Ecosse [ʌ]</p>

Tableau 5. Géographie de la lénition de /h/ dans les clusters /hw/, d'après Wells (1982 : 228)

Deuxièmement, le traitement de *H dropping* varie dans les parlers britanniques. McMahon (2002 : 65) rappelle que /h/ est éliidé dans certains accents, et peut même être totalement absent du système phonologique du *Cockney* par exemple. Trask (2003 : 105) reproduit une carte de Milroy (1992), dont les données de 1960 indiquent que *H dropping* est très répandu en Angleterre, avec un maintien de /h/ dans quelques enclaves anglaises comme le Somerset, le Norfolk, le Suffolk, l'Essex et le Nord. Le critère géographique semble peu déterminant cependant. En effet, *H dropping* reflète plutôt une opposition entre l'accent RP et l'anglais basilectal des classes populaires. Pour la variété d'anglais britannique standard, l'accent RP est, depuis Jones (1928), à la fois la référence, utilisée par les dictionnaires, et la variété de prestige, parlée par les lettrés. Pourtant, la prononciation RP est censée être celle qui est comprise par le plus grand nombre d'auditeurs : « *widely understood pronunciation* » (Jones 1956 : 4). En tant qu'acrolecte (variété de prestige) de l'anglais britannique standard, l'accent RP exclut la chute de /h/ initial, tout en autorisant, dans certains parlers précieux, d'une part, l'éliision de /h/ dans des syntagmes comme *an historical fact*, et d'autre part, le maintien de /h/ dans les agrégats /hw/ par exemple.<sup>11</sup> On peut considérer qu'aux franges de la prononciation RP, certains locuteurs font varier leur prononciation, de façon très subtile, pour se distinguer socialement des locuteurs des basilectes (ruraux et urbains).

Trudgill (2000 : 32) donne ainsi dix-neuf prononciations différentes du mot *home* dans les accents RP et autres d'Angleterre. Dans les variantes des classes populaires, /h/ tend à disparaître. Ainsi [jem] serait-il attesté pour *home* dans certains parlers de Newcastle.

Chevillet (1991) précise que la chute de /h/ initial est un marqueur social caractéristique des dialectes urbains de l'Angleterre, tandis que les franges rurales sont épargnées. Selon Chevillet (1991 : 87),

---

la chute du /h/ initial dans *hat*, *hill*, *hedge*, etc. n'est pas attestée dans les zones rurales du Norfolk et du Suffolk. Il n'en va pas de même à Norwich, où le phénomène rappelle encore une fois la prononciation cockney. La chute du /h/, que Daniel Jones qualifiait de « suicide social » est reconnue depuis longtemps comme un marqueur social infaillible, quelle que soit la variété d'anglais considérée.

---

<sup>11</sup> Voir J. Wells : <http://www.phon.ucl.ac.uk/home/wells/accentsanddialects/>, URL consultée le 10/03/2019.

Tollfree (1999 : 173) explique que, dans le dialecte urbain du sud-est de Londres, l'élision de /h/ conduit à l'émergence de mots homophones. Un allongement vocalique pourrait compenser la chute de /h/. En cas d'élision de /h/ initial, le mot peut être précédé du déterminant *an*, avec des variantes en *a* suivi de [ʔ] (p. ex. *a<sup>ʔ</sup>orse*), ou *a* suivi d'un [ɪ] intrusif plus rarement (p. ex. *a [ɪ] (h)orse*). L'hypercorrection par insertion de [h] est attestée, en particulier chez les locuteurs âgés. Dans le nord de l'Angleterre et dans le centre, le maintien et la chute de /h/ sont variables (Foulkes & Docherty 1999 : 51). Dans la région du Wirral, autour de Liverpool, /h/ chute dans plus de 70 % des cas (Newbrook 1999 : 98). En apparence, on peut associer le maintien de /h/ aux zones rurales et son élision aux zones urbaines. Il existe des exceptions cependant. On trouve des zones urbaines où /h/ est préservé, comme dans le Tyneside (Watt & Milroy 1999 : 30). En parallèle, une région rurale comme l'Est-Anglie, où *H dropping* est longtemps resté rare, a connu pourtant assez tôt un fort taux d'élision (sans son pendant qui est l'hypercorrection) dans la zone plus restreinte de Norwich, d'après Trudgill (1999 : 133). Ceci montre premièrement que *H dropping* s'installe d'abord dans les zones urbaines, puis rurales par contamination. Deuxièmement, il existe une corrélation entre *H dropping* et le style du locuteur et sa classe sociale.

Au XIX<sup>e</sup> siècle, Ellis (1869 : 461) indique que l'omission de /h/ est universelle et se retrouve même dans les classes aisées. Conservé dans les campagnes, il est éliminé à la ville (542) et dans les villages de pêcheurs par les jeunes locuteurs (777). Les élisions de /h/ seraient plus fréquentes quand le locuteur se sent nerveux (307), et les insertions plus fréquentes quand il est en colère (312). Ellis (1869 : 739) recense des insertions fortuites et accidentelles (*<sup>h</sup>owlet* au lieu de *owlet*), et des occurrences emphatiques comme *she<sup>h</sup>is* (476) ou *<sup>h</sup>us* (599). Dans une certaine mesure ces analyses anciennes trouvent leur écho dans les études plus récentes résumées ci-dessous.

Au XX<sup>e</sup> siècle, Crépin (1972 : 29, 37) indique que la variation dialectale de l'anglais est davantage sociale que géographique. Wells décrit *H dropping* comme un trait saillant de la classe ouvrière, à la suite de Hudson et Holloway (1977) et de Trudgill (1974). Il considère que le phonème /h/ est acquis par l'enfant suite à la pression sociale, et fait de l'effacement de /h/ un des plus puissants *shibboleths*<sup>12</sup> de l'anglais (254). Jones (1956 : 116) décrit l'effacement de /h/ comme une particularité de nombreux dialectes de l'anglais, en particulier celui de Londres. L'insertion de [h] serait, elle aussi, spécifique aux locuteurs non lettrés. Ces deux variantes (élision et insertion) sont vigoureusement stigmatisées. A l'inverse, certains lexèmes, comme *historical*, *hysterical*, *hotel*, seraient parfois prononcés sans [h] par des locuteurs RP, d'un certain âge, et lettrés (Gimson 2001 : 192).

Selon certaines sources, la tendance à insérer ou éliminer /h/ pourrait varier selon le sexe du locuteur. En effet, dans le dialecte de Sheffield des années 90, les locuteurs masculins plutôt jeunes et âgés ont tendance à éliminer /h/ mais les jeunes femmes le

---

<sup>12</sup> Un *shibboleth* est un mot, une prononciation ou une coutume propres à un groupe de personnes, créant une différence distinctive permettant d'identifier ce groupe par opposition à un autre.

préservent (Stoddart, Upton & Widdowson 1999 : 76). Le sexe peut donc sembler déterminant par rapport à l'âge par exemple. Cependant, au sein de la classe ouvrière, si l'âge n'est pas un facteur à Hull, il est discriminant à Milton Keynes et Reading, où les adolescents tendent à maintenir /h/, par comparaison avec les personnes âgées qui l'élide (Williams & Kerswill 1999 : 157–158). Dans les dialectes pidgins, l'effacement de /h/ est plus fréquent chez les hommes que chez les femmes, comme l'indiquent Horvath (2004 : 101) pour l'anglais d'Australie, et Gut (2012) pour l'anglais du Nigeria. En revanche, Choon *et al.* (2012 : 15) ne détectent pas d'influence significative du sexe sur les insertions de [h]. Les auteurs soulignent une possible corrélation idiosyncrasique entre insertion et effacement : 73,1 % des locuteurs qui insèrent [h] l'effacent également. Cette corrélation est unidirectionnelle, c'est-à-dire qu'il serait faux de dire que la majorité des locuteurs qui effacent /h/ l'insèrent également. Robb et Chen (2009) précisent que le sexe du locuteur n'influe pas sur la durée de [h] mais influe sur les réalisations voisées [ɦ].

Les variations dans la prononciation de /h/ sont différemment perçues, selon l'origine sociolinguistique des auditeurs. Au Royaume-Uni, l'effacement de /h/ devant une voyelle initiale est fortement stigmatisé, et ce, depuis le XVIII<sup>e</sup> siècle (Wells 1982 : 255). Il faut noter que le maintien de /h/ dans les agrégats, qu'il soit naturel ou forcé (nuance introduite par Wells), est au contraire la preuve d'une prononciation « correcte, soignée et belle » (Wells 1982 : 229). Cette tendance serait davantage marquée chez les femmes que chez les hommes. Elle serait un trait spécifique des locuteurs qui adoptent la prononciation RP, et idéale pour la pratique théâtrale ou la versification. Choon *et al.* (2012) montrent qu'au Nigeria, deux paramètres influent sur la recevabilité des insertions et élisions de /h/ : l'âge et le sexe de l'auditeur. L'effacement est plutôt bien accepté, surtout parmi les jeunes auditeurs, et chez les hommes. Les élisions sont davantage acceptées que les insertions.

En conclusion, la variation dans les réalisations de /h/ s'exprime différemment selon les paramètres extralinguistiques. Incontestablement, le maintien de /h/, ainsi que son élision et son insertion illicites, dépendent de la classe sociale du locuteur. Les paramètres de genre, d'âge et de zone géographique ont une influence moins claire dans les réalisations de /h/. La stigmatisation sociale semble moins importante pour l'effacement que pour l'insertion. La variation extralinguistique, qui concerne les caractéristiques des locuteurs (et leurs origines sociales, géographiques, leur âge et leur sexe), s'accompagne d'une variation au plan strictement linguistique.

## 2.2. Variation au plan linguistique

La langue anglaise n'utilise pas toutes les combinaisons de phonèmes possibles. Son emploi de /h/ est limité en raison de contraintes phonotactiques importantes. Les réalisations de /h/ varient en fonction du lexique, de la catégorie grammaticale, et de l'environnement phonologique et phonétique.

### 2.2.1. /h/ et les contraintes phonotactiques de l'anglais

Les contraintes phonotactiques de l'anglais sont l'ensemble des regroupements licites de consonnes et de voyelles en fonction de leur place dans le mot, dans la syllabe, ou

dans un groupement consonantique. Les contraintes phonotactiques sont fortes pour /h/ dans le cadre de l'accent RP. Le phonème apparaît à l'initiale des mots (*who, horn*) ou en milieu de mot, où la coupe syllabique se fait juste avant /h/ : *ahead, behave, perhaps, behind, spearhead, anyhow, manhood, abhor, adhere* (exemples de Gimson 2001 : 191). Il apparaît exclusivement devant une voyelle ainsi que /j/ et /w/ en RP. Un principe de l'accent RP, à l'origine, est l'absence de /h/ dans une syllabe inaccentuée. Ce principe se retrouve dans les prononciations désuètes de *historic ou hysteria*, mots dans lesquels /h/ pouvait être éliminé dans un passé récent, même s'il a été restauré de nos jours (Wells 1982 : 255). De plus, /h/ fonctionne comme un *glide* dans le sens où, comme /r/, /j/ et /w/, il ne se combine avec aucune autre consonne (hormis /j/ en RP, comme dans *Huston*, et parfois /w/).

### 2.2.2. Lexique : cas des emprunts

Les emprunts montrent bien les compromis à l'œuvre entre les contraintes phonotactiques de la langue source et celles de la langue d'arrivée, dans les processus de lexicalisation. Par exemple, dans le mot *Messiah* issu de l'hébreu, /h/ est éliminé en position finale, pour être conforme avec la contrainte phonotactique de l'anglais qui interdit /h/ en position finale. Au contraire, en anglais d'Irlande, /h/ final de mot peut être réalisé dans des emprunts au gaélique irlandais, comme par exemple avec *McGrath* [mə'grah] (Wells 1982 : 44).

Le graphème <h> est muet dans les mots empruntés au grec, qui comportent le digraphe <rh> comme dans *rhapsody*. Il ne se prononce pas dans les mots issus du latin par le français, qui ont été assimilés selon les contraintes de la langue source (*hour, honor, honest, heir*, et leurs dérivés). De même, <h> est muet dans certains mots dérivés avec un préfixe venant du latin, comme *exhaust, exhibit, exhilarate*, ou un suffixe issu du scandinave comme dans *Durham, Clapham*. En revanche, <h> correspond à la réalisation obligatoire de la fricative glottale qui réapparaît dans des emprunts assimilés au français, qu'ils soient d'origine latine (*horror, harass, hospital, host, humour*) ou d'origine germanique (*hardy, haste, herald*) (Gimson 2001 : 192).

Cependant, des usages variables existent. La variante libre [əʊ'tel] pour *hotel*, signalée par le *Cambridge Pronouncing Dictionary* (Jones 2006), n'est pas mentionnée dans le *Longman Pronunciation Dictionary* (Wells 2008). De plus, contrairement à la prononciation stable de *honor, honoris causa* se prononce avec /h/ initial, mais avec une variante sans /h/ (Wells 2008 : 385). Le nom propre *Honorius* retient /h/ (Jones 2006 : 243). Wells (1982 : 255) indique que les prononciations particulières de *Birmingham* ou *Nottingham* avec [h] sont une tentative des locuteurs des classes moyennes pour se distinguer socialement en se rapprochant d'une variété ressentie comme prestigieuse.

La chute potentielle de /h/ s'explique premièrement par l'étymologie du mot, issu du français, et deuxièmement par la contrainte phonotactique qui fait chuter /h/ dans toute syllabe inaccentuée. Le maintien ou la résurgence de /h/ s'explique par le respect des contraintes phonotactiques de l'anglais, probablement renforcées par le rôle de l'orthographe.

### 2.2.3. Morphosyntaxe : rôle de la catégorie grammaticale

La catégorie grammaticale d'un mot et sa position dans la phrase déterminent la réalisation de /h/. En effet, dans les accents standards, /h/ dans les auxiliaires *has, have, had* n'est pas réalisé si les mots ne sont ni accentués ni précédés d'une pause. Il en va de même pour les pronoms et les déterminants *he, him, her, his* (Wells 1982 : 254).

	Forme forte	Forme faible
Déterminant <i>his</i>	[hɪz] en début de phrase	[ɪz]
Pronom sujet <i>he</i>	[hi] en tête de phrase [hi:] pour le contraste et l'emphase	[i] p. ex. "I'm fine," <i>he</i> said.
Déterminant <i>her</i>	[hə] en tête de phrase	[ə] : ailleurs
Pronom objet <i>him</i>	--	[ɪm]
Auxiliaires <i>have, has, had</i>	en position finale : [hæv], [hæz], [hæd] p. ex. <i>Yes, she has.</i>	i) en position initiale [həv], [həz], [həd] p. ex. <i>Have you seen John ?</i> ii) ailleurs dans la phrase : [əv], [əz], [əd]

Tableau 6. Formes faibles et fortes des pronoms, des déterminants, et des auxiliaires commençant par la lettre <h> (Roach 2009 : 91-94).

Pour Jones (1956 : 116), cependant, l'alternance entre forme pleine et forme lénifiée est aléatoire. Les paramètres grammaticaux peuvent être secondaires par rapport aux contraintes stylistiques. Wells (1982 : 255) écrit qu'il a l'impression que certains locuteurs des classes moyennes, pour éviter de faire quelque chose d'aussi mauvais goût que de faire chuter un /h/, ont une tendance affirmée à affecter un [h] aux pronoms et auxiliaires inaccentués, comme dans « ['tel<sup>h</sup>ɪm] » par exemple.

Gimson (2001 : 252-254) confirme la labilité de /h/ en contexte inaccentué. C'est ainsi que dans le domaine du traitement automatique des données, Auran & Bouzon (2003) sont prudents pour énoncer les principes qui régissent les règles d'élision dans l'alignement du corpus AIX-MARSEC.

---

Principe n° 2 : élision de [h] dans les formes **he, he'd he'll, he's, his, him, et her**

En parole continue, la fricative [h] dans les pronoms et / ou contractions énumérés ci-dessus est souvent élidée ; toutefois, cette consonne est supprimée dans la transcription à condition qu'aucune marque prosodique ne précède le mot en question, dans ce cas, on imagine que le pronom fortement accentué sera réalisé avec sa forme pleine, sans élision du /h/.

---

### 2.2.4. Facteurs morphophonologiques et phonétiques

La réalisation de /h/ est conditionnée par sa position dans le mot et dans la syllabe, et par son environnement phonétique. L'orthographe joue un rôle important. /h/ est

réalisé [h] à l'initiale absolue des mots anglais, devant une voyelle. Il est réalisé [ɦ] en position intervocalique, à l'initiale d'une syllabe accentuée (p. ex. *ahead*) (Koenig 2000 : 1223). Il existe des exceptions. Dans *at home*, /h/ peut être éliidé bien qu'il soit dans une syllabe sous l'accent (Jones 1956 : 115).

Une règle d'effacement de /h/ en syllabe inaccentuée (*Unstressed H Dropping*) est optionnelle selon Wells (1982 : 67). Que /h/ soit totalement éliidé ou bien que la réalisation soit la fricative voisée [ɦ], on peut affirmer que /h/ est lénifié en position faible, c'est-à-dire dans une syllabe inaccentuée.

Dans les parlars créoles ou pidgins, l'effacement de /h/ peut être lié à l'environnement consonantique ou vocalique. Par exemple, dans l'anglais des Bahamas, Childs & Wolfram (2004) expliquent que l'élision est plus fréquente après une consonne ou une pause qu'après une voyelle.

En revanche, en anglais du Nigeria, Gut (2012 : 12) observe moins d'élisions de /h/ quand le mot est précédé d'une pause dans son corpus de parole spontanée. Choon *et al.* (2013 : 15) ne trouvent pas d'effet du contexte phonétique précédent sur l'insertion de [h]. Gut (2012) montre que les insertions sont presque exclusivement observées dans la parole lue, par opposition avec la parole spontanée. En revanche, les élisions sont plus fréquentes dans la parole spontanée que dans les informations télévisées et dans la parole lue ou préparée. On trouve des insertions plus souvent à l'initiale d'un mot commençant par la lettre <h> qu'à l'initiale d'un mot commençant par une voyelle. Quant aux élisions, elles sont plus fréquentes quand le mot commence par le digraphe <wh> (26,7%) que quand le mot commence par <h> (17,7%). L'auteure ne donne pas d'indication précise sur la réalisation des mots dans le premier cas : sont-ils articulés avec une semi-voyelle ou non ? Elle infère que l'orthographe joue un rôle dans les élisions et intrusions de /h/. Les observations rapportées par Gut (2012) et Choon *et al.* (2012) doivent être considérées avec précaution parce que les auteurs observent une interaction entre trois paramètres (le sexe, le style de parole et le contexte phonétique), de sorte que les généralisations sont difficiles.

La réalisation du phonème /h/ de l'anglais montre une variabilité importante en synchronie. Des invariants peuvent cependant être identifiés. Des correspondances existent entre sa labilité et l'étymologie des mots, le style de parole, ainsi que l'origine linguistique et sociale du locuteur. Si l'insertion semble plus propice en contexte de hiatus, son élision est fréquente en position faible ou inaccentuée et à l'intervocalique. Les élisions et les aspirations illicites sont souvent stigmatisées. Des études sur de grands corpus oraux (p. ex. IPCE-IPAC)<sup>13</sup> sont en cours pour confirmer ces facteurs de variation. Une difficulté est que /h/ est un mauvais candidat pour des analyses semi-automatiques, en raison de ses caractéristiques acoustiques. Il s'analyse mieux grâce à des analyses manuelles fines, donc coûteuses en temps.

---

<sup>13</sup> <https://www.pacprogramme.net/IPCE-IPAC?lang=en>; URL consultée le 01/09/2019.

## Conclusion

Considéré dans la littérature comme le résultat de la lénition de la fricative vélaire /x/ du moyen anglais, le phonème /h/ apparaît comme intrinsèquement labile en anglais. Sa distribution devient lacunaire au fil des siècles. En vieil anglais le graphème <h> correspond au phonème /x/ qui, en fin de période, se réalise par son allophone [h] en position initiale, par [ɣ] en position médiane et par [x] en position finale. Le son [h], qui n'est pas encore un phonème en vieil anglais, se phonémise en moyen anglais. Robuste en position initiale, il subit une vocalisation complète en position finale et une lénition partielle, sous la forme du voisement, en position intervocalique sous l'accent. En syllabe inaccentuée, la lénition peut être complète ou partielle selon le lexème et le contexte grammatical et phonétique. L'existence de /h/ dans les agrégats initiaux du vieil anglais, dont on ignore s'ils doivent être décomposés en deux sons ou en un seul, ainsi que son alternance remarquable mais aléatoire avec Ø, posent la question de la pérennité et du statut du phonème /h/. Au plan linguistique, les variations dans les réalisations apparaissent essentiellement en creux, par sa disparition dans des contextes obligatoires. Des insertions illicites sont rapportées, qu'elles soient non étymologiques dans des textes anciens, ou idiosyncrasiques dans le parler de certains locuteurs. *H dropping* semble être un processus en cours d'achèvement, qui a commencé à l'époque du vieil anglais, avec la réduction des agrégats initiaux, dans lesquels /h/ se combinait à des sonantes. Des traces de ce processus inachevé se retrouvent dans des dialectes de l'anglais (comme l'écossais) où [h] est maintenu devant [w] dans des parlers perçus comme prestigieux. À l'inverse, la chute de /h/ en contexte initial devant une voyelle est encore aujourd'hui stigmatisée au plan social. Au plan extralinguistique, le jugement normatif et prescriptif est plus fort dans les basilectes urbains et ruraux qu'aux franges de l'ancien empire anglo-saxon, où l'anglais se créolise et où les variantes sont mieux acceptées, surtout par les jeunes locuteurs.

En décrivant l'évolution de /h/ en synchronie, on scrute un petit segment de la réalité linguistique, une sorte de fractale de l'ensemble de ses caractéristiques en diachronie. On retrouve, à la fois en synchronie et en diachronie, le rôle de /h/ comme marqueur de hiatus ou de coupe syllabique (Scragg 1970 : 180 et 186), sa confusion avec /ʔ/, sa robustesse à l'initiale, sa variabilité idiosyncrasique et sa labilité qui s'oppose à une forte stigmatisation de son mauvais usage.

Il semblerait que le phonème /h/ de l'anglais soit voué à disparaître (peut-être à l'instar du <h> disjonctif français). Des usages résistent cependant, et les réalisations de /hw/ chez certains natifs aujourd'hui rappellent l'existence des digraphes médiévaux. La lénition graduelle de /h/ dans les agrégats évoque l'idée selon laquelle les réalisations de /h/ sont graduelles selon le contexte accentuel (Faure 1975 : 71). En parallèle, Pierrehumbert & Talkin (1991) s'interrogent sur le caractère binaire ou graduel des réalisations phonétiques de /h/ en position inaccentuée.

Les similitudes dans l'évolution de /h/, entre synchronie et diachronie, trouvent leur écho dans une autre opposition : celle de l'anglais langue maternelle à l'anglais langue seconde. En effet, dans l'anglais langue seconde des francophones, des intrusions phonétiques de [h], souvent en contexte de hiatus externe, peuvent créer des

confusions phonologiques, comme dans *the <sup>h</sup>art*, qui s’oppose à *the heart*, ou phonétiques, comme dans *the <sup>h</sup>animal* (Exare : 2017). Ces aspirations intrusives, qui ne sont pas systématiques pour un locuteur donné, se comprennent mieux en observant le statut de la consonne dans l’histoire de l’indo-européen et de l’anglais.

## Références bibliographiques

- Aceto, Michel. 2006. Caribbean Englishes. In Braj. B. Kachru, Yamuna Kachru, Cecil L. Nelson. Malden (dir.), *The Handbook of World Englishes*. 203-222. MA: Blackwell Publishing.
- Auran, Cyril & Caroline Bouzon. 2003. Phonotactique prédictive et alignement automatique: Application au corpus MARSEC et perspectives. *TIPA. Travaux interdisciplinaires du laboratoire Parole et Langage* 22.
- Bacquet, Paul. 1974. *Le vocabulaire anglais*. Paris : Presses universitaires de France.
- Bauer, Laurie. 2002. *An Introduction to International Varieties of English*. Edinburg : Edinburgh University Press.
- Chevillet, François. 1991. *Les Variétés de l’anglais*. Paris : Nathan.
- Childs, Becky & Walt Wolfram. 2004. Bahamian English: Phonology. In Edgar Schneider, Kate Burridge, Berndt Kormann, Rajend Mesthrie & Clive Upton (dir.). *A Handbook of Varieties of English*, 643-656. Berlin et New York : Walter de Gruyter.
- Choon, Anja, Robert Fuchs, Ulrike Gut, Presley Ifukor & Taiwo Soneye. 2012. /H/-deletion and /h/-insertion in Nigerian English. *International Computer Archive of Modern and Medieval English (ICAME 33)*, Leuven, 30 mai – 3 juin 2012.
- Crépin, André. 1972. *Histoire de la Langue Anglaise*. Paris : Presses Universitaires de France.
- Crisma, Paola. 2007. Were they *dropping their haitches* ? A quantitative study of h-loss in Middle English. *English Language and Linguistics* 11 (1). 51-80.
- Crisma, Paola. 2009. Word-initial *h-* in Middle and Early Modern English. In Donka Minkova (ed.) *Phonological Weakness in English*. 130-167. New York : Palgrave Macmillan.
- Ellis, Alexander John. 1869. *On early English pronunciation: with especial reference to Shakspeare and Chaucer*. London : Asher & Co. <https://archive.org/details/onearlyenglishpr00elliuft> (29 août 2019).
- Exare, Christelle. 2017. *Les aspirations intrusives dans l’anglais des apprenants francophones*. Paris : Université Paris 3 Sorbonne Nouvelle. doi:<https://hal-univ-paris3.archives-ouvertes.fr/tel-015111510/document> (15 mars 2018).
- Faure, Georges. 1975. *Manuel pratique d’anglais parlé*. Paris : Hachette.

- Foulkes, Paul & Gerard Docherty (dir.). 1999. *Urban Voices: Accent Studies in the British Isles*. London : Routledge.
- Gimson, Alfred C. (1917-1985). 6<sup>e</sup> éd 2001. *Gimson's Pronunciation of English*. Alan Cruttenden (éd.) London : Arnold.
- Gut, Ulrike. 2012. Towards a codification of Nigerian English - The ICE Nigeria Project. *Journal of the Nigeria English Studies Association (JNESA)* 15 (1). 1-17.
- Häcker, Martina. 2004. Intrusive [h] in present-day English accent and <h>-insertion in medieval manuscripts. Hypercorrection or functionally-motivated language use? *New perspectives on English Historical Linguistics II*, 109-123. Amsterdam : John Benjamins.
- Hogg, Richard. 2002. *An Introduction to Old English*. Edinburgh : Edinburgh University Press.
- Horobin, Simon & Jeremy Smith. 2002. *An Introduction to Middle English*. Edinburgh : Edinburgh University Press.
- Horvath, B. M. (2008). Australian English: Phonology. *Varieties of English*, 3, 89-110.
- Hudson, Richard Anthony & A. F. Holloway. 1977. *Variation in London English*. London : University College London.
- Jones, Daniel. 1928. *An English Pronouncing Dictionary*. London : J.M. Dent and sons.
- Jones, Daniel. 17<sup>e</sup> éd. 2006. *Cambridge English Pronouncing Dictionary*. 17th edn. Cambridge : Cambridge University Press.
- Koenig, Laura L. 2000. Laryngeal Factors in Voiceless Consonant Production in Men, Women, and 5-Year-Olds. *Journal of speech, language, and hearing research : JSLHR* 43(5). 1211-28.
- Labov, William. 1963. The social motivation of a sound change. *Word* 19. 273-303.
- Larrea, Paul. & Jean-Philippe Watbled. 2004. *Linguistique générale et langue anglaise*. Paris : Armand Colin.
- Lass, Roger & Margaret Laing. 2010. In celebration of early middle English “H.” *Neuphilologische Mitteilungen* 111 (3). 345-354.
- Lejeune, Michel. 1965. *Traité de phonétique grecque*. Paris : Klincksieck.
- Lutz, Angelika. 1994. Vocalisation of “post-vocalic r.” An Early Modern English sound change? *Studies in Early modern English*, 167-185. Berlin : Mouton de Gruyter.
- McMahon, April M. 2000. *Lexical phonology and the history of English*. Cambridge : Cambridge University Press.
- McMahon, April M. 2002. *An Introduction to English Phonology*. Edinburgh : Edinburgh University Press.

- Mees, Inger M. & Beverley Collins. 1999. Cardiff: a real-time study of glottalization. In Paul Foulkes & Gerard Docherty (dir). *Urban Voices: Accent Studies in the British Isles*, 185-202. London : Routledge.
- Millward, C. M. & Mary Hayes. 2011. *A Biography of the English Language*. Wadsworth, Boston : Cengage Learning.
- Milroy, James. 2009. Variability, Language Change, and the History of English. *International Journal of English Studies* 5(1). 1-11.
- Minkova, Donka. 2003. *Alliteration and Sound Change in Early English*. Cambridge : Cambridge University Press.
- Newbrook, Mark. 1999. West Wirral: norms, self reports and usage. In Paul Foulkes & Gerard Docherty (dir). *Urban Voices: Accent Studies in the British Isles*, 90-106. Routledge.
- Pierrehumbert, Janet & David Talkin. 1991. Lenition of /h/ and glottal stop. *Papers in Laboratory Phonology II*, 90-117. Cambridge : Cambridge University Press.
- Ramisch, Heinrich. 2010. Analysing Linguistic Atlas data: the (Socio)-linguistic Context of H-Dropping. *Dialectologia Special Issue* (1). 175-184.
- Roach, Peter. 2009. *English Phonetics and Phonology. A Practical Course*. Cambridge : Cambridge University Press.
- Robb, Michael P. & Yang Chen. 2009. Is /h/ phonetically neutral? *Clinical Linguistics & Phonetics* 23 (11). 842-855.
- Scragg, Donald G. 1970. Initial H in Old English. *Anglia-Zeitschrift für englische Philologie* (88). 165-196.
- Stoddart, Jana, Clive Upton & J.D.A Widdowson. 1999. Sheffield dialect in the 1990s: revisiting the concept of NORMs. In Paul Foulkes & Gerard Docherty (dir.), *Urban Voices: Accent Studies in the British Isles*, 72-89. London : Routledge.
- Tollfree, Laura. 1999. South East London English: discrete *versus* continuous modelling. In Paul Foulkes & Gerard Docherty (dir.). *Urban Voices: Accent Studies in the British Isles*, 163-184. London : Routledge.
- Trask, Robert Lawrence. 1996. *A Dictionary of Phonetics and Phonology*. London : Routledge.
- Trask, Robert Lawrence. 2003. *Language: The Basics*. London : Routledge
- Trudgill, Peter. 1974. *The Social Differentiation of English in Norwich*. Cambridge : Cambridge University Press.
- Trudgill, Peter. 1999. Norwich: endogenous and exogenous linguistic change. In Paul Foulkes & Gerard Docherty (dir.), *Urban Voices: Accent Studies in the British Isles*, 124–140. London : Routledge.
- Trudgill, Peter. 2000. *Sociolinguistics*. London : Penguin.
- Vaissière, Jacqueline. 2001. Changements de sons et changements prosodiques: du latin au français. *Parole* (15/16). 53-88.

- Watt, Dominic & Lesley Milroy. 1999. Patterns of variation and change in three Newcastle vowels: is this dialect levelling? In Paul Foulkes & Gerard Docherty (dir.). *Urban Voices: Accent Studies in the British Isles*, 25-46. London : Routledge.
- Williams, Ann & Paul Kerswill. 1999. Dialect levelling: change and continuity in Milton Keynes, Reading and Hull. In Paul Foulkes & Gerard Docherty (dir.). *Urban Voices: Accent Studies in the British Isles*, 141-162. London : Routledge.
- Wells, John. 1982. *Accents of English 1. An Introduction*. Cambridge : Cambridge University Press.
- Wells, John. 1970. Local accents in England and Wales. *Journal of Linguistics* 6 (02). 231-252.
- Wells, John. 2008. *Longman Pronunciation Dictionary*. Harlow : Pearson Longman.

## Des mots pour dire l'insulte (de la naissance du français à nos jours)

Laurie Raymond  
Université Toulouse Jean Jaurès  
Laboratoire CLLE-ERSS (UMR 5263)  
[laurie.raymond@laposte.net](mailto:laurie.raymond@laposte.net)

### Résumé

Cet article<sup>1</sup> mène une réflexion autour du réseau hypéronymique relatif au vocable *insulte*, de la naissance du français au XII<sup>e</sup> s. à nos jours. Le but est d'aboutir à un schéma d'utilisation des termes précis pour se démarquer des définitions parfois floues et redondantes des dictionnaires. La plupart des lexicographes (Alain Rey *GRLF*, Paul Imbs *TLF*, entre autres) nous renvoient vers des mots qui créent une boucle synonymique autour de termes orduriers désignant des termes orduriers sans rendre compte de la singularité sémantique de chacun des termes définissant un acte de langage injurieux. Notre étude – centrée sur le Moyen-Âge (XII<sup>e</sup>-XV<sup>e</sup> s.) – rend compte du sémantisme des mots *insulte*, *injure*, *gros mot*, *juron* et *blasphème* dans une perspective diachronique, jusqu'à nos jours.

**Mots-clés :** insulte ; injure ; gros mot ; lexicographie ; diachronie ; synonyme ; hypéronyme

### Abstract

This paper deals with the hypernymic system of insults from the apparition of French language (12<sup>th</sup> century) onward. It aims at distinguishing the use of each word from the sometimes vague and redundant definitions one can find in dictionaries. While most lexicographers tend to redirect the reader to synonyms, thus simply creating a synonymic loop of terms defining acts of abusive language (without taking into account their respective semantic diversity), this study – focused on the Middle Ages (12<sup>th</sup> -15<sup>th</sup> centuries) – examines the semantics of the following words: *insult* (*insulte*), *offense* (*injure*), *coarse word* (*gros mot*) *swearword* (*juron*), and *blasphemy* (*blasphème*). The point is to extend our observations, in a diachronic perspective, to nowadays.

**Keywords:** insult; offense; coarse word; lexicography; diachrony; synonym; hypernym

Notre recherche porte sur la réception et la perception de l'insulte au Moyen-Âge<sup>2</sup>. À partir du postulat qu'« une prise en compte plus affinée du sens conduit à des avancées en matière d'analyse grammaticale », et « que le sens, inaccessible à l'observation, ne peut être construit que par hypothèse à partir des indices formels – eux observables – fournis par la langue » (Vaguer 2017), nous souhaitons délimiter précisément délimiter

---

<sup>1</sup> Nous remercions Céline Vaguer pour ses précieux conseils tant éditoriaux que scientifiques, ainsi que Thomas Verjans pour l'accompagnement dont il a fait preuve tout au long de la rédaction de cet article.

<sup>2</sup> Raymond, Laurie, 2019. *L'insulte : étude et enjeux du phénomène au Moyen-Âge*, mémoire de Master, codir. Céline Vaguer et Thomas Verjans, Université Toulouse Jean-Jaurès.

ce que l'on entend par le mot *insulte* de façon à pouvoir l'identifier dans un corpus donné. Définir l'insulte n'est pas chose aisée lorsque l'on sait que, bien souvent, les ouvrages lexicographiques<sup>3</sup> tout comme les locuteurs, opèrent une boucle synonymique autour de cette notion, en renvoyant tour à tour à des termes appartenant au même champ sémantique, à savoir : *injure*, *gros mot*, *juron* et *blasphème*. Nous avons choisi de travailler sur un corpus de données fournies par les ouvrages encyclopédiques car ce sont eux qui sont à l'origine des définitions des mots. Or, la lecture de ces ouvrages nous a permis de nous rendre compte que ces derniers – lus indépendamment les uns des autres – étaient d'une précision assez incertaine. Nous avons ainsi choisi de travailler sur ce corpus de textes (cf. *supra* note 3) afin d'en confronter les différentes données ; le but est de les mettre en relation de complémentarité pour parvenir à une macro-définition satisfaisante d'un point de vue sémantico-synchronique et sémantico-diachronique.

Dans cet article, nous souhaitons démontrer que ces termes, considérés à tort comme de purs synonymes, ne le sont pas en démêlant les liens subtils de leur rapprochement sémantique, et ce, dans une perspective diachronique du latin à nos jours. Ce n'est que par ce travail de sémantique lexicale approfondie que nous serons en mesure de saisir pleinement la notion d'*insulte*, et que nous pourrions alors proposer la définition qui la caractérise pour la période spécifique du Moyen-Âge.

Pour ce faire, nous confronterons les données lexicographiques de cinq termes (*insulte*, *injure*, *gros mot*, *blasphème*, *juron*). Le choix de ces lexèmes s'explique par le fait que ce sont ceux que nous avons le plus fréquemment rencontrés lors de nos lectures, et dont les sens nous semblaient proches, voire parfois confus. Dans un souci d'accessibilité et de représentativité de la langue (particulièrement pour la langue médiévale), les occurrences qui servent d'appui à nos recherches sont issues de la base textuelle Frantext ; notre corpus d'étude se compose de l'ensemble des textes attestés pour la période <Ancien français (1000 à 1300)> (composé de 80 textes<sup>4</sup>) et <Moyen français (1300-1499)> (composé de 339 textes). Nous avons ainsi soumis les lemmes sous forme des requêtes suivantes : <injure>, <insulte>, <gros mot>, <juron> et <blasphème>. Par ailleurs, afin d'assurer l'homogénéité des données, nous avons choisi de poursuivre la récolte des données sur Frantext pour les périodes suivantes, autrement dit du XVI<sup>e</sup> s. à nos jours.

Le but de notre étude consiste en la différenciation sémantique de ces cinq termes au regard de l'évolution de leur acception. Cela nous amène, ensuite, à redéfinir les intersections sémantiques de chacun des lexèmes et à proposer une vue schématique de la notion d'*insulte* qui permet de pallier les imprécisions des dictionnaires et de

---

<sup>3</sup> Cette étude s'est focalisée principalement sur des dictionnaires et a été amendée par des lexiques spécialisés et des encyclopédies (cf. références bibliographiques) ; il s'agit d'ouvrages soit descriptifs (le Greimas, le Godefroy, le Gorog, le DHLF, le TLF), soit normatifs (le GLLF, le GRLF, le DAALF, le Furetière, le DEHLF, le DELF).

<sup>4</sup> Il s'agit d'une donnée antérieure à la mise à jour de la base textuelle Frantext, donc antérieure à 2018. Aujourd'hui, Frantext compte 58 textes dans le corpus déterminé sous le nom <Ancien français>, période qui regroupe les textes antérieurs à 1300.

pouvoir saisir pleinement les variations terminologiques et sémantiques de l’insulte à travers les siècles pour une étude approfondie des textes littéraires.

## 1. Démêler l’injure et l’insulte

### 1.1 L’injure

Le terme *injure* est un substantif féminin dont les dictionnaires<sup>5</sup> recensent la première attestation en ancien français en 1155. Or, nous avons pu identifier une attestation antérieure<sup>6</sup>, non mentionnée par les ouvrages lexicographiques, dès 1150, illustré dans l’exemple (1) :

- (1) *Jamés nule creature n’ose penser si grant injure.* (Anonyme, *Le Roman de Thèbes*, 1150, p. 110)  
*Jamais nulle créature n’eut l’audace de s’imaginer une si grande injure.*

Étymologiquement, *injure* est un emprunt, issu du latin classique, de *injuria* formé par dérivation de *jus*, *juris* ‘droit, justice’ et par ajout du préfixe à valeur négative *in-* impliquant alors une ‘violation du droit, de la justice’. Le terme a donné lieu au sens global et synthétique de ‘tort, injustice’ [1<sup>re</sup> acception]. Dans les premiers textes en ancien français, *injure* apparaît également sous les formes *injurie*, *enjurie*<sup>7</sup> (2) et *injuire*<sup>8</sup> (3) :

- (2) *Ço fu par Diocletian, / Qui enveia Maximian, / Par cruelté e par **enjurie**, / Pur tuz les crestiens destruire.* (Wace, *Le Roman de Brut*, 1155, v. 5579-5582)  
*Ce fut par Dioclétien, / Qui envoya Maximien, / Par cruauté et par **injustice**, / Pour détruire tous les chrétiens.*
- (3) *Dunc pristrent la terre a destruire ; / Deus, quel dolur e quel **injuire**.* (Wace, *Le Roman de Brut*, 1155, v. 13473-13474)  
*Ils se mirent donc à détruire la terre ; / Dieu, quelle douleur et quel **tort** ce fut !*

<sup>5</sup> s.v. INJURE dans le GRLF, le DHLF, le DAALF et le DEHLF.

<sup>6</sup> Attestation trouvée grâce à une recherche effectuée sur la base textuelle *Frantext* à la requête <injure> sur l’ensemble des textes attestés pour la période <Ancien français (1000-1330)> (80 textes ; 9 occurrences d’*injure* ; requête effectuée le 18/06/2018).

<sup>7</sup> s.v. INJURE dans le GRLF, le DHLF, le DAALF, le DHLF, le Godefroy et le TLF. Notons que le préfixe *en-* est alors en usage car *in-* était encore étranger à la langue de l’époque. Par ailleurs, nous n’avons trouvé aucune attestation d’*injurie* dans *Frantext* sur la période considérée <Ancien français (1000-1330)>.

<sup>8</sup> *Injuire* serait l’aboutissement phonétique attendu : les deux dernières syllabes d’*injurie* (*-urie*) auraient subi une métathèse (Molinu 1999) en *-uire*, comme la plupart des vocables latins dont la syllabe finale se formait sur le schéma *-rius* (*injurius* > *injurie* > \**injuire*). Selon le DHLF, *injuire* serait une réfection des formes archaïques latines employées à partir de 1155, chez WACE, avec un sens repris du latin de ‘violation du droit, d’injustice’. Nous avons pu relever une seule attestation d’*injuire* dans *Frantext* pour la période considérée <Ancien français (1000-1330)>.

Sur la période médiévale, *injure* connaît deux autres acceptions, peu usitées à l'époque mais encore en usage aujourd'hui : celle de 'tort immérité' et celle d' 'offense grave'. Dans la première, *injure* désigne un 'tort immérité et injuste que l'on fait subir à quelqu'un' [2<sup>e</sup> acception] et que l'on peut rapprocher de celle de *préjudice*. Elle se retrouve dans des expressions comme *faire injure à quelqu'un*, *faire l'injure de quelque chose à quelqu'un* (4) :

- (4) Sire Dieu, je suis en divers penssement de mon nepveu Roland, qui a voulu **faire injure a ma personne**, au quel j'avoye plus de fiance que en homme vivant. (Bagnyon Jean, *L'Histoire de Charlemagne*, 1465-1470, p. 35)  
Seigneur Dieu, je suis en pleine méditation au sujet de mon neveu Roland, qui a voulu **faire injure à ma personne**, en lequel j'avais placé plus de confiance qu'en n'importe quel autre homme vivant.

Dans la seconde acception, *injure* désigne une 'offense grave et délibérée commise par un individu à l'égard d'un tiers' et qui blesse [3<sup>e</sup> acception] et que l'on peut rapprocher de celle d'*outrage*. Cette offense peut se traduire par un affront (parole ou procédé écrit), voire un geste ou un coup et est utilisée pour *venger une injustice*<sup>9</sup> :

- (5) Sachié que **m'avez fait injure**, / Car mon clerc ert quant il vivoit / Et devotement me servoit. (La Marchand Jean, *Miracle de Notre-Dame de Chartres*, 1262)  
Sachez que vous **m'avez fait injure**, car il était mon clerc quand il vivait, et me servait avec dévotion.

Ce n'est qu'à partir du XVI<sup>e</sup> siècle que *injure* se spécialise et prend sa valeur courante et populaire encore usitée aujourd'hui de 'parole outrageante'<sup>10</sup> [4<sup>e</sup> acception]. Au singulier, *injure* réfère à « une parole qu'on dit pour offenser quelqu'un, en lui reprochant quelque défaut, ou quelque vice vrai ou faux »<sup>11</sup> Au pluriel, *injures* désigne ces « paroles offensantes », ces « interjections grossières », et ces « termes de mépris<sup>12</sup> ». *Injure* et *parole* sont complémentaires, puisque *injure* est assimilé à une parole offensante tels le *blasphème*, l'*insulte*, l'*invective*, voire la *mesparole* ou la *vilipendation*<sup>13</sup> ; il s'agit d'autant de termes qui renvoient à l'attaque, l'insolence ou la vilénie, et non plus seulement à un acte offensant (au sens physique), mais véritablement à un acte de parole (au sens pragmatique). *Injure* sert alors à caractériser les paroles injurieuses adressées à autrui (6) :

<sup>9</sup> s.v. INJURE dans le DAALF.

<sup>10</sup> s.v. INJURE dans le DHLF. Cet emploi de *injure*, qui semble émerger au XVI<sup>e</sup> siècle, existait déjà en latin chrétien (latin ecclésiastique usité par les clercs et les moines dans les livres religieux), en usage restreint, pour désigner une « parole blessante ». Autrement dit, le sens actuel de *injure* n'est qu'une réminiscence de l'une de ses acceptions en latin.

<sup>11</sup> s.v. INJURE dans le Furetière et le DAALF.

<sup>12</sup> s.v. INJURE dans le TLF.

<sup>13</sup> Termes plus archaïques que l'on retrouve dans le Gorog (s.v. INJURE).

- (6) Jehan de Saint Disier appella ledit Jordanet **filz de Champisse et maillet, traistre** et plusieurs autres paroles injurieuses. (Du Cange, s.v. CAMPENSES, 1394, reg. 146, chap. 117)  
*Jehan de Saint Disier appela le dit Jordanet ‘bâtard trouvé dans les champs, révolté traître’, et plusieurs autres propos injurieux.*

Une autre acception de *injure*, également en usage à partir du XVI<sup>e</sup> siècle, réfère au sens spécifique (d’origine latine) de ‘dommages’ [5<sup>e</sup> acception] « causés par le temps, les évènements, le destin ou les intempéries »<sup>14</sup>.

- (7) C’est l’**injure** du temps et les gens mal zelés / Qui vous font prosperer sous un Roy fait de Paille. (L’Estoile Pierre de, *Registre-journal du regne de Henri III*, 1576-1578, p. 185)

À l’issue d’un bref examen, il apparaît que, si le terme *injure* apparaît dès le XII<sup>e</sup> siècle sous plusieurs graphies (*injure, enjurie, injurie, injuire*), empruntant au latin la désignation d’une ‘violation du droit, de la justice’ [1], il se décline en deux sens nouveaux sur la période médiévale : un ‘tort immérité que l’on fait subir à quelqu’un’ [2] et une ‘offense grave et délibérée causée à autrui’ [3]. Au XVI<sup>e</sup> siècle, *injure* prend sa valeur, encore actuelle, de ‘parole outrageante’ visant à marquer au fer rouge un individu à qui l’on reproche un défaut [4] et connaît une acception plus marginale de ‘dommage causé par le temps’, entre autres [5].

## 1.2 L’insulte

L’apparition du terme *insulte* en tant que substantif féminin n’est attesté qu’en 1664<sup>15</sup>. Son origine reste floue : il serait emprunté au supin *insilire*, au bas latin *insultus* (DAALF), au latin classique ou médiéval<sup>16</sup> *insultum* (DHLF), duquel serait issu le verbe *insulter* dont *insulte* serait le « déverbal » (GLLF). Le mot *insulte* n’apparaît dans les textes qu’à partir de 1370 sous les formes *insultation* et *insult*. *Insultation* (substantif féminin) désignait alors ‘une attaque’ (8) mais également une ‘parole ou attitude outrageante’ (9) – encore en usage aujourd’hui :

<sup>14</sup> s.v. INJURE dans le DEHLF.

<sup>15</sup> s.v. INSULTE dans le DHLF.

<sup>16</sup> Latin médiéval identifié comme tel avant 1125.

- (8) Icellui Vigier se lança au dit Robin, lequel, doutant la puissance et **insultation** du dit Vigier, le feri un seul coup. (*Arch. JJ* 102, pièce 83, 1370)  
Vigier lança ainsi au dit Robin, lequel doutant de la puissance et **de la capacité d’attaque** du dit Vigier, le frappa d’un seul coup.
- (9) Toutesfois resistoit il tres grandement aux insolences et **insultations** de ceulx de Grimberghe. (Vauquelin de la Fresnaye, trad. *Chronique de Edmond. de Dynter* IV, 28, Xavier de Ram. 1560)  
Toutefois, il résista très longtemps aux insolences et **aux insultes** de ceux de Grimbergen.

*Insult*<sup>17</sup> substantif masculin dès 1380, et ce, jusqu’au XVII<sup>e</sup> siècle désignait une ‘révolte’, un ‘soulèvement’ [1<sup>re</sup> acception] :

- (10) Le mardy 25 d’octobre au soir toute la nuit ensuivant fut fait ung grand **insult** a Montpellier par aucuns du popullaire, auquel furent occis plusieurs grandz officiers de nostre sire le roy. (*Procédure faite contre ceux de Montpellier*, Dupuy, I, 172, Richel, 1380)  
Le mardi 25 octobre au soir et durant toute la nuit, une grande **révolte** fut faite à Montpelier par quelqu’un du peuple, durant cette nuit furent tués plusieurs grands officiers de notre roi.

Durant le XVI<sup>e</sup> et le XVIII<sup>e</sup> siècles, *insult* désigne une ‘attaque’ [2<sup>e</sup> acception] – reprenant alors la première acception de *insultation* – que celle-ci soit une ‘agression militaire’, une ‘attaque armée contre quelque chose ou quelqu’un’ ou un ‘assaut mené par surprise’. Dans cet emploi, *insult*, *attaque* et *assaut* étaient perçus comme « synonymes » :

- (11) Et Mr le mareschal l’accompagne partout ; ils estoient hier à \*Berre, et aujourd’huy sont à Marseille d’où l’on avoit cez jours passez envoyé [...] à Tollon pour se premunir contre **un insult** de cette armée navale qui tient le monde en grande allarme. (Peiresc, Nicolas de, *Lettres*, t. III : *Lettres aux frères Dupuy*, 1634, p. 185)  
Monsieur le maréchal l’accompagne partout : hier, ils étaient à Berre et aujourd’hui à Marseille d’où on l’avait, les jours passés, envoyé [...] à Tollon pour se prémunir contre **un assaut** de cette armée navale qui terrorise tout le monde.

Au XVII<sup>e</sup> siècle, *insult* est plus spécifiquement employé pour désigner une ‘querelle ou une attaque physique dirigée par surprise envers un individu’ [3<sup>e</sup> acception]. Cependant, nous n’en avons trouvé aucune illustration dans Frantext comme dans les ouvrages encyclopédiques consultés. Cette acception d’‘agression physique’ n’est plus attestée, mais *insulte* en garde la connotation, lorsque l’on parle d’« agression » ou

---

<sup>17</sup> s.v. INSULTE dans le DELF.

d'« attaque verbale » dirigée avec véhémence envers quelqu'un, une sorte d'*invective* en somme.

Dans le même temps au XVII<sup>e</sup> s., *insulte* devient un substantif féminin, mais les deux genres restent en usage. Le nom ne désigne donc plus une 'attaque physique', mais une 'parole', une 'attitude' « qui porte atteinte à l'honneur ou à la dignité de quelqu'un » (TLF 1971) [4<sup>e</sup> acception] – reprenant alors la seconde acception de *insultation* :

- (12) Les [pêcheurs] voyez-vous, ces superbes [...] à qui la pudeur même semble indigne d'eux, parce que c'est une espèce de crainte : si bien qu'ils ne méprisent pas seulement, mais qu'ils font **un insulte** public à toute l'Église, à tout l'Évangile, à toute la conscience des hommes ? (Bossuet Jacques-Bénigne, *Sermon sur l'intégrité de la Pénitence*, 1662, p. 223)

De cette acception découlera celle d'*insulte* qui consiste à caractériser des 'attaques en paroles' irrespectueuses et méprisantes [5<sup>e</sup> acception], mais aussi des écrits, des gestes ou des attitudes, qui nuisent à la réputation d'un individu. L'insulte ainsi proférée porte atteinte à la *fama* [réputation], valeur centrale dans la société médiévale (Gonthier 2007 : 17), et est perçue comme un véritable acte de langage (au sens pragmatique) :

- (13) **Bastarda, filli de moyne et de chapellain !** (*Archives Départementales de la Loire*, B 1189, f<sup>o</sup> 7, *La Tour en Jarez*)  
**Bâtarde, fille de moine et de chapelain !**

En somme, terme *insulte* apparaît au XIV<sup>e</sup> siècle et prend une forme graphique plurielle (*insult*, *insultation*, *insulte*). Désignant initialement une 'révolte', un 'soulèvement' [1], son sens se spécialise au XVI<sup>e</sup> siècle et renvoie à une 'attaque (militaire, armée) menée par surprise' [2] (*assaut*) ; attaque qui sera 'dirigée envers quelqu'un' [3] au XVII<sup>e</sup> siècle (*querelle*). Au XVI<sup>e</sup> siècle, *insulte* connaît une première restriction de sens, désignant alors une 'attaque en paroles' [4], véritable joute verbale (*invective*), qui donnera lieu à sa valeur, encore actuelle, d'attaque en paroles, en gestes et en attitudes 'qui nuisent à la réputation d'un individu' [5].

### 1.3. *Injure* et *insulte*, synonymes ?

Il n'est pas aisé d'identifier la nuance sémantique entre *injure* et *insulte*. Le rapprochement sémantique des deux termes ne date que du XVII<sup>e</sup> siècle, comme le montre l'illustration 1. En effet, le schéma dresse une chronologie qui rend compte des nuances sémantiques des deux mots lorsqu'ils désignent des « paroles » (et non plus des « actes »). Pour l'*injure* [4], il s'agit de « parole outrageante que l'on dit pour offenser quelqu'un, en lui reprochant quelque défaut, ou quelque vice vrai ou faux ». Pour l'*insulte* [4] et [5], il s'agit de « parole, attitude qui porte atteinte à l'honneur ou à la dignité de quelqu'un ».

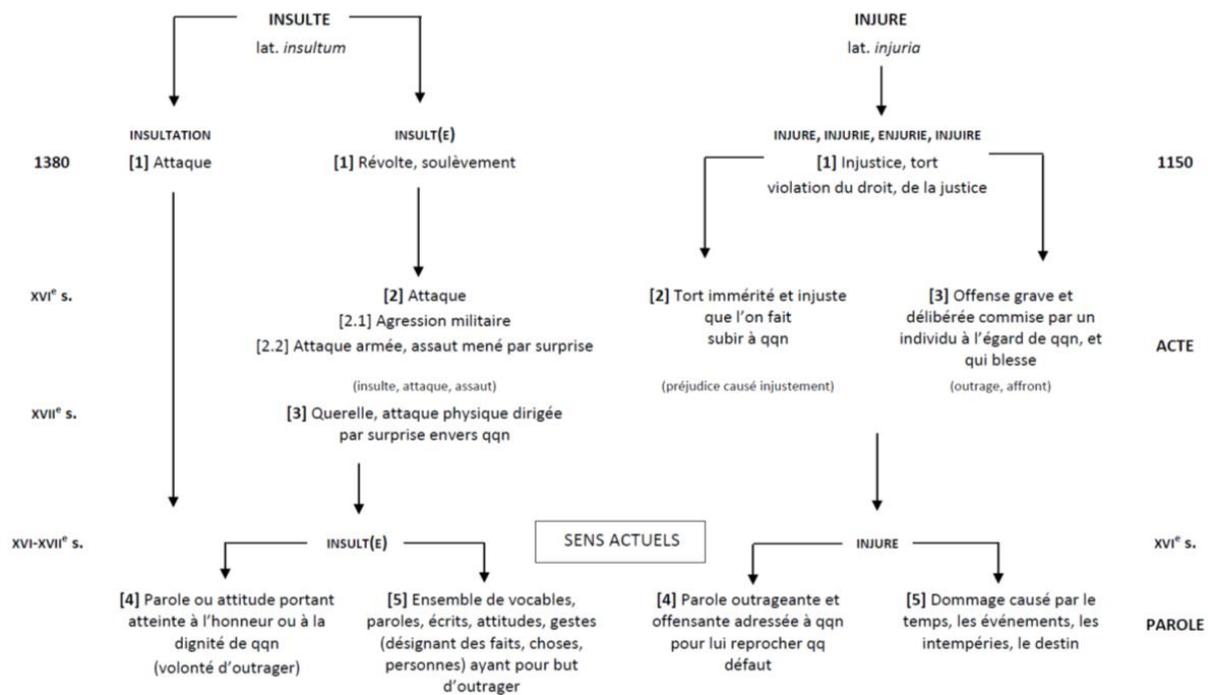


Figure 1. Synthèse des acceptions de insulte et injure

Autrement dit, les deux termes relèvent d'un « phénomène plutôt oral, comportemental, spontané, immédiat » (Fracchiolla 2011, citée par Flory 2016 : 18) et, « dans un sens commun, populaire, *insulte* et *injure* se valent à peu près l'un l'autre » (*ibid.*).

Malgré ce rapprochement d'ordre sémantique, certains chercheurs ont essayé de les distinguer, en s'appuyant sur la relation *hyperonyme/hyponyme*. Ainsi, *injure* est-il perçu comme « un terme plus large » (Flory 2016 : 15), « moins précis » (Hammer 2009 : 17 citée par Flory 2016 : 15) ou englobant, au point que F. Hammer en vient à se demander si « le terme *injure* [ne serait pas] hyperonyme de *insulte* » ; elle propose donc de :

---

réserver le terme d'*insulte* aux agressions verbales tandis qu'*injure* serait applicable à toutes les formes d'agressions gestuelles et/ou comportementales. La différenciation des deux notions se fonde ainsi sur l'intensité de la violence exercée. (Hammer 2009 : 17 citée par Flory 2016 : 15)

---

Cette analyse de F. Hammer mérite discussion. Au regard de l'étude que nous venons de proposer *supra*, il semble que ce soit *insulte* qui serait le plus à même de désigner tout type d'agressions (gestuelles, verbales, comportementales, etc.), alors que *injure* désignerait, plus spécifiquement, les actes de violence verbale.

*Injure* admet une connotation relative au préjudice moral ; cette parole est proférée dans l'intention de juger, de reprocher des vices/défauts à quelqu'un. Les injures sont le reflet avant tout des propos évaluatifs et subjectifs d'un locuteur concernant l'attitude, le comportement d'un tiers. Les conséquences de ces paroles proférées sont certes blessantes, offensantes, mais non destructrices, pour celui qui les reçoit. Leur intensité n'est donc pas associée aux conséquences subies ; elle repose sur leur

émission même, la violence des mots tels qu'ils sortent de la bouche de celui qui les prononce. L'injure est d'ailleurs considérée comme un délit, puisque le fait même de dire une injure se retourne contre celui qui la prononce.

*Insulte* conserve sa connotation d'attaque physique ; cette parole est proférée dans l'intention de porter atteinte à la dignité d'autrui, et ce, alors qu'il ne s'y attend. Les insultes sont d'une extrême violence, destructrices socialement, puisqu'elles nuisent à la réputation de celui qui est insulté et, « par ricochet » (Lagorgette 2008) à celle du délocuté par le recours, notamment à la construction <filz de X> / <filz a X> ou par le sémantisme associé à *bastarda* (impliquant l'abandon de l'enfant par le parent). L'intensité de cette violence verbale peut être apparentée, d'une part aux conséquences dévastatrices, fâcheuses pour celui qui les reçoit et, d'autre part, à l'intention de nuire de celui qui les émet, relevant d'une marque d'irrespect et de « mépris envers quelque chose ou quelqu'un » (Lagorgette 1994 : 318) dans sa volonté d'outrager.

## 2. Démêler le gros mot, le blasphème et le juron

### 2.1 Le gros mot

Les termes *injure* ou *insulte* désignent un certain type de paroles proférées qui ne renvoient pas exclusivement à des mots reconnus comme grossiers, contrairement aux *gros mots* ou aux *noms d'oiseaux*. En effet, à l'instar des *gros mots* reconnus comme tels, les termes constitutifs de l'*insulte* et de l'*injure* ne sont pas toujours des lexèmes orduriers. Si *gourgandine* (désignant péjorativement une fille de joie à l'époque médiévale) est d'emblée associé à un vocabulaire jugé ordurier, le segment injurieux *gros menton !* (14) n'est lexicalement pas reconnu comme ce que l'on qualifierait aujourd'hui de *gros mot* alors que, dans son contexte initial, il s'agit bel et bien d'une injure :

- (14) H : Pourpoint gras ! et vous, dame orda, / On vous appelle Giroflée. / [...] F : **Gros menton !** (ANONYME, « Le Chaudronnier », *Farces du Moyen-Âge*, 1547-1557, v. 38-48)  
H : Pourpoint gras ?! Et vous, sale putain ! / On vous appelle clou de girofle. / [...] F : **Gros menton !**

*Gros mot* est attesté, dès le XII<sup>e</sup> siècle, avec une acception spécifique de l'adjectif *gros* (dans le sens de 'rude') sous la forme *gros moz*. Il désignait alors de 'rudes paroles' (15) [1<sup>re</sup> acception].

- (15) **Gros moz** dire gueres ne couste ! (Anonyme, *Roman de Thèbes*, 1150, v. 4902)  
Prononcer **un/des gros mot/s** [de rude(s) parole(s)] ne coûte rien !

Cette acception est abandonnée au XVI<sup>e</sup> siècle au profit de 'grossier' qui admet un sémantisme identique à celui que nous connaissons aujourd'hui. *Gros mot* désigne alors de 'grossières paroles' [2<sup>e</sup> acception], soit des mots « qui manque[nt] à la bienséance, à la pudeur, à la décence » (GLLF, TLF) :

(16) Le villain, syllogisant à part luy, faisoit de terribles et enormes figures, une fois haulsant son chapeau, passant le doigt à sa ceinture, grondoit certains **gros mots** entre les machoueres, [...]. (Matthieu Pierre, *La Guisiade*, acte III, scène 2, 1589, p. 133)

Le paysan, syllogisant en lui, mimait de terribles et énormes personnages, une fois haussant son chapeau, passant le doigt à sa ceinture, murmurant plaintivement certains **gros mots** [de grossières paroles] entre ses dents, [...].

Au XVII<sup>e</sup> siècle, *gros mot* est employé dans l'expression *dire des gros mots* [3<sup>e</sup> acception] pour désigner l'action de se quereller en paroles / par la langue ('avoir de grosses paroles avec quelqu'un') :

(17) Après avoir essayé les paroles douces, j'en vins aux **gros mots**, je l'appellai coquin, maraud, fripon. (Diderot Denis, *Lettres à Sophie Volland*, t. 1, 1762, p. 143)

*Gros mot* était utilisé dans la langue littéraire [4<sup>e</sup> acception] comme « mot emphatique » (GRLF, TLF), au sens de 'grandiloquent'<sup>18</sup> (et était alors rapproché de *grand mot*) ; il désignait alors un « mot qui exprime une réalité importante, grave » (GRLF) :

(18) L'honneur... l'honneur... Avec toi, tout de suite les **gros mots** ! (ANOUILH Jean, *Ornifle*, II, 1955, p. 102)

Plus récemment, *gros mot* est considéré comme une locution qui dénote un mot cru et brutal, qui manque de finesse. *Gros mot* réfère alors à un « mot obscène » (Furetière) [5<sup>e</sup> acception], voire un mot « choquant » (GRLF) qui est alors rapproché (en tant qu'emploi restreint et moderne) des « termes scatologiques ou sexuels prohibés notamment dans le vocabulaire enfantin » [6<sup>e</sup> acception].

(19) Allez en arriere **merdaille** ! (*La Farce de Maistre Pathelin*, 1456-1469, v. 850)  
*Reculez, tas de merde !*

Nous résumons, dans l'illustration 2, les différentes acceptions du lexème *gros mot* en synchronie et en diachronie. Ainsi, le terme *gros mot* apparaît au XII<sup>e</sup> siècle et désigne de 'rudes paroles' [1] puis, au XVI<sup>e</sup> siècle, des 'paroles grossières' voire 'triviales' [2] ou, dans la langue littéraire, de 'grands mots exprimant une réalité grave' [4]. Au XVII<sup>e</sup> siècle, *dire des gros mots* signifie 'se quereller' [3]. Ce n'est qu'à partir du XVII<sup>e</sup> siècle que la locution *gros mot* désigne le 'mot obscène et choquant' [5], qui sera associé au vocabulaire sexuel et scatologique [6].

---

<sup>18</sup> Au Québec, l'on ne parle pas de *gros mot* mais de *sacre*, qui désigne tout autant le juron que le blasphème. Ainsi les expressions suivantes sont-elles considérées, en français québécois, comme des « gros mots », des « jurons », en somme des « sacres » : *hostie de calice !*, *tabarnak !*, *viarge !*, *calvaire !*, etc.

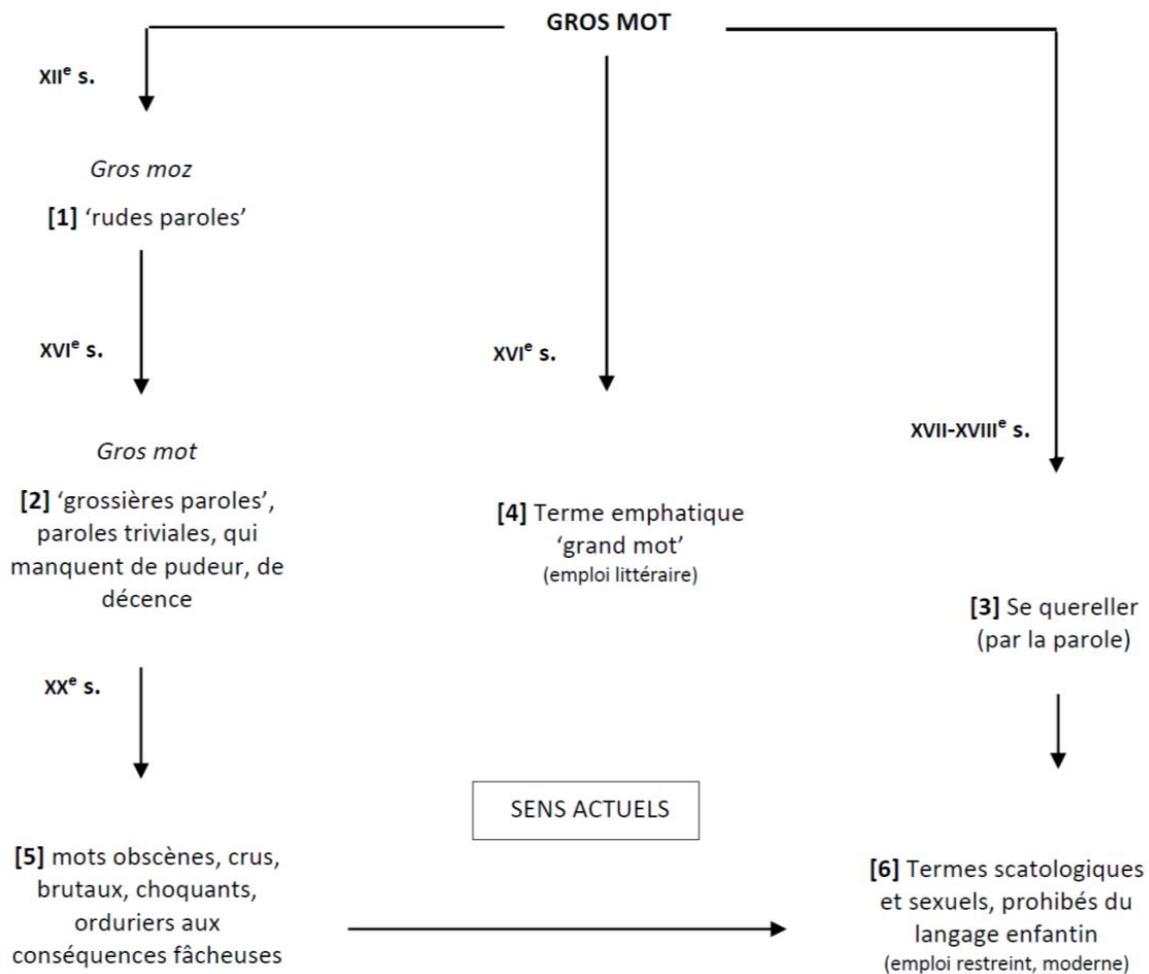


Figure 2. Synthèse des acceptions de gros mot

Enfin, quelle est la distinction à opérer entre *injure*, *insulte* et *gros mot* qui désignent tous trois des paroles outrageantes ? La spécificité de *gros mot* est qu'il qualifie la nature même de ces paroles axiologiques négatives : elles sont grossières, ordurières, obscènes, triviales alors que celles liées à *injure* et *insulte* étaient identifiées au regard, d'une part, de leur réception par un tiers (*insulte*) et, d'autre part, de leur émission dans l'intention de nuire (*insulte*). Autrement dit, *gros mot* n'est pas systématiquement assimilable à l'*injure* ou à l'*insulte* ; ils ne sont pas synonymes, comme le note C. Rouayrenc :

Il est des injures qui ne sont pas des gros mots (p. ex. *pintade*, *blaireau*) et, inversement, certains des mots considérés comme *gros mots* (p. ex. *branlette*, *pine*) ne sont par eux-mêmes ni des jurons, ni des injures. (Rouayrenc 1998 : 3-7)

Si *gros mot*, *insulte* ou *injure* ne sont certes pas synonymes, il n'en demeure pas moins que, dans certains contextes, les *gros mots* peuvent être interprétés/perçus comme des

*insultes*<sup>19</sup>. Nous proposons donc de considérer *gros mot* comme un « hyponyme » de *injure* et d'*insulte*.

## 2.2 Le blasphème

Dater avec précision la première attestation de *blasphème* est une tâche délicate, puisque les ouvrages lexicographiques tels que le DEHLF, le DHLF et le DELF mentionnent son origine relativement floue et dont l'étymologie ambivalente est à la fois grecque<sup>20</sup> et latine<sup>21</sup>.

Quoi qu'il en soit, *blasphème* désigne une parole 'outrageant la divinité' ou 'méprisant la religion (parole impie)' (pour son origine latine) – qui serait à rapprocher de *jurement*<sup>22</sup>, *sacrilège* – ou 'de mauvaise augure (causant du tort, des dommages) ne devant pas être prononcée' (pour son origine grecque) [1<sup>re</sup> acception] :

(20) Li enduremenz e li **blaphemes** el saint esprit. (*Sermon de Saint Bernard*, 1190, 117, 20)

L'endurcissement et le **blasphème** du Saint Esprit.

Au XV<sup>e</sup> siècle, *blasphème* ne désigne plus uniquement un outrage à la divinité ; il est euphémisé en plus d'être laïcisé, et doit être rapproché des termes *médiance*, *calomnie* ou *accusation* qui désignent une 'parole malveillante, mensongère, qui cause du tort à quelqu'un' [2<sup>e</sup> acception] :

(21) Et en ce commandement est deffendue toute mençonge, tout parjure, toute **blapheme**. (DMF s.v. BLASPHEME, Christine de Pizan, *Epître Othea L.*, c.1400-1401, 205)

Ce commandement interdit tout mensonge, tout parjure et tout **blasphème**.

Au XVI<sup>e</sup> siècle, par extension de l'acception de 'parole outrageant la divinité', *blasphème* désigne un discours 'outrageant à l'égard de tout ce qui est considéré comme sacré' [3<sup>e</sup> acception]. Dans l'exemple (22), le blasphème ne porte pas atteinte à un unique Dieu, mais à un ensemble de divinités païennes considérées comme sacrées :

(22) Quelle folie ! mais quel impudent **blapheme** eust-ce esté de dire Minos avoir esté fils de Jupiter, Ascalaphe de Mars, Enée de Venus, et Autolice de Mercure, qui n'entendroit ces Planettes avoir eu le principal pouvoir à leurs naissances ?

<sup>19</sup> Ce que révèlent les structures syntaxiques dans lesquelles ils apparaissent et les effets qu'ils produisent : ils peuvent avoir de fâcheuses conséquences pour celui qui les reçoit.

<sup>20</sup> *Blasphème*, pour le DHLF, aurait pour étymon grec *blasphêmia* ('injure, calomnie'), construit sur *blabos-* ('dommage, tort') et *-phême* ('divulgateur de la parole'). *Blasphêmia* s'oppose à *euphêmia* ('euphémisme'). *Blasphêmia* serait le déverbal de *blasphêmein* ('injurer', 'calomnier', 'dire du mal de quelqu'un', 'maudire', 'renier' dans *La Bible* hébraïque).

<sup>21</sup> *Blasphème*, selon le DELF et le DEHLF, serait emprunté au latin chrétien *blasphemia* (fin du XII<sup>e</sup> siècle).

<sup>22</sup> Notion sur laquelle nous serons amenées à revenir lors de notre étude lexicale de *juron*.

(TYARD Pontus de, *Mantice : discours de la vérité de Divination par Astrologie*, 1587, p. 146)

Le blasphème est alors considéré comme un « crime énorme qui se commet contre la divinité par des paroles ou des sentiments qui choquent sa majesté ou les *mystères de la vraie religion* » (Furetière)<sup>23</sup>. Le blasphème désigne donc également, dans cette acception, des actes préjudiciables commis à l'encontre du Roi, reconnu comme lieutenant de Dieu sur terre. Outrager le roi, c'est insulter Dieu et donc faire preuve de blasphème :

(23) Et ilec sans avoir chef ne homme de conduite, se fortifierent fort, et firent de grans **blasphemes** au roy, comme faire gibetz en ladicte ville et sur les murs, et y pendre croix blanches, monstrent leur cul et autres villenyas. (ROYE Jean de, *Chronique Scandaleuse*, t. 2, 1460, p. 45)

En ce lieu-là, sans même avoir ni de chef ni d'homme pour les diriger, ils devinrent plus forts et firent de grands **blasphèmes** au Roi, comme ceux d'établir un gibet en ladite ville, pendre des croix blanches sur les murs, montrer leur cul, et autres vilénies.

Au XVI<sup>e</sup> siècle et par extension, *blasphème*, signifie outrage 'porté à une chose ou à une personne considérée comme très respectable, quasi sacrée' [4<sup>e</sup> acception] :

(24) Il esclatta tout à fait, et prononce de si estranges **blasphemes** contre la sacree humanité de Jesus Christ que Fracon [...] mit deux fois la main sur son poignard pour luy plonger dans le sein. (GARASSE François, *La Doctrine curieuse des beaux-esprits de ce temps*, 1623, p. 146)

Il s'emporta et prononça de si étranges **blasphèmes** contre l'humanité sacrée de Jésus-Christ que Francon [...] porta deux fois la main à son poignard pour lui plonger dans la poitrine.

De cette acception découlera celle de *blasphème* désignant 'une insulte, une imprécation, contenue dans une parole, un propos, un acte injuste, injurieux' [5<sup>e</sup> acception].

Comme nous l'avons résumé dans l'illustration 3, issu d'un double étymon grec et latin, *blasphème* désigne, au XII<sup>e</sup> siècle, une 'parole outrageant la divinité' ou 'méprisant la religion' [1], puis au Moyen-Âge une 'parole malveillante, mensongère' [2]. Ne se limitant plus aux divinités, le blasphème peut alors concerner 'tout ce qui est considéré comme sacré' (le Roi par exemple) [3] ou 'respectable' [4]. De façon plus générale, le blasphème réfère à des 'insultes contenues dans des propos ou actes injurieux' (sémantisme proche de *injure* [2]).

---

<sup>23</sup> *Crime* ne doit pas être perçu comme hyperbolique, puisqu'au Moyen-Âge celui qui blasphème est puni par mutilation (« on perce, on arrache la langue », Lagorgette 2003 : 180), voire par la mort.

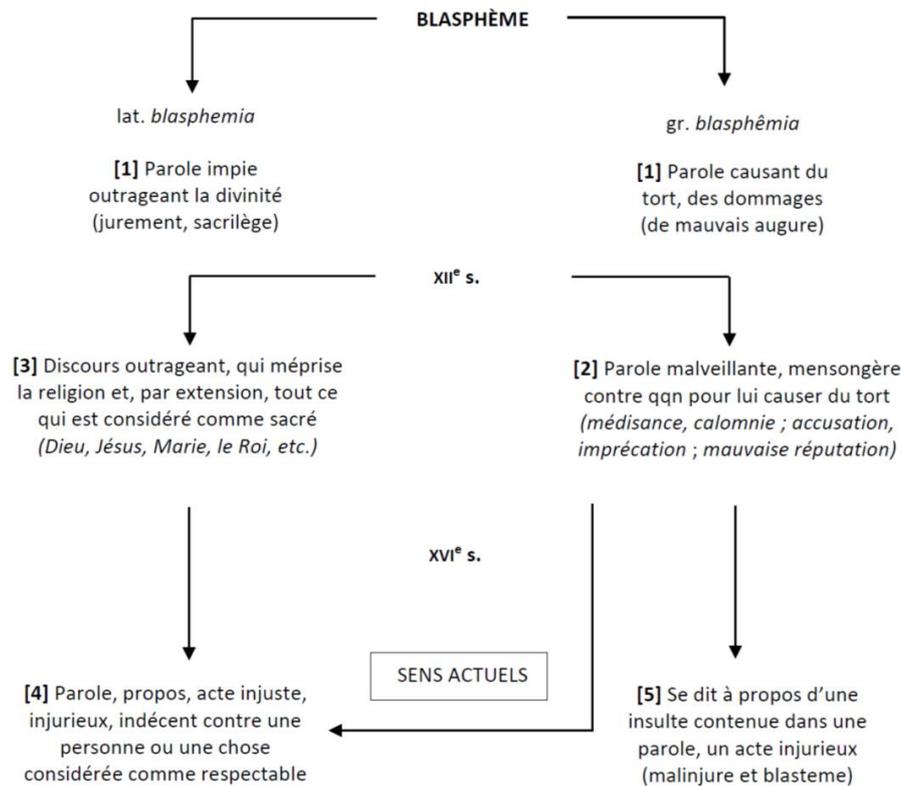


Figure 3. Synthèse des acceptions de blasphème

Enfin, alors qu'ils désignent tous trois des paroles proférées, quelle est la distinction à opérer entre *injure*, *insulte* et *blasphème* ? La spécificité de *blasphème* repose sur la nature même de la parole qui est perçue comme outrageante vis-à-vis du sacré, du divin, de ce qui doit être respecté. Toutefois, *blasphème* se rapproche de *injure* en tant que ces paroles sont offensantes, irrévérencieuses et qu'elles constituent un délit. Ce faisant, nous proposons d'identifier *blasphème* comme un hyponyme de *injure*.

Le *Lexique* de R. P. De Gorog vient étayer notre propos, dans la mesure où y figure *malinjure* dont l'acception est une synthèse de celles de *blasphème* et de *injure* :

- (25) Qui doresnavant diront, de mauvais et felon courage, **malinjure** ou blasfemie de Dieu. (1480, Ord., XIV, 499)  
 Ils diront dorénavant ceux, qui avaient de mauvaises et de terribles intentions, des **blasphèmes** à l'encontre de Dieu.

### 2.3 Le juron

Les dictionnaires consultés sont unanimes pour dater la première attestation de *juron*, déverbal de *jurere*, aux prémices de la période classique (1599). Cette apparition tardive de *juron* doit être corrélée au fait qu'un autre terme était en usage jusqu'alors, et ce, depuis le XIII<sup>e</sup> siècle – *jurement* – et qu'il avait pour acceptions celles désignées par *serment* (26), *blasphème* (27) et *juron* (28). La collision synonymique de *jurement* et *juron* conduira à la disparition progressive de *jurement*, qui déléguera alors tous ses sens à *juron*.

(26) Les chevaliers furent premierement fais par leur **jurement**. (Foulechat Denis, *Le policratique de Jean de Salisbury*, Livres VI et VII, 1372, p. 154a)  
Les chevaliers furent faits premièrement par la prononciation de leur **serment**.

(27) **Juremens** sont donc de tel sorte Qu'il fault que grant malheur en sorte En ame, corps, bien ou amis. (DMF, s.v. JUREMENT, Meschin., *Lun. Princes M.-G.*, c.1461-1465, 41)  
Les **blasphèmes** ont un tel pouvoir qu'il faut éprouver une grande souffrance d'âme et de corps pour les expier.

(28) **Jurement** est une attestation de Dieu pour confermer la vérité de nostre parole. (Calvin Jean, *Institution de la religion chrestienne*, 1560, p. 156)  
Le **juron** est une attestation de Dieu pour confirmer la véracité de nos dire.

*Juron* connaît deux acceptions principales : celle de serment [1<sup>re</sup> acception] désignant l'« action de jurer » jusqu'au XVII<sup>e</sup> s. (29) et celle de « jurement grossier » à partir de 1606 [2<sup>e</sup> acception] recouvrant alors les acceptions de « blasphème » et « juron » initialement dévolues à *jurement*.

(29) Je jure mon grand **juron**, qu'ils s'en repentiront. (Coypeau d' Assoucy Charles, *L'Ovide en belle humeur*, 1650, p. 42)

La première acception (serment « action de jurer ») donnera lieu à la désignation de *juron* comme « terme dont on se sert pour jurer » [3<sup>e</sup> acception] : *Par Dieu !* ou *Par la vierge Marie !* seront employés pour « faire serment », pour jurer. Puis, il désignera plus spécifiquement, dès 1690, une « façon particulière de jurer » [4<sup>e</sup> acception] propre à un individu, à des peuples (p. ex. *Ventre Saint Gris !* était le juron de prédilection de Henry IV).

La seconde acception de *juron* (« jurement grossier »), encore en usage aujourd'hui, s'est dotée d'une coloration religieuse désignant une « parole offensante à l'égard de Dieu pris à témoin d'une affirmation » [5<sup>e</sup> acception]. Autrement dit, cet emploi de *juron* peut être rapproché de celui de *blasphème* [1]. Toutefois, pour éviter de tomber dans le *blasphème*, « le juron constitue souvent en un euphémisme vidé de son sens » (GRLF). Ainsi, on trouve couramment des formes en *-bleu* ou *-bieu* pour euphémiser *Dieu* (comme *palsambieu !* pour *par le sang de Dieu !* ou *sacrebleu !* pour *sacré Dieu !*). De même, *par la mer Dé !* (Rabelais, *Gargantua*, 1534, p. 131), euphémisation cocasse, puisqu'elle sert le comique grivois scatologique, thème de prédilection de François Rabelais. En effet, *par la mer Dé !* est un jurement dont la restitution entière serait *par la mère de Dieu !* ; toutefois, pour éviter de commettre l'acte de jurer, le locuteur prononce une forme euphémisée. Finalement, *par la mer Dé !* peut également être compris comme *par la merde !*, et ainsi être lu sous le prisme

d'un registre scatologique. Il était de coutume à l'époque de véhiculer l'idée suivante : « on ne doit pas croire les marchands avec tous leurs jurements<sup>24</sup> » (Furetière) :

(30) Tout beau ! Ne jurez pas ; et principalement ce **juron**, qui est toujours en la bouche des putains : si on vous oyoit, que diroit-on de vous ? (Béroalde de Verville François, *Le Moyen de parvenir*, 1610, p. 193)

Tout beau ! Ne jurez pas, surtout avec ce **juron** qui est souvent prononcé par la bouche des putains : si l'on vous entendait, que dirait-on de vous ?

*Juron* finit par désigner, plus familièrement, un 'terme d'emportement, d'exécration' (« une interjection ou exclamation grossière ou familière » TLF) que l'on prononce pour 'exprimer des violentes passions' (joie, peine, colère, surprise, dépit etc.) [6<sup>e</sup> acception].

(31) Mon grand-père employait son grand **juron** contre cette Mme Vignon : **le diable te crache au cul !** (Stendhal, *Vie de Henry Brulard*, 1836, p. 204)

Ainsi, le mot *juron* vient remplacer *jurement*, à la fin du XVI<sup>e</sup> siècle, et hérite de ses acceptions : celle de serment 'action de jurer, prier, promettre' [1] et celle de 'jurement grossier', à partir du XVII<sup>e</sup> siècle [2]. La première acception aboutira à désigner l'ensemble des termes qui permettent de jurer [3] ; puis, par restriction au XVII<sup>e</sup> siècle, ceux spécifiques à une personne [4]. La seconde, doté d'une coloration religieuse, désignera les paroles prenant Dieu à témoin (le locuteur espérant ainsi rendre son propos vraisemblable) ; puis, le *juron* servira de support pour traduire des émotions [6].

---

<sup>24</sup> Précepte qui rappelle la scène de Pathelin et du drapier (dans la *Farce de Maistre Pathelin*), lorsque le protagoniste tente de convaincre le marchand de lui faire crédit de quelques aunes de draps. Pour ce faire, il ponctue ses propos de quelques jurements pour attester de sa bonne foi en tant que bon chrétien.

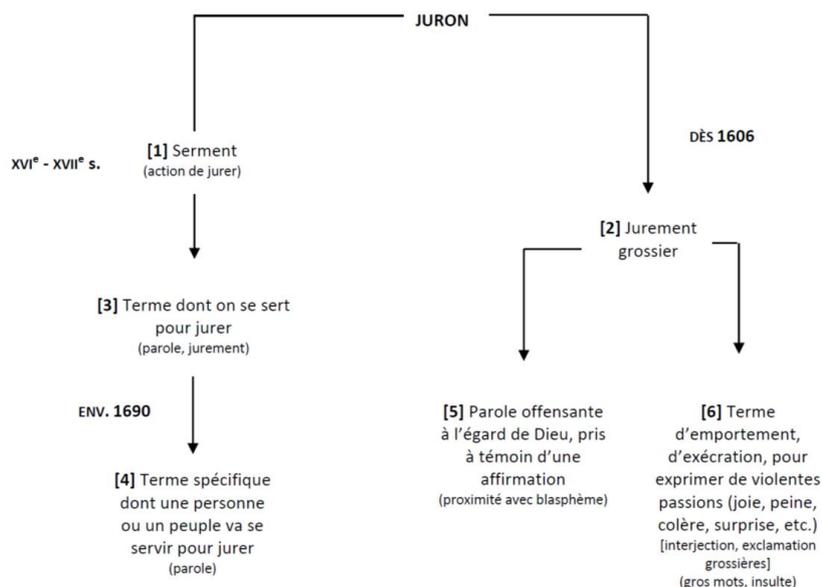


Figure 4. Synthèse des acceptions de juron

On peut donc s'interroger sur la distinction à opérer entre *injure*, *insulte* et *juron*. La spécificité de *juron* repose sur la nature même de la parole qui est perçue, soit comme offensante à l'égard de Dieu (le juron se rapproche alors du blasphème et de l'injure en ce qu'il constitue un délit<sup>25</sup>), soit qui exprime l'emportement de celui qui l'énonce traduisant ainsi ses « violentes passions ». La violence associée au juron permettrait alors de mettre à nu son énonciateur<sup>26</sup> ; le juron permet l'expression de la « fonction émotive du langage » (Jakobson 1960) et serait à rapprocher de l'injure :

[...] il est à supposer que les *jurons*, qui sont des exclamations entièrement dépourvues de sens, ont été inventés comme instinctivement pour donner issue à la colère sans rien dire de blessant ni d'irréparable. (Alain, *Propos sur le bonheur*, Paris : Éditions Gallimard, 1928 ; s.v. LXXII. Injures, 17 novembre 1913)

Notons que, dans un sens moderne, certains termes injurieux ayant perdu leur force injurieuse initiale peuvent être considérés comme des jurons, admettant une acception émotive, puisqu'ils ont subi une grammaticalisation (Marchello-Nizia 2006) : ils sont passés de la catégorie lexicale de *nom* à celle grammaticale d'*interjection*. C'est le cas, par exemple, de *putain !*, *enculé !* prononcés dans un contexte de surprise (p. ex. **putain/enculé ! tu m'as fait peur !**), d'énervement (p. ex. **putain/enculé ! tu m'agaces !**), de dépit (p. ex. **putain/enculé ! j'en ai marre !**), de joie (p. ex. **C'était génial, putain/enculé, c'est super ça !**), etc.

<sup>25</sup> Le juron, tout comme le blasphème, peut être rapproché de *sacre* en québécois : « Le gars qui échappait un **sacre** dans ses colères. » (Lemelin Roger, *Au pied de la pente douce*, 1944, p. 40). Ainsi, *Pute Marie !* peut référer à un juron exprimant un sentiment, mais également se rapprocher du blasphème et du juron, puisqu'associer le terme injurieux *pute* à la figure divine de *Marie*, c'est offenser Dieu. En effet, proférer de telles paroles, c'est profaner une des figures sacrées qui l'entoure mais également, en usurpant le droit divin, renommer la traditionnelle Vierge Marie en *Pute Marie* (autrement dit, en *Marie Putain*).

<sup>26</sup> La violence associée au juron se retrouve également dans des expressions comme {*lâcher, proférer, lancer*} *un juron*, témoignant d'un véritable acte de langage (Armengaud 1985) produit par le fait même de parler.

## Conclusion

Cet article a permis de démontrer que *injure*, *insulte*, *gros mot*, *blasphème*, *juron* ne peuvent être considérés comme de purs synonymes, même s'ils peuvent être regroupés sous le sème de 'parole outrageante, offensante'. Autrement dit – tel est le cas dans l'illustration 5, ils peuvent tous être sous la coupe sémantique de l'*invective*, sorte d'hyperonyme dont ils seraient les hyponymes<sup>27</sup> à coloration axiologique négative, mais chacun avec leurs nuances, que nous rappelons *infra* :

- L'**injure** est proférée afin d'apporter un jugement subjectif (préjudice moral) sur autrui concernant son attitude, son comportement qui justifie l'injure. D'un point de vue lexicographique, les conséquences de cette parole injurieuse seront certes blessantes, offensantes, mais non destructrices, pour celui qui la reçoit. Son intensité ne repose donc pas sur les dommages subis par le tiers, mais sur son émission même, sa production effective. Autrement dit, l'injure, par sa violence, se retourne contre le locuteur puisqu'il s'agit d'un délit.
- L'**insulte** est proférée dans l'intention de porter atteinte à quelqu'un alors même qu'il ne s'y attend pas (attaque physique) ; elle nuit à la réputation de celui qui est insulté et, « par ricochet », à celle du délocuté (<filz de X>). Aussi, les conséquences de cette parole insultante seront-elles destructrices, dévastatrices notamment socialement pour celui qui la reçoit. Son intensité excessive repose à la fois sur les dommages subis par le tiers mais également sur l'intention délibérée de nuire/outrager de celui qui l'émet.
- Le **gros mot** qualifie la nature même de la parole proférée (grossière, ordurieuse, obscène, triviale) et, en fonction, des dommages subis par celui qui le reçoit, il sera à considérer comme « un type de » soit d'insulte soit d'injure.
- Le **blasphème** qualifie la nature même de la parole proférée (outrageante, offensante, irrévérencieuse vis-à-vis du sacré, du divin, de ce qui doit être respecté) qui ne s'adresse pas forcément à quelqu'un. Il serait à identifier comme « un type d' » injure (en ce qu'il constitue un délit).
- Le **juron** qualifie la nature même de la parole proférée : soit elle est offensante à l'égard de Dieu (le juron serait à identifier comme « un type d' » injure, au même titre que le blasphème, en ce qu'il constitue un délit), soit elle exprime l'affect de celui qui l'énonce (colère, joie, peine, dépit, etc.) tout en le mettant à nu (puisque c'est ce dernier qui en subit les dommages). Il ne s'adresse pas forcément à quelqu'un (là encore, il serait à identifier comme « un type d' » injure).

---

<sup>27</sup> Nous rejoignons en cela Dubreuil (2011 : 17).

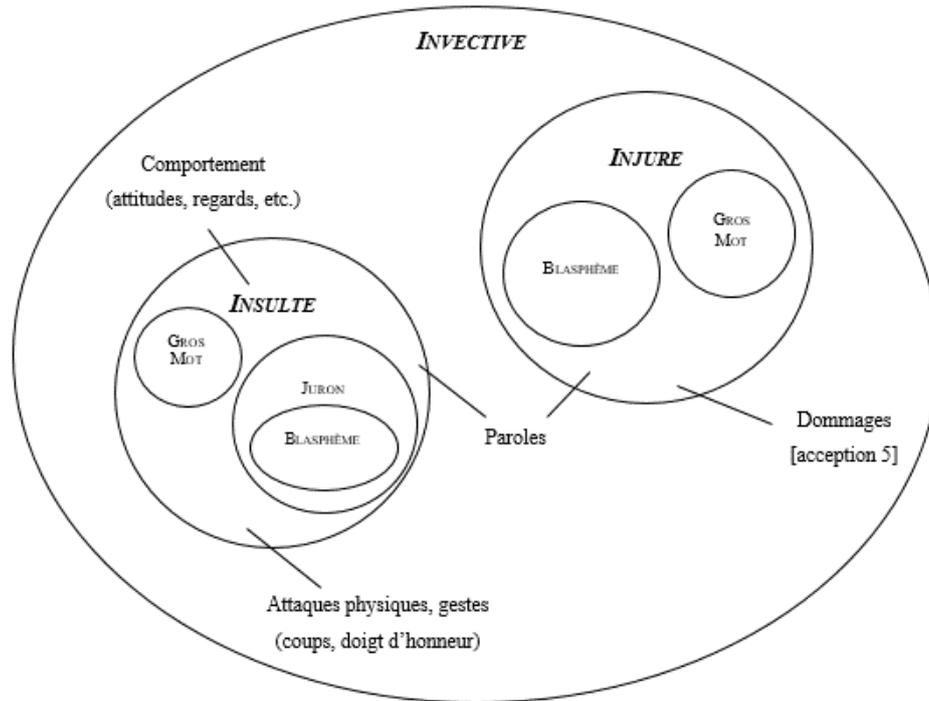


Figure 5. Réseau hypéronymique de l'insulte

## Références bibliographiques

### *Ouvrages théoriques*

- Armengaud, Françoise. 1993. *La pragmatique*. Paris : Presses Universitaires de France.
- Dubreuil, Philippe. 2013. *Le marché aux injures à Rome : injures et insultes dans la littérature latine*. Paris : L'Harmattan.
- Flory, Julienne. 2016. *Injurez-vous ! Du bon usage de l'insulte*. Paris : La Découverte.
- Fracchiolla, Béatrice. 2011. « Injure » dans *Dictionnaire de la violence*. Paris : Presses Universitaires de France. 706-710.
- Gonthier, Nicole. 2007. “*Sanglant coupaul!*” “*Orde ribaude!*”: les injures au Moyen Âge. Rennes : Presses Universitaires de Rennes.
- Hammer, Françoise, 2009. *Cherchez l'insulte ! Trouvez l'outrage ! Une approche du champ vexatoire*, dans Lagorgette, Dominique (dir). *Les insultes en français : de la recherche fondamentale à ses applications (linguistique, littérature, histoire, droit)* Chambéry : Université de Savoie.
- Jakobson, Roman. 1960. “Linguistics and Poetics” in Sebeok, Thomas A. (eds.). *Style in Language*. Cambridge, Massachusetts : MIT Press.
- Lagorgette, Dominique. 1994. *Termes d'adresse, acte perlocutoire et insultes : la violence verbale dans quelques textes des XIV<sup>e</sup>, XV<sup>e</sup> et XVI<sup>e</sup> siècles*, dans Jean Arrouye, Anne Berthelot, Catherine Blons-Pierre, Marie-Odile Bodenheimer, Marie Rose

- Bonnet, Danielle Buschinger, *et al.* 2013. *La violence dans le monde médiéval*. Aix-en-Provence : Presses Universitaires de Provence. 317-332.
- Lagorgette, Dominique. 2003. « Les syntagmes nominaux d'insulte et de blasphème : analyse diachronique du discours marginalisé », *Thélème, Revista Complutense de Estudios Franceses*. 171-188.
- Lagorgette, Dominique. 2008. *Les insultes par ricochet (fils de, cocu et consorts) : de quelques avanies du lexique insultant – quels critères pour l'outrage verbal ?*, in Desmons, Eric, Marie-Anne Paveau & Eric Beaumatin (dir.). *Outrages, insultes, blasphèmes et injures : violences du langage et polices du discours*. Paris : Harmattan. 7-30.
- Marchello-Nizia, Christiane. 2006. *Grammaticalisation et changement linguistique*. Bruxelles : De Boeck & Larcier.
- Molinu Lucia, 1999. « Métathèse et variation en sarde », *Cahiers de Grammaire* 24. 153-181.
- Raymond, Laurie, 2019. *L'insulte : étude et enjeux du phénomène au Moyen-Âge*, mémoire de Master, codirection Céline Vaguer et Thomas Verjans. Université Toulouse Jean-Jaurès.
- Rouayrenc, Catherine. 1998. *Les gros mots*. Paris : Presses Universitaires de France.
- Vaguer, Céline. 2017. *Une foule de témoins, un homme de cœur, cet abruti de Romero, un panier de crabes*. <Dét N1 de N2>, un syntagme nominal sous contraintes ?, dans C. Lachet, L. Meneses-Lerín & A. Roig (dir.), *Contraintes linguistiques : à propos de la complémentation nominale*, Bruxelles : Peter Lang. 139-156.

### **Corpus d'étude**

- [DAALF] Robert, Paul. 1969. *Le Robert : Dictionnaire alphabétique et analogique de la langue française*, Paris : Imprimerie Paul Dupont.
- [DEHLF] Baumgartner, Emmanuèle & Menard, Philippe. 2004. *Dictionnaire étymologique et historique de langue française*, Italie : Livre de poche.
- [DELFL] Bloch, Oscar & Wartburg, Walther von. 2008. *Dictionnaire étymologique de la langue française*, Paris : Presses Universitaires de France.
- [DHLF] Rey, Alain, Tomi Marianne, Hordé Tristan et Tanet Chatal, 1992. *Le Robert : Dictionnaire historique de la langue française*, Paris : Dictionnaire Le Robert.
- [Frantext] *Base textuelle FRANTEXT*, ATILF - CNRS & Université de Lorraine, [en ligne], URL : <http://www.frantext.fr>. Version décembre 2016.
- [Furetière] Furetière, Antoine. 1773. *Le Dictionnaire universel*, Paris : SNL – Le Robert.
- [GLLF] Guilbert, Louis, Lagane, René & Niobey, Georges. 1971. *Le Grand Larousse de la Langue Française*, Paris : Larousse.
- [Greimas] Greimas, Algirdas Julien. 2007. *Grand dictionnaire Ancien français*, Paris : Larousse.

- [GRLF] Rey, Alain. 2001. *Le Grand Robert de la Langue Française*, Paris : Dictionnaire Le Robert.
- [Godefroy] Godefroy, Frédéric. 1892. *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX<sup>e</sup> au XV<sup>e</sup> siècle*, Chartres : Imprimerie Durand.
- [Gorog] De Gorog, Ralph P. 1973. *Lexique français moderne-ancien français*, Athènes : Presses Universitaires de Géorgie.
- [TLF] Imbs, Paul. 1971. *Le Trésor de la langue française*, Paris : Éditions du CNRS, [en ligne], URL : <http://atilf.atilf.fr/>.

## Éléments de comparaison pour une correspondance entre unités macro-syntaxiques et interactionnelles

François Delafontaine

Université d'Orléans

Laboratoire Ligérien de Linguistique (UMR 7270)

### Résumé

La macro-syntaxe est l'étude des relations grammaticales au-delà du cadre rectionnel. Elle implique, à ce niveau d'analyse, de nouveaux types d'unités délimitant un domaine entre la syntaxe classique, ou micro-syntaxe, et les domaines de l'analyse du discours ou conversationnelle. Nous nous concentrerons ici sur l'une des approches macro-syntaxiques, le modèle fribourgeois de la pragma-syntaxe, et sur ses rapports avec le domaine de l'interaction. Nous testerons, à l'aide d'annotations expérimentales réalisées dans le cadre du projet SegCor, l'hypothèse selon laquelle il existe une articulation entre les deux niveaux, autrement dit qu'à une unité maximale de la pragma-syntaxe correspondrait une unité minimale de l'interaction.

**Mots-clés :** macro-syntaxe ; pragmatique ; analyse conversationnelle ; interaction ; clause ; période ; TRP ; action ; corpus oral ; oralité

### Summary

Macro-syntax is the study of grammatical relations beyond the scope of syntactic dependency. It implies, at this level of analysis, new types of units within a domain between classical syntax, or micro-syntax, and domains such as discourse or conversational analysis. We will focus here on one of the macro-syntactic approaches, the Fribourg model of pragma-syntax, and on its relations with the domain of interaction. We will test, using experimental annotations produced within the framework of the SegCor project, the hypothesis of an articulation between both levels, meaning that to a maximal pragma-syntactic unit would correspond a minimal interactional one.

**Keywords:** macro-syntax; pragmatics; conversational analysis; interaction; clause; period; TRP; action; oral corpus; orality

### 1. L'unité maximale en macro-syntaxe

La macro-syntaxe, ou l'étude des relations grammaticales au-delà du cadre de la rection (Blanche-Benveniste 1990 : 113-114), s'est intéressée principalement à la frontière entre micro- et macro-syntaxe, entre justement le cadre de la rection (Blanche-Benveniste *et al.* 1984 ; Berrendonner 2012) et celui de la parataxe

(Deulofeu 2013). Cet intérêt est explicite dans l'approche aixoise<sup>1</sup> (Groupe de Fribourg 2012 : 341) :

---

Pour les tenants de l'*Approche pronominale*, la description de la syntaxe reste un objectif prioritaire. Ils estiment que tout n'a pas été décrit dans ce domaine et qu'il reste encore beaucoup à faire avant d'envisager le contrôle d'un domaine plus vaste et de le théoriser convenablement. Les relations entre micro et macro-syntaxe, insuffisamment développées à leur avis, ne permettent pas encore de faire des hypothèses valides sur la nature des « grandes unités ».

---

L'étude de la frontière entre macro-syntaxe et discours ou interaction, elle, a fait l'objet de moins d'attention.

L'approche aixoise a proposé que, dans leur cadre, « l'unité maximale ne peut pas être définie » (Blanche-Benveniste 1990 : 144) et parle depuis d'énoncé (Blanche-Benveniste 2010), de construction (Sabio 2006) ou encore d'unité illocutoire (Benzitoun *et al.* 2010). Dans d'autres approches, cette unité tend à être mieux définie. L'approche de Florence ou Théorie de la langue en acte (Cresti 1999, 2000 ; Cresti *et al.* 2011) parle de patron illocutif et de strophe (*stanza*) dont les frontières sont toujours identifiées par un contour prosodique dit terminal. L'approche fribourgeoise, avec la période (Berrendonner 2003, Avanzi 2005), exploite ce même contour.

Les divergences entre les approches macro-syntaxiques dépassent l'identification d'une unité maximale : elles concernent aussi le rapport entretenu par les unités avec celles des autres niveaux d'analyse : « [l'*Approche pronominale*] accepte volontiers les décalages entre les niveaux de l'analyse » (Groupe de Fribourg 2012 : 343). Cela signifie qu'il n'y a pas, en macro-syntaxe aixoise, de frontière nette entre micro et macro-syntaxe, et par extension entre macro-syntaxe et discours ou interaction.

L'approche fribourgeoise, aussi appelée grammaire de la période, établit une telle frontière et la compare à celle séparant phonème et morphème (Berrendonner & Béguelin 1989). Dans ce modèle, la macro-syntaxe ou pragma-syntaxe constitue une troisième articulation du langage à fonction communicative, que nous décrirons brièvement, et qui suppose une quatrième articulation que le modèle ne spécifie pas, pouvant correspondre à l'analyse du discours ou de l'interaction (Groupe de Fribourg 2012 : 38).

---

<sup>1</sup> Cette approche macro-syntaxique exploite le cadre de l'Approche pronominale (Blanche-Benveniste *et al.* 1984, 1990) qui définit l'unité syntaxique à partir du verbe, comme tout élément pouvant y être pronominalisé.

Phonologie	Morpho-syntaxe	Pragma-syntaxe	Interaction
phonème			
syllabe	morphème		
	clause	énonciation	
		période	...

Tableau 1. Tableau simplifié des quatre articulations

À l'unité maximale de l'articulation précédente doit correspondre une unité minimale de l'articulation suivante. On passe de la syllabe au morphème<sup>2</sup>, puis de la clause à l'énonciation, puis de la période à une unité minimale en discours ou interaction.

Notre travail porte sur cette frontière entre troisième et quatrième articulation.

Nous faisons l'hypothèse qu'à l'unité maximale de la pragma-syntaxe, c'est-à-dire la période, correspond une unité minimale au niveau de l'interaction (Mondada 1998, 2000 ; Pekarek 2004 ; Traverso 2012, 2016). Nous laissons ici de côté la question de l'analyse de discours, notamment du modèle de Genève (Roulet *et al.* 2001). En vue de tester cette hypothèse, nous comparerons les unités de la pragma-syntaxe avec celle de la linguistique interactionnelle en exploitant les annotations expérimentales de ces deux niveaux d'analyse. Nous réfèrerons dans la suite de ce travail à la linguistique interactionnelle à l'aide du terme d'interaction (voir 2.2).

Pour ce faire, nous commencerons par présenter ces annotations, le corpus sur lequel elles ont été réalisées ainsi que leur cadre avec le projet SegCor. Pour la pragma-syntaxe, nous présenterons la clause-énonciation et la période ; pour l'interaction il s'agira du TRP (*Transition Relevance Place*) et de l'action.

Une fois ces unités et frontières présentées, nous reviendrons sur les protocoles d'annotation avant d'en présenter les résultats et, de là, nous comparerons les frontières obtenues aux deux niveaux, avec l'attente que ces frontières correspondent pour une ou plusieurs de ces unités. Nous terminerons par les perspectives offertes par nos résultats, perspectives tant pour la pragma-syntaxe que pour l'interaction.

## 2. Annotations du projet SegCor

Notre travail s'inscrit dans le projet ANR-DFG SegCor (Ursi *et al.* 2018) pour la segmentation de corpus oraux français et allemands à de multiples niveaux, avec pour objectif la production de protocoles et d'outils d'annotation, ainsi que d'annotations de référence. Pour les niveaux d'annotation, le lecteur peut se rapporter au tableau 2 listant, pour chaque colonne, le nom des laboratoires impliqués et les différents niveaux traités par ces laboratoires :

<sup>2</sup> Ce passage est la double articulation de Martinet (1960) : les unités significatives formées d'unités distinctives. Cela ne suppose pas une équivalence stricte.

LLL Orléans, ICAR Lyon	IDS, Mannheim
Prosodie	Pauses
Chunking	Champs topologiques /constructions
Micro-syntaxe	
Macro-syntaxe (aixoise, fribourgeoise)	Unités illocutoires
Interaction	

Tableau 2. Annotations du projet SegCor

Le corpus exploité pour ces annotations a été constitué à moitié par la base de données allemande Folk, et pour l'autre moitié par les bases de données françaises CLAPI et ESLO<sup>3</sup>. La partie française de ce corpus, exploitée ici, compte 100 minutes d'enregistrements transcrits avec alignement au son, exploités ici avec le logiciel Praat (Boersma & Weenink 1999-), et représentent un vaste panel de situations comprenant : conversation téléphonique (PC, *Phone Call*), préparation de repas (PM, *Preparing Meal*), interaction commerciale (SE, *Service Encounter*), réunion (SM, *Social Meeting*), discussion de table (TT, *Table Talk*), conférence (Conf), entretien (Ent), lecture (LE, Livre Enfant), débat radiophonique (Media) et table ronde (TR).

Nom	PC	PM	SE	SM	TT	Conf	Ent	LE	Media	TR	Total
Mots	2104	986	2139	986	2283	3873	2356	1631	2334	2202	20894
Durée	577	658	540	605	618	1230	681	569	602	682	6762

Tableau 3. Détail du corpus

Dans le tableau 3, le nombre de mots correspond aux *tokens* et la durée est en secondes. Le nom est celui des transcriptions, abrégé. Nous avons exploité 20 minutes de conférence au lieu de 10 minutes, déséquilibre qui n'existe pas pour les autres équipes du projet SegCor.

Par la suite, nos exemples donneront le nom de la transcription suivi des bornes temporelles du passage en secondes. Le paraverbal est indiqué entre parenthèses, dont les pauses en secondes, tandis que les crochets « [] » signaleront les chevauchements.

Les deux niveaux d'annotation que nous décrivons sont ceux de la pragma-syntaxe (macro-syntaxe fribourgeoise) et de l'interaction (linguistique interactionnelle). Nous nous en tiendrons, pour la description de ces niveaux, à ce qui est essentiel et pertinent pour le présent travail.

## 2.1. Pragma-syntaxe

Pour la pragma-syntaxe, les unités retenues sont la clause-énonciation et la période.

---

<sup>3</sup> Voir la sitographie.

La « clause-énonciation » est une unité ambiguë issue du travail d’annotation réalisé dans le cadre du projet SegCor. L’énonciation est l’unité minimale de pragma-syntaxe, et correspond à une opération sur une mémoire discursive (M), laquelle représente un état du discours à un moment donné. Chaque énonciation modifie M et crée un nouvel état. Néanmoins, du point de vue de l’annotation (Groupe de Fribourg 2012 : 43) :

---

En pratique, identifier les énonciations se ramène alors à une tâche de segmentation de chaînes : on doit être capable de délimiter dans un continuum textuel chacune des clauses successivement actualisées.

---

La clause est l’unité maximale de la morpho-syntaxe<sup>4</sup>, c’est-à-dire l’unité au-delà de laquelle il n’y a plus de rapport de rection. L’annotation des énonciations crée alors une confusion entre clause et énonciation, d’où la création d’une unité hybride que nous appellerons clause-énonciation. Par exemple :

- |     |   |   |  |
|-----|---|---|--|
| (1) | cette semaine elle<br>nous a appelés (0.3)<br><b>pour nous dire tu<br/>sais que</b> | normalement quand tu<br>fais une formation<br>comme ça tu es prise<br>automatiquement [...] | et du coup (0.6) <b>là</b><br><b>ANON ils vont pas</b><br><b>en garder</b><br>(TableTalk, 497-516) |
| a.  |   | clause1   | clause2  |
| b.  | énonciation1  | énonciation2  | énonciation3   |
| c.  | clause-énonciation1   | clause-énonciation2   | clause-énonciation3  |

En (1) l’extrait, réparti sur trois colonnes, introduit un discours indirect (première colonne), préfacé à son introduction par une forme de parenthèse (deuxième colonne) avant le discours rapporté lui-même (troisième colonne). Les deux premières parties sont régies et forment donc une seule clause (1c), mais la préface communique tout autre chose et constitue donc une énonciation à part (1b). Rection et actualisation entrent ici en conflit.

Par clause-énonciation nous entendons un segment textuel continu au-delà duquel, dans son actualisation, ne se trouve plus de rapport de rection. La clause-énonciation peut être complète ou inachevée mais ne peut pas, suivant le protocole d’annotation, être discontinuée. Autrement dit, en (1c) la frontière d’unité est avérée dès que l’un des deux critères est rempli, rection ou actualisation, est rempli. La validité de cette unité vis-à-vis du modèle fribourgeois dépasse l’objet de notre présent travail : nous arguons néanmoins qu’elle est en l’état l’unité d’annotation la plus efficace pour correspondre au modèle.

La période, définie comme l’unité maximale de la pragma-syntaxe, peut avoir trois critères définitoires (Avanzi 2005 : 3) :

---

<sup>4</sup> Pour une discussion sur la phrase et la différence avec la clause, nous reportons le lecteur à l’article de l’Encyclogram en ligne (Berrendonner 2017).

- a. Un programme communicatif complet<sup>5</sup>
- b. Un intonème conclusif
- c. Une fonction interactive<sup>6</sup>

Le protocole d'annotation retient les second et troisième critères, et nous devons souligner ici la nature interactive de cette unité dans le modèle (Berrendonner 2003 ; Groupe de Fribourg 2012) : la fin d'une période correspond théoriquement à un lieu de transition pertinent du tour de parole. Cela signifie inversement que tout changement du tour de parole implique nécessairement une frontière de période.

Par période, nous entendons un segment textuel continu s'achevant sur un intonème conclusif. Un intonème est un contour intonatif en fin d'énonciation et porteur d'une fonction, notamment soit continuative, soit conclusive. Une période ne s'achevant pas sur un intonème conclusif est considérée inachevée. Quant aux critères prosodiques de l'intonème conclusif, nous ne chercherons pas à les décrire ici (Lacheret-Dujour & Victorri 2002 ; Lacheret, Avanzi & Victorri 2010 ; Avanzi 2010). Le modèle fribourgeois les décrit ainsi (Groupe de Fribourg 2012 : 97) :

---

La marque par défaut de fin de période, F, se présente comme un complexe associant : (i) une descente progressive de la F0 au niveau infra-grave, à partir d'un sommet de hauteur placé sur une syllabe antérieure à la pénultième ; (ii) une baisse d'intensité sur la syllabe finale ; (iii) une pause subséquente.

---

Il s'agit bien là d'un contour par défaut, non d'un contour unique et exclusif.

À titre d'exemple :

- (2) ELI    a. j'ai mis le chauffage  
           b. mais je crois qu'il fait un peu chaud  
           c. mais bon (PhoneCall, 81-84)

---

<sup>5</sup> La mémoire discursive M peut avoir un état dit instable, nécessitant des opérations supplémentaires. Un programme communicatif complet est un ensemble d'opérations entre deux états stables.

<sup>6</sup> Cette fonction serait de signaler à sa fin un lieu de transition pertinent. Essentiellement, ce troisième critère est à l'origine de notre travail.

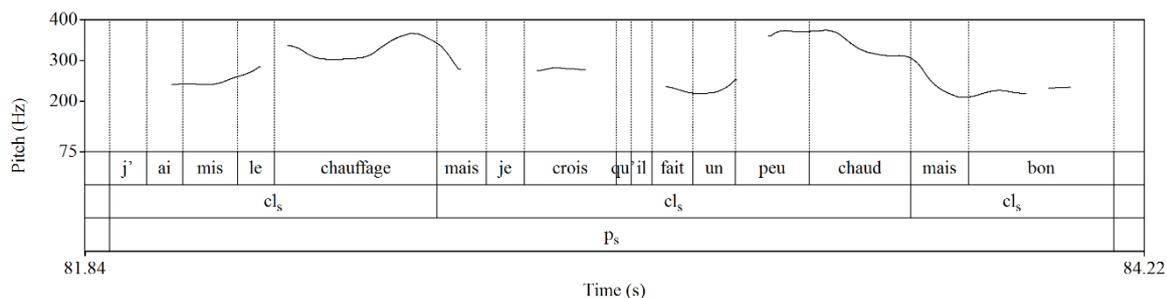


Figure 1. Prosodie de l'exemple (1)

En (2), les trois clauses (a-c) forment une période complète. Chaque clause a un intonème visible avec montée de l'intonation dans la figure 1 : la dernière clause représente une chute relative, ce contour signalant la conclusion.

## 2.2. Interaction

Suivant Kerbrat-Orecchioni (2005 : 14), nous rassemblons sous le terme d'interaction un objet étudié en contexte interactif aussi bien par l'analyse du discours que par l'analyse conversationnelle. La linguistique interactionnelle (Mondada 1998) semble partager la même volonté de regroupement. L'important est qu'en parlant d'interaction, nous ne nous cantonnons pas spécifiquement à l'analyse conversationnelle (Sacks *et al.* 1974), quand bien même pour l'essentiel nous en suivrons les méthodes et travaux (Schegloff 1996 ; Selting 2000 ; Auer 2010).

Pour l'interaction, les unités retenues sont l'action et tout type d'unité délimitée par le TRP (*Transition Relevant Place*).

L'action est inspirée des actes illocutoires d'Austin (1970) et Searle (1972), c'est-à-dire des tâches, des activités réalisées par le discours, indépendamment des conséquences réelles qui elles relèvent de l'acte perlocutoire (Austin 1970 : 114). L'acte illocutoire correspond au fait de dire, demander, ordonner, supplier, remercier, etc. Il se limite au cadre de la « phrase ». L'action y ajoute un contexte interactif et constitue donc un « acte conversationnel » (Kerbrat-Orecchioni 2008 : 54) incluant des commentaires, continueurs, corrections, phatiques, etc. L'action est la tâche, l'activité accomplie durant le tour de parole, ainsi que l'acte conversationnel pertinent pour la gestion de ce tour (Ford 2001 : 56). A titre d'illustration (exemple tiré de Traverso 2016 : 72) :

- (3) A     quelle heure est-il ?  
           (silence)  
       A     hein c'est quelle heure ?

Chaque ligne en (3) représente une action. La première, une requête, projette l'attente d'une réponse. La seconde, paraverbale, est un silence transcrit entre parenthèses qui pousse A à réaliser une nouvelle action. Les actions s'enchaînent sous forme de paires adjacentes et permettent d'estimer le degré de complétion du tour de parole.

L'autre unité retenue devrait être la TCU (*Turn-Constructional Unit*) ou unité de construction du tour. Il s'agit de l'unité minimale en analyse conversationnelle (Sacks *et al.* 1974) correspondant à une action et au terme de laquelle il devient pertinent de gérer le changement de tour de parole. Toute TCU a donc théoriquement pour frontière un TRP, ou lieu de transition pertinent, où le tour de parole est réattribué. En pratique, Selting (2000) remarque que toutes les TCUs n'ont pas nécessairement de TRP : ce dernier peut être neutralisé par des indices prosodiques notamment. Un tour de parole peut être constitué de plusieurs TCUs s'il y a une telle neutralisation, ou si le tour est réattribué au même locuteur. Là encore, pour illustrer (Traverso 2016 : 45) :

- (4) DOC bon | alors voyons voir | i bouge bien vot' bébé  
PAT ouais | (pause) enfin pas tous les jours hein

Le tour de DOC en (4) est formé de trois unités distinctes que nous signalons par des « | ». Chacune de ces unités est réputée constituer une TCU, bien que les deux premières n'offrent pas, par la prosodie entre autres, un lieu où le tour de parole peut changer de locuteur (TRP). A l'inverse, chez PAT la frontière après « ouais » constitue bien un lieu de transition pertinent, où DOC pourrait reprendre la parole.

Les tentatives d'annoter les TCUs existent (Ford & Thompson 1996 ; Selting 1996) mais n'ont pas permis d'établir de protocole, principalement à cause de la nature de cette unité : il s'agit d'une unité co-construite et révisable par les interlocuteurs durant la production. L'unité annotée, elle, tend à être finie et ne rend pas compte des révisions. Face à cette difficulté théorique, le projet SegCor a préféré se concentrer sur la frontière d'unité, c'est-à-dire le TRP, sans chercher à déterminer ce que cette frontière délimitait comme unité.

Les annotations expérimentales du projet SegCor se sont concentrées sur le faisceau d'indices syntaxiques, prosodiques et pragmatiques pour le repérage des TRPs. Il s'agit, essentiellement, de :

1. la complétion syntaxique,
2. d'indices prosodiques (débit, contour, pause, etc),
3. et des unités d'action tel que définies plus haut.

Nous mentionnerons les annotations du projet SegCor pour les actions, tandis que pour les TRPs nous avons, pour les besoins de ce travail, réalisé notre propre annotation adaptée du protocole du projet.

La comparaison de l'annotation des périodes avec celle en interaction nous a fait émettre l'hypothèse que la période pouvait correspondre à certains types de TRPs, ce que vise à démontrer le présent travail. Un précédent travail d'annotation a permis de cerner plus précisément les types de TRPs qui seront considérés ici pertinents pour cette comparaison.

Nous conservons autrement le protocole SegCor pour le repérage des TRPs, en partageant tant la nature expérimentale de ce protocole que les limitations du corpus : ce dernier ne fournit que des données sonores, alors que le TRP demanderait une prise en compte de données multimodales (regards, gestes, environnement, etc). De même, il faut rappeler l'importance des réactions des interlocuteurs qui participent à l'identification des TRPs.

Pour les types de TRPs, nous nous appuyons sur notre premier travail d'annotation pour proposer la typologie suivante, détaillée dessous :

TRP	Type	Description
0	Neutralisation	TRPs neutralisés par la prosodie
1	Hésitation	Cas d'auto-sélection avec allongement, « euh », etc
2	Interruption	Cas d'auto-sélection avec inachèvement de l'unité
3	Continueur	Pas de prise de parole (mh, oui, d'accord, etc)
4	Expansion	Continueur suivi d'une prise de parole effective
5	Ouverture	Abandon du tour, avec ou sans sélection

Tableau 4. Types de TRPs

La « neutralisation » correspond au cas de Selting (2000) : un TRP identifié par la syntaxe et l'action, mais neutralisé par la prosodie. Le « continueur » correspond à des accusés de réception et continueurs de type « mh mh, oui, d'accord » qui ne constituent pas des prises du tour (de parole). Parler de TRP ici est du reste ambigu. Enfin, « l'ouverture » correspond aux cas d'abandon du tour, qu'il y ait ou non sélection du prochain locuteur. Cela peut couvrir le cas d'auto-sélection (c'est-à-dire où le locuteur se sélectionne lui-même à la transition du tour) d'un point de vue pragmatique (l'action projetée une suite), si d'autre part la syntaxe et surtout la prosodie suggère un potentiel abandon.

À ces trois types s'en ajoutent trois autres. « Hésitation » et « interruption » sont des sous-types d'auto-sélection. L'hésitation autorise des collaborations (au moins pour la recherche lexicale) mais sans abandon du tour au-delà, et avec généralement un maintien du tour en pratique. L'interruption est soit due à une tentative de prise de parole d'un interlocuteur, soit due à un changement complet d'action du locuteur. Quant à « l'expansion », il s'agit d'un sous-type de continueur exploité par le participant pour prendre la parole. Parler de continueur à ce titre est ambigu.

À titre d'exemple (pour rappel, les parenthèses indiquent le paraverbal et les pauses en secondes, et les crochets signalent les chevauchements) :

- (5) ELI a. c'est eux qui paient la formation (0.2)  
       b. [ils ont plutôt intérêt de te]  
    MAR a. [(inaudible)]  
    ELI a. voilà  
       b. ils ont plutôt intérêt de l'embaucher à la fin (PhoneCall, 507-512)

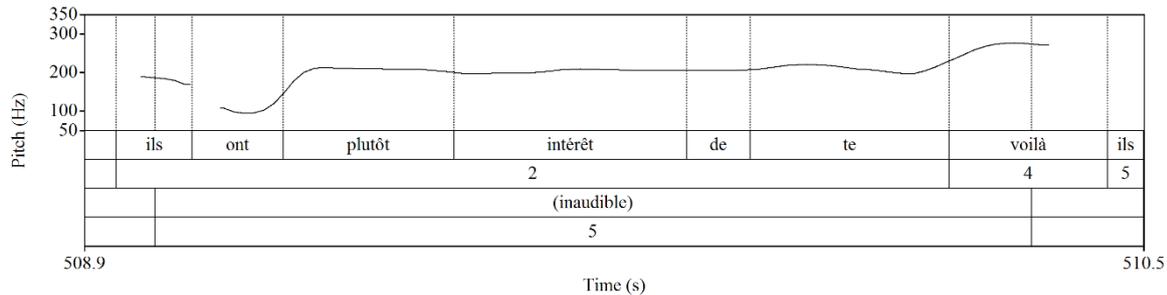


Figure 2. Prosodie de l'exemple (5)

Le participant MAR tente une prise du tour. Le chevauchement crée une interruption : le locuteur, bien que n'ayant pas abandonné son tour, réagit au tour de MAR et agit donc comme s'il y avait eu alternance. Le « voilà » est traité comme un continueur permettant à ELI de reprendre son action inachevée.

### 3. Comparaison des frontières

Nous devons encore présenter les résultats des annotations obtenues avant d'en comparer et discuter les frontières. Notamment, nous avons parlé d'annotations expérimentales : les protocoles d'annotation exploités n'ont pas fait l'objet de mesures de fiabilité (accord inter-annotateurs) et annotent des objets (unités et frontières) dont le traitement est sujet à débat dans leurs domaines respectifs.

Nous ferons donc les remarques qui s'imposent sur les annotations avant d'en présenter les résultats, puis de discuter la comparaison de leurs frontières.

#### 3.1. Remarques sur les annotations

L'annotation pragma-syntaxique a été menée par un seul annotateur sur l'ensemble du corpus présenté dans le tableau 3. Clause-énonciation et période ont été annotées séparément. Nous devons insister sur le caractère hybride de la clause-énonciation : l'objectif était que les frontières correspondent au mieux aux lieux d'actualisation de la mémoire discursive. L'annotation est donc expérimentale à deux égards : d'une part pour l'identification des rapports de rection, d'autre part pour le repérage des lieux d'actualisation. Nous avons également révisé les frontières périodiques pour *PhoneCall* et *TableTalk* dans le cadre de ce travail.

Pour l'interaction, une première annotation a été menée dans le cadre du projet SegCor par trois annotateurs sur quatre enregistrements, deux français et deux allemands. Nous y référerons seulement pour les actions. L'annotation des TRPs présentée ici a été menée par un seul annotateur et s'inspire de leur protocole et travail d'annotation en l'adaptant au niveau des types de TRPs. Cette annotation couvre deux transcriptions (*PhoneCall* et *TableTalk*).

### 3.2. Résultats des annotations

Pour la pragma-syntaxe (MSF, Macro-Syntaxe Fribourgeoise), nous donnons ci-dessous (tableau 5) le nombre de clause-énonciations et de périodes obtenues par enregistrement :

MSF	PC	PM	SE	SM	TT	Conf	Ent	LE	Media	TR	Total
Clause	443	242	303	405	492	343	334	264	264	224	3314
Période	266	177	208	228	321	109	102	195	116	79	1801

Tableau 5. Résultat de l'annotation en pragma-syntaxe

Un résultat en faveur de la nature interactive des périodes vient des périodes inachevées, dont les causes sont essentiellement interactives. Sur 1801 périodes, 299 sont inachevées avec trois causes : soit (a) le locuteur a été interrompu par une tentative de prise de parole, soit (b) il a lui-même interrompu sa tentative de prise de parole, soit (c) le locuteur a cherché à achever son tour.

- (6) ELI *et le vendredi normalement j'ai une fois sur deux tu sais int- euh*  
 BEA *analyse conversationnelle* (TableTalk, 232-238)

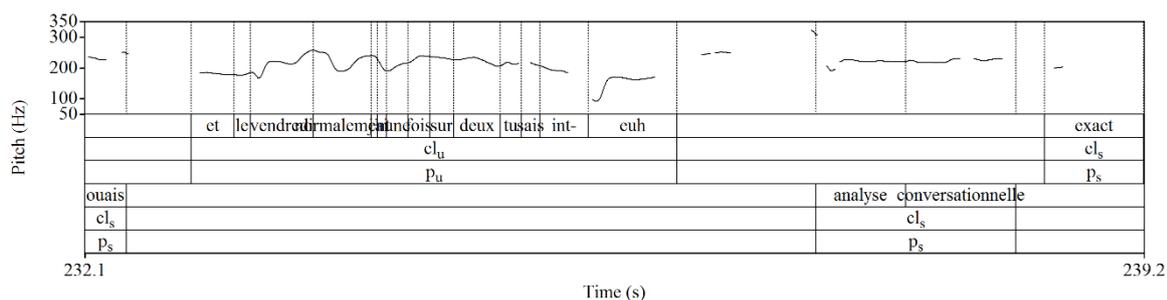


Figure 3. Annotation de l'exemple (6)

- (7) spk1 *euh ça ça vous est venu comment c'est un (0.3) c'est un rêve d'enfant ou c'est par coïncidence ou c'est*  
 spk2 *ouais ouais non non moi depuis tout petit hein voilà euh*  
 (Ent, 115-120)

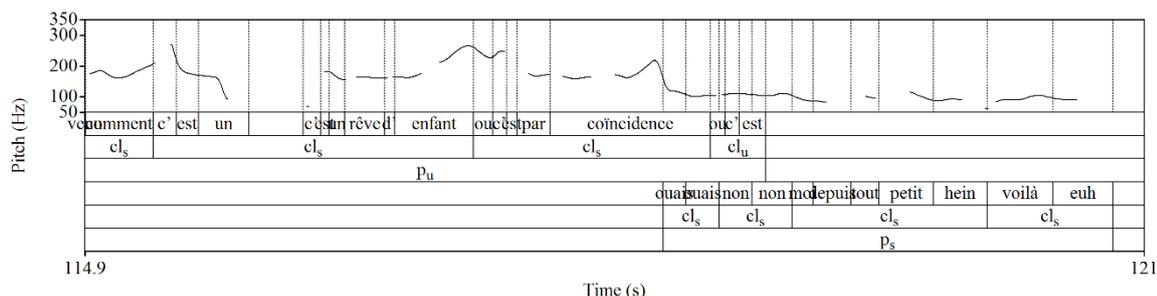


Figure 4. Annotation de l'exemple (7)

Dans les figures 3 et 4, les clause-énonciations sont annotées « cl » et les périodes « p », suivies d'un « s » pour les unités complètes et « u » pour celles inachevées. Nous mettons en italique dans les exemples les passages illustrés par les figures. L'exemple (6) illustre un cas où le locuteur laisse brièvement son tour (cas (c)) pour une recherche lexicale : seul le changement de tour qui suit, avec BEA, justifie cette frontière : du point de vue prosodique, il n'y a pas ici d'intonème conclusif. L'exemple (7) est un cas classique (a) d'interruption par un interlocuteur qui ne demande pas de commentaire.

Les résultats pour l'interaction, pour rappel, couvrent deux enregistrements (*PhoneCall* et *TableTalk*) et sont le fait d'un seul annotateur :

TRP	Neutr.	Hésitation	Interrupt.	Contin.	Expans.	Ouverture	Total
PhoneCall	178	52	46	54	29	172	531
TableTalk	137	59	26	74	16	225	537
Total	315	111	72	128	45	397	1068

Tableau 6. Segmentation des TRPs

Grâce à un travail préliminaire, les nombres du tableau 6 sont relativement stables, notamment parce qu'ils écartent les types de TRPs problématiques, telle la « demande de continuer »<sup>7</sup> ou l'auto-sélection, laquelle, selon la perspective adoptée, peut se retrouver aussi bien dans les TRPs neutralisés que dans l'abandon de tour.

Même en écartant ces types, leur problématique demeure :

- (8) ANA *dis je voulais te demander (0.2) demain vu que on a qu'une heure pour manger euh ça te dit que on aille manger dehors*  
(PhoneCall, 6-13)

<sup>7</sup> Ce type « TRP+ », venant du protocole SegCor, correspond à un TRP n'autorisant pas la prise du tour, mais signalant aux interlocuteurs de produire des continuateurs.



et	tout	tu	sais	parc	q	ell	en	a	marre	arrivée	à	vingt	quatre	ans	tu	te	dis	euh	_	ELI[ort]
			++																/	ELI[act]
			explication																/	ELI[typ]
			cl_s																	ELI[cla]
																				ELI[per]

Figure 6. Annotation de l'exemple (9)

Dans le cas de (9), l'action d'explication est suivie d'une justification en deux temps, visible par les deux clauses (au sens de la pragma-syntaxe) du tableau 6. Le type d'action n'est pas annoté du fait de l'inachèvement (noté « / ») mais nous pouvons noter que l'action elle-même chevauche les deux clauses. Au niveau de la pragma-syntaxe, les deux clauses correspondent à la même routine communicative.

Cet exemple (9) nous permet aussi de revenir sur les périodes inachevées. La frontière de période inachevée n'est pas motivée par un changement effectif de tour de parole. Il serait possible de ranger cet exemple dans l'abandon de tour (c) similaire à (5), avec l'incrément (Schegloff 1996) « c'est c'est compréhensif quand tu » lorsque les interlocuteurs ne collaborent pas (Lerner 1991). Néanmoins, du point de vue du protocole, l'exemple (9) illustre un inachèvement fondé sur le critère prosodique, non celui interactif. La cause est une rupture prosodique (Delafontaine *et al.* 2018) révisant ce lieu comme une frontière. Nous fermons là cet aparté.

En l'état, la comparaison des clauses-énonciations avec les actions interactionnelles suggèrent qu'aucune clause-énonciation ne peut contenir plus d'une action. Nous proposons ici que les cas où une action contient plusieurs clauses-énonciations est dû à un choix d'annotation qui fusionne des actions d'un même type ; nous arguons qu'il y a bien là plusieurs actions interactionnelles et non une seule.

Le choix d'annotation est motivé par les attentes vis-à-vis de l'action interactionnelle. Sa fonction attendue est de former des paires adjacentes pour la gestion interne du tour de parole (voir 2.2), et deux actions d'un même type, projetant la même attente, n'offrent pas de pertinence de ce point de vue. Mais si nous nous intéressons au repérage des TRPs, la frontière entre deux actions d'un même type devient pertinente. Dans l'exemple (4) de Traverso (2016), « bon » et « alors voyons voir » constituent des actions projetant essentiellement la même attente. La neutralisation des TRPs, en (4) comme en (9), n'empêche pas d'y parler d'unités constitutives du tour de parole.

Nous soutenons donc ici que la clause-énonciation correspond bien à l'action interactionnelle. Néanmoins, l'action interactionnelle considérée perd en pertinence pour la formation de paires adjacentes et nous pouvons donc questionner s'il s'agit bien d'une unité pertinente pour le domaine de l'interaction ; ou, dans le modèle fribourgeois, d'un bon candidat comme unité minimale d'une quatrième articulation.

### 3.3.2. TRP et période

Notre annotation compte 1068 TRPs pour 587 périodes. Nous pouvons comparer ces frontières dans le tableau suivant :

TRP	0	1	2	3	4	5	Total
PhoneCall	178	52	46	54	29	172	531
p_s	3			<b>50</b>		<b>172</b>	225
p_u		3	<b>38</b>				41
TableTalk	137	59	26	74	16	225	537
p_s				<b>73</b>		<b>225</b>	298
p_u			<b>23</b>				23

Tableau 7. Comparaison des frontières

Les lignes *PhoneCall* et *TableTalk* répètent les nombres de TRPs pour chaque transcription, suivi des frontières périodiques correspondantes (complètes « p\_s » et inachevées « p\_u ») pour ces TRPs. Nous mettons les correspondances en gras. A toute période correspond un TRP, principalement des continueurs (type 3) et des ouvertures (type 5) pour les périodes complètes et principalement des interruptions (type 2) pour les périodes inachevées. Inversement, les TRPs constituant une neutralisation, une hésitation ou une expansion (types 0, 1 et 4) tendent à ne jamais correspondre à une période.

La quasi-équivalence qui se dégage appelle à la prudence : elle peut être due aux protocoles d'annotation. Le protocole pour la période prévoit explicitement une frontière périodique en cas de changement effectif de tour, ce qui a été étendu aux cas où le locuteur, indépendamment des interlocuteurs, signale un abandon de tour. La correspondance entre le type « ouverture » (type 5) ou le type « interruption » (type 2) était donc attendu.

Si nous résumons cette comparaison : (a) toute frontière périodique correspond à un TRP ; (b) quasiment toute période inachevée correspond à un TRP d'interruption ; (c) tout TRP d'ouverture et quasiment tout TRP continueur correspond à une frontière périodique complète ; (d) quasiment aucun TRP d'autres types ne correspond à une frontière périodique.

Un cas de divergence peut être directement expliqué : il s'agit des TRPs de type « continueur » sans frontière périodique. Cela est dû à un choix d'annotation traitant une série de continueurs comme des unités séparées, mais comme une seule période : il n'y a pas de réelle divergence en pratique.

Les divergences sur les types 0 et 1 sont dues soit à la difficulté de catégoriser les TRPs, soit à une sur-annotation des frontières périodiques :

- (10) MEL *mais euh* (0.6)  
 NAT *y a pas de [souci euh]*  
 MEL *[mais euh] donc* (0.2) *[ou alors] sinon on peut aller dans un petit resto*  
 NAT *[ouais]* (PhoneCall, 127-132)

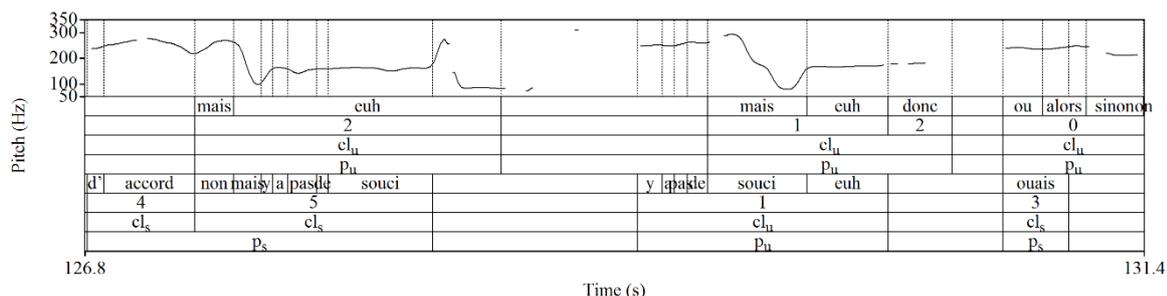


Figure 7. Annotation de l'exemple (8)

L'exemple (8) illustre le cas où une frontière de période inachevée, « y a pas de souci euh », correspond à un TRP de type « hésitation ». Ce TRP aurait aussi bien pu être annoté en type 2, du fait du chevauchement, puis du continueur signalant la prise en compte du tour de MEL. Néanmoins le « euh » d'hésitation a forcé une autre interprétation, créant une situation où l'hésitation correspond à un abandon de tour.

- (11) JUL et que je regarde surtout le prix des billets de train *parce que c'est ça qui va* (0.8) *qui va jouer* (0.4) *ben [là je suis au] boulot donc je peux pas trop regarder maintenant*  
 CAT [c'est sûr] (PhoneCall, 41-49)

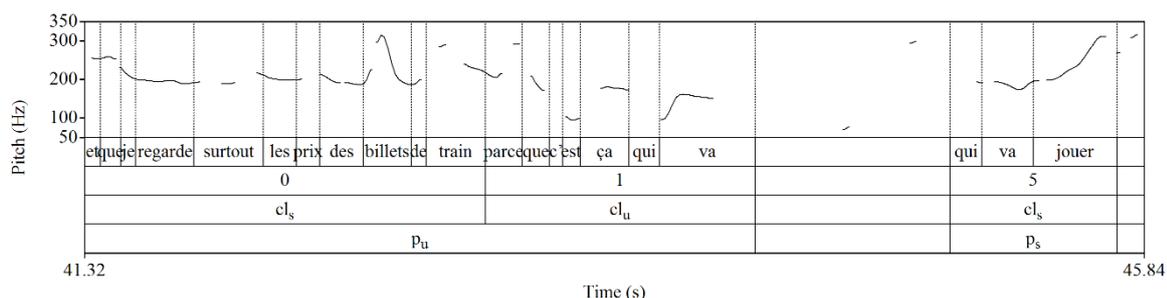


Figure 8. Annotation de l'exemple (9)

En (9), c'est au contraire, à notre avis, la frontière périodique qui est en cause. Il n'y a pas ici d'intonème conclusif en soi, mais le protocole prévoit de considérer une frontière de clause-énonciation à partir de 0.8s, où la pause devient un geste et donc une opération sur M (paraverbal, donc échappant à notre annotation). L'interruption de la période considère donc un intonème sous forme de rupture prosodique, mais cela est tout à fait contestable et nous pouvons considérer l'annotation comme fautive, la période continuant jusqu'à « qui va jouer ».

Quant au cas de TRP de type 0 correspondant à une frontière périodique, il peut être illustré par l'exemple (6).

- (6) ANA *dis je voulais te demander* (0.2) *demain vu que on a qu'une heure pour manger euh ça te dit que on aille manger dehors*  
 (PhoneCall, 6-13)

Les TRPs après « dis » et « je voulais te demander » ont été annotés en type 0, comme TRPs neutralisés, du fait de la proximité des contours prosodiques avec les frontières suivantes, lesquelles ne sont pas considérées ambiguës. Mais l'intonème à la fin de « je

voulais te demander » est ambigu : il a été, dans l’annotation actuelle, annoté comme conclusif, ce qui pourrait comme en (9) être abusif.

Cette difficulté à repérer l’intonème conclusif peut être étendue à quelques frontières périodiques liées aux TRPs de type 5. Faute d’annotateurs supplémentaires et de par la nature expérimentale du protocole, nous maintenons l’appel à la prudence : la correspondance n’est sans doute pas aussi systématique que le tableau 7 le suggère.

Nous donnerons un dernier exemple avant d’en venir aux perspectives :

- (12) ELI mardi j’ai pas cours (0.8) le mercredi j’ai euh (0.5) *qu’est-ce que j’ai j’ai syntaxe bon comment j’ai pu oublier ça* (0.4) à une heure du matin trois heures tu sais (1.2) le jeudi j’ai juste euh psycho (TableTalk, 219-231)

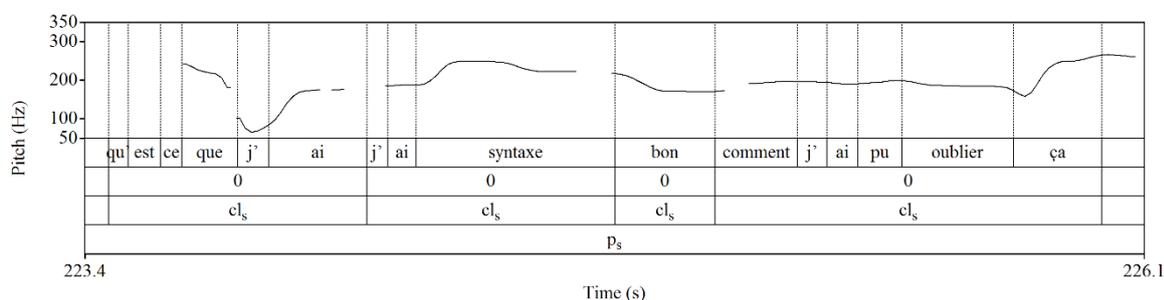


Figure 9. Annotation de l'exemple (10)

Le passage illustré par la figure 9 est une parenthèse formée d’une question, puis d’une réponse à soi-même immédiatement étendue par une remarque.

Le contour de « qu’est-ce que j’ai » correspondrait à un intonème conclusif, de même que celui de « comment j’ai pu oublier ça ». La montée prosodique en fin de remarque est causée par du bruit de fond, non par le locuteur ELI. Cela signifie qu’il y a ici potentiellement plusieurs périodes qui n’ont pas été annotées car constituant une parenthèse au sein d’une période plus large. Il pourrait s’agir d’un cas de sous-annotation.

Du point de vue des TRPs, toutes ont été considérées comme neutralisées malgré la variété d’indices, de par la nature là encore de la parenthèse. Il est difficile d’envisager par exemple que la question puisse être de type 5 d’un point de vue pragmatique. La correspondance est dû au fait que la parenthèse est un phénomène problématique pour les deux niveaux d’analyse : un changement de traitement affecterait également les deux niveaux, et maintiendrait la correspondance.

## Perspectives

Les résultats obtenus par la comparaison des unités de la pragma-syntaxe et de l’interaction sont encourageants. Malgré le caractère expérimental des annotations et la faible quantité de données annotées, il est possible d’observer un rapport entre clause-énonciation et action, et surtout entre frontière de période et TRP.

Nous avons exploité ici notre propre annotation en TRPs, à la suite d'autres travaux préliminaires permettant déjà d'observer certaines correspondances entre types de TRPs et frontières de période, sans que ces travaux préliminaires n'aient pu établir de systématicité. Le présent travail affine suffisamment la typologie des TRPs pour repérer celles correspondant aux frontières périodiques du modèle fribourgeois. Cela ne permet pas d'établir un rapport entre période et TCU cependant : l'annotation ne dit rien de l'unité, s'il y en a une, délimitée par le TRP.

Les résultats obtenus suggèrent qu'à la frontière périodique correspond systématiquement : pour une période inachevée, un TRP de type interrompu ; pour une période complète, un TRP de type continueur ou ouverture (abandon du tour). Inversement, aucun TRP visant au maintien du tour ne devrait correspondre à une frontière de période.

Les divergences observées sont dues aux difficultés d'annotation : soit pour distinguer les types de TRPs, soit pour identifier les intonèmes conclusifs, soit dû à des phénomènes de l'oral, tel la parenthèse. Des protocoles mieux établis et des annotations concurrentes devraient permettre de confirmer et/ou infirmer ces résultats.

Il nous est néanmoins possible de confirmer la nature interactive de la période pragma-syntaxique mais aussi, en allant plus loin, de proposer que, au niveau de l'interaction, les types de TRPs correspondant à des frontières périodiques pourraient aussi correspondre à des frontières de TCUs. La question du traitement des incréments, et celle de la nature même de la TCU, se posent évidemment. Mais les annotations réalisées tendent à confirmer notre hypothèse de départ : il y aurait bien une unité minimale en interaction correspondant à l'unité maximale de la pragma-syntaxe dans le modèle fribourgeois<sup>8</sup>.

## Références bibliographiques

- Auchlin, Antoine & Zezone, Anna. 1980. Conversations, actions, actes de langage : éléments d'un système d'analyse. *Cahiers de linguistique française* 1, 6-41.
- Auer, Peter. 2010. Zum Segmentierungsproblem in der Gesprochenen Sprache. *InList [En Ligne]* 46, <[inlist.uni-bayreuth.de/issues/49/InList49.pdf](http://inlist.uni-bayreuth.de/issues/49/InList49.pdf)>.
- Austin, John Langshaw. 1970. *Quand dire c'est faire - How to do things with words*. Paris : Seuil.

---

<sup>8</sup> Quant à l'unité maximale de l'interaction (ou de la grammaire du discours), elle s'étend classiquement au discours entier en tant qu'activité : l'incursion pour le modèle de Genève (Auchlin *et al.* 1980), la conversation pour l'analyse conversationnelle (Sacks *et al.* 1974). Cette unité ne concerne pas le modèle fribourgeois.

- Avanzi, Mathieu. 2005. Quelques hypothèses à propos de la structuration interne des périodes. In C. Auran et al. (dir.). *Actes du symposium interface discours-prosodie*, Aix-en-Provence, 2005, CD-ROM.
- Avanzi, Mathieu. 2010. Rattachement et fragmentation de la syntaxe par la prosodie. *Travaux de linguistique* 60, 145-166.
- Benzitoun, Christophe, Dister, Anne, Gerdes, Kim, Kahane, Sylvain, Pietrandrea, Paola & Sabio, Frédéric. 2010. Tu veux couper là faut dire pourquoi – Propositions pour une segmentation syntaxique du français parlé. *CMLF 2010*, 2075-2090.
- Berrendonner, Alain & Béguelin, Marie-José. 1989. Décalages : les niveaux de l'analyse linguistique. *Langue française* 81, 99-125.
- Berrendonner, Alain. 1993. Périodes. In Herman Parret (éd.). *La temporalité du discours*. Louvain : Presses Universitaires de Louvain, 47-61.
- Berrendonner, Alain. 2012. Autour de la rection. In Sandrine Caddéo, Marie-Noëlle Roubaud, Magalie Rouquier & Frédéric Sabio (éds.). *Penser les langues avec Claire Blanche-Benveniste*. Aix-en-Provence : Presses Universitaires de Provence, 83-91.
- Berrendonner, Alain. 2017. La notion de phrase. In *Encyclopédie grammaticale du français*, <encyclogram.fr>.
- Blanche-Benveniste, Claire, Deulofeu, José, Stefanini, Jen & Eynde, Karen van den. 1984. *Pronom et syntaxe. L'approche pronominale et son application au français*. Paris : Société d'Études Linguistiques et Anthropologiques de France.
- Blanche-Benveniste, Claire, Bilger Mireille, Rouget, Christine & Eynde, Karen van den. 1990 (2005). *Le français parlé. Etudes grammaticales*. Paris : CNRS.
- Boersma, Paul & Weenink, David. 1999-. *Praat: doing phonetics by computer [Computer program]*. <[praat.org](http://praat.org)>.
- Cresti, Emanuela. 1999. Force illocutoire, articulation topic/comment et contour prosodique en italien parlé. *Faits de langue* 13, 168-181.
- Cresti, Emanuela. 2000. Critère illocutoire et articulation informative. In Mireille Bilger (éd.). *Corpus. Méthodologie et applications linguistiques*. Paris : Champion, 350-367.
- Cresti, Emanuela, Moneglia, Massimo & Tucci, Ida. 2011. Annotation de l'entretien d'Annita Musso selon la Théorie de la langue en acte. *Langue française* 170, 95-110.
- Delafontaine, François, Ursi, Biagio & Acosta, Luisa. 2018. Annotation des proéminences pour la segmentation de corpus oraux : l'expérience du projet SegCor. *CMLF 2018*, <[shs-conferences.org/articles/shsconf/pdf/2018/07/shsconf\\_cmlf2018\\_11001.pdf](http://shs-conferences.org/articles/shsconf/pdf/2018/07/shsconf_cmlf2018_11001.pdf)>
- Deulofeu, José. 2013. L'approche macro-syntaxique : sources et controverses. In Jeanne-Marie Debaisieux (éd.). *Analyses linguistiques sur corpus : subordination et insubordination en français*. Paris : Lavoisier, 427-497.
- Ford, Cecilia, Fox, Barbara & Thompson, Sandra. 1996. Practices in the construction of turns: the 'TCU' revisited. *Pragmatics* 6 (3), 427-454.

- Ford, Cecilia. 2001. At the intersection of turn and sequence: Negation and what comes next. In Elizabeth Couper-Kuhlen & Margaret Selting (éds.). *Studies in interactional linguistics*. Amsterdam : Benjamins, 51-79.
- Groupe de Fribourg. 2012. *Grammaire de la période*. Berne : Peter Lang.
- Kerbrat-Orecchioni, Catherine. 2005. *Le discours en interaction*. Paris : Armand Colin.
- Kerbrat-Orecchioni, Catherine. 2008. *Les actes de langage dans le discours : théories et fonctionnement*. Paris : Armand Colin.
- Lacheret-Dujour, Anne & Victorri, Bernard. 2002. La période intonative comme unité d'analyse pour l'étude du français parlé : modélisation prosodique et enjeux linguistique. *Verbum* XXIV, 55-72.
- Lacheret, Anne, Avanzi, Mathieu & Victorri, Bernard. 2010. A corpus-based learning method for prominence detection in spontaneous speech. *Speech prosody 2010*. Chicago, 20-30.
- Lerner, Gene H.. 1991. On the syntax of sentences-in-progress. *Language in society* 20, 441-458.
- Martinet, André. 1960. *Éléments de linguistique générale*. Paris : Armand Colin.
- Mondada, Lorenza. 1998. Pour une linguistique interactionnelle. *ARBA* 8, 113-130.
- Mondada, Lorenza. 2000. Analyse conversationnelle et grammaire-pour-l'interaction. In Anne-Claude Berthoud & Lorenza Mondada (éds.). *Modèles du discours en confrontation*. Berne : Peter Lang, 23-42.
- Pekarek Doehler, Simona. 2004. Une approche interactionniste de la grammaire : réflexions autour du codage grammatical de la référence et des topics chez l'apprenant avancé d'une L2. *Acquisition et Interaction en Langue Étrangère* 21, 123-166.
- Roulet, Eddy, Filliettaz, Laurent & Grobet, Anne. 2001. *Un modèle et un instrument d'analyse de l'organisation du discours*. Berlin : Peter Lang.
- Sabio, Frédéric. 2006. Phrases et constructions verbales : quelques remarques sur les unités syntaxiques dans le français parlé. In Daniel Lebaud, Catherine Paulin & Katja Ploog (éd.). *Constructions verbales et production de sens*. Besançon : Presses Universitaires de Franche-Comté, 127-139.
- Sacks, Harvey, Schegloff Emanuel A. & Jefferson, Gail. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50(4), 696-735.
- Schegloff Emanuel A. 1996. Turn-organization: One intersection of grammar and interaction. In Elinor Ochs, Emanuel A. Schegloff & Sandra A. Thompson (éds.). *Interaction and grammar*. Cambridge : Cambridge University Press, 52-133.
- Searle, John R. 1972. *Les actes de langage*. Paris : Hermann.
- Selting, Margaret. 1996. On the interplay of syntax and prosody in the constitution of turn-constructive units and turns in conversation. *Pragmatics* 6 (3), 371-388.

Selting, Margaret. 2000. The construction of units in conversational talk. *Language in society* 29, 477-517.

Traverso, Véronique. 2012. Analyses interactionnelles : repères, questions saillantes et évolution. *Langue française* 175, 3-17.

Traverso, Véronique. 2016. *Décrire le français parlé en interaction*. Paris : Ophrys.

Ursi, Biagio, Delafontaine, François, Etienne, Carole, Badin, Flora, Baldauf-Quilliatre, Heike, Eshkol-Taravella, Iris, Jouin-Chardon, Emilie, Kanaan-Caillol, Loyal, Lambert, Margot, Rossi-Gensane, Nathalie, Skrovec, Marie & Traverso, Véronique. 2017. Segcor : vers une segmentation multiniveaux pour le français parlé. *Colloque Syntaxe et discours III – Types d'unités et procédures de segmentation*, Paris, 2017.

### **Sitographie**

FOLK : <[agd.ids-mannheim.de/folk.shtml](http://agd.ids-mannheim.de/folk.shtml)>

CLAPI : <[clapi.icar.cnrs.fr](http://clapi.icar.cnrs.fr)>

ESLO : <[eslo.huma-num.fr](http://eslo.huma-num.fr)>

SegCor : <[segcor.cnrs.fr](http://segcor.cnrs.fr)>

## Constitution et annotation d'un corpus micro-diachronique de français oral en vue de l'étude du subjonctif

Fatma Ben Barka Messaoudi

Université d'Orléans

Laboratoire Ligérien de Linguistique (UMR 7270)

### Résumé

Nous proposons dans cet article d'exposer la méthodologie que nous avons suivie pour constituer et annoter un corpus oral échantillonné de 1 164 998 mots et de 89 heures d'enregistrements, puisé dans les Enquêtes Sociolinguistiques à Orléans : ESLO1, 1968-1971, ESLO2, 2008- et ESLO-MD, 2018. Notre regard sera focalisé ici sur l'ensemble des procédures qui ont été mises en œuvre pour l'élaboration et l'exploitation de notre corpus dans l'objectif de vérifier une éventuelle évolution en micro-diachronie du subjonctif et de créer une cartographie des usages du subjonctif en français parlé à Orléans.

**Mots-clés :** subjonctif ; français oral ; corpus ; constitution ; annotation ; micro-diachronie

### Abstract

This article sets out the methodology followed to build and annotate a sampled oral corpus of 1 164 998 words and 89 hours of recordings, collected in the Sociolinguistic Surveys in Orléans: ESLO1, 1968-1971, ESLO2, 2008- et ESLO-MD, 2018. The focus will be on all the procedures that have been implemented for the elaboration and exploitation of the corpus in order to verify a potential evolution in micro-diachrony of the subjunctive and to create a mapping of the uses of the subjunctive in spoken French in Orléans.

**Keywords:** subjunctive; French; spoken language; corpus; constitution; annotation; micro-diachrony

Le subjonctif en français constitue un objet d'étude qui peut d'une façon pertinente bénéficier de l'apport de la nouvelle jonction entre données orales et outils du TAL. En effet, la plupart des analyses s'appuient encore aujourd'hui sur des descriptions opérées principalement sur de l'écrit (Nordhal 1969 ; Nølke 1985 ; Soutet 2000). Malgré un nombre considérable d'études, le subjonctif n'a pas disposé, à notre connaissance<sup>1</sup>, d'examen sur des données orales authentiques situées. De ce fait, le débat sur ses contextes d'emploi et ses valeurs sémantiques est loin d'être clos. Ces considérations nous ont poussée à proposer un réexamen de l'emploi de ce mode verbal en français en nous basant sur des données du français oral contemporain. Pour ce faire, nous nous sommes reposée sur le réservoir des Enquêtes Sociolinguistiques à Orléans (ESLO)<sup>2</sup> pour constituer notre propre corpus. Dans cet article, il sera question des principaux choix méthodologiques et techniques que nous avons opérés lors de la construction et l'exploitation de ce corpus, afin de procéder par la suite à l'analyse

---

<sup>1</sup> Seuls Laurier (1989) et Kastronic (2016) se sont penchés sur le français du Québec à partir de corpus oraux.

<sup>2</sup> <http://eslo.huma-num.fr/index.php/pagecorpus/pagepresentationcorpus>, consulté le 20 août 2019.

quantitative et qualitative de nos données.

Notre article s'organise en deux parties : une première partie où il sera question de la méthodologie de la constitution de notre sous-corpus d'étude et une seconde partie dans laquelle nous aborderons la démarche suivie dans l'exploitation de nos données à partir de la présentation et la démonstration de l'annotation.

## 1. Sous-corpus d'étude

Défini par Charaudeau et Maingueneau (2002 : 661) comme « un recueil large quelquefois exhaustif de documents et de données », et par Sinclair (cité par Habert *et al.* 1997 : 144) comme « une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites », le corpus se présente comme un échantillon de langage utilisé pour un objectif linguistique spécifique : « tester des hypothèses, confronter un modèle postulé aux réalisations effectives. » (Habert *et al.* 1997 : 8). Cette partie est dédiée à la présentation des données sur lesquelles notre étude doctorale (en cours) sur le subjonctif a été menée.

Depuis peu, un grand corpus de français oral a été mis à la disposition des chercheurs, qui « [a] pour vocation de permettre des études linguistiques à partir de données situées » (Baude & Dugua 2015 : 9) : les Enquêtes Sociolinguistiques à Orléans (ESLO<sup>3</sup>). Collecté en deux temps, à quarante ans de distance, ESLO est composé de deux grandes enquêtes :

- ESLO1 (1968 - 1970), corpus clos de 470 enregistrements et 4,5 millions de mots ;
- ESLO2 (depuis 2008), dont le but est d'atteindre plus de six millions de mots (avec actuellement 450 heures d'enregistrements).

Grâce à sa taille importante (actuellement autour de sept millions de mots) et à ses genres interactionnels assez diversifiés (entretiens, repas, conférences universitaires...), ce corpus offre la possibilité de mener des recherches sur des données orales diversifiées, situées et enrichies par des métadonnées informant sur la situation de communication et précisant le profil de chaque locuteur (son âge, son sexe et sa catégorie socioprofessionnelle). Nous avons choisi de fonder nos analyses sur l'exploration exhaustive de ce corpus, et plus particulièrement ESLO-MD<sup>4</sup> (ESLO Micro-Diachronie). Constitué initialement en 2014 pour l'étude du futur en français parlé, cet échantillon composé d'un million de mots et enrichi par des métadonnées et des annotations syntaxiques et sémantiques permet d'examiner l'évolution des différents phénomènes linguistiques en français oral contemporain. Nous entendons contribuer à l'étude de l'évolution des usages du subjonctif en nous basant sur un sous-

---

<sup>3</sup> Ce corpus est accessible et disponible à l'ensemble de la communauté scientifique. Tous ses documents sont délivrés sous licences Creative Commons.

<sup>4</sup> <https://www.ortolang.fr/market/corpora/eslo-md>, consulté le 15 août 2019.

corpus oral micro-diachronique puisé dans les ESLO.

Nous avons construit un sous-corpus échantillonné de 1 164 998 mots (ou 89 heures d'enregistrements) formé à parts égales d'enregistrements d'ESLO1, ESLO1 - MD et d'ESLO2, ESLO2 - MD. Le choix de nos locuteurs a été basé sur les principes suivants :

- se reposer sur trois catégories socioprofessionnelles: cadres, employés, ouvriers ;
- sélectionner trois catégories d'âge dans chaque catégorie socioprofessionnelle : 15-35 ans, 35-55 ans et plus de 55 ans ;
- choisir un homme et une femme de chaque catégorie d'âge.

	Enregistrement	Durée	Code locuteur	Sexe	Catégorie d'âge	CSP
ENT				H	plus 55	Employés
				H	15-35	Cadres
				H	15-35	Employés
				F	15-35	Cadres
				H	15-35	Ouvriers
				F	plus 55	Cadres
				F	35-55	Cadres
				H	plus 55	Cadres
				F	15-35	Employés
				F	15-35	Ouvriers
				F	35-55	Ouvriers
				F	plus 55	Employés
				F	plus 55	Ouvriers
				H	35-55	Employés
				F	35-55	Employés
				H	plus 55	Ouvriers
			H	35-55	Ouvriers	
			H	35-55	Cadres	

Figure 1. Échantillonnage des locuteurs

Le tableau ci-dessous résume les propriétés quantitatives (durée en minutes et nombre de locuteurs) du corpus et sa structure diaphasique :

Sous corpus ESLO-MD	ESLO1 / ESLO1-MD	ESLO2 / ESLO2- MD	Total
Durée (min)	2675 (44h35min)	2677 (44h37min)	5352 (89h12min)
	- ENT : 2236 - REP : 224 - CONF : 215	- ENT : 2230 - REP : 236 - CONF : 211	- ENT : 4466 - REP : 460 - CONF : 426
Nombre de mots	447 484	486 693	934 177
Nombre de locuteurs	37	35	72

Tableau 1. Composition du sous-corpus ESLO

Nous avons essayé, autant que possible, d'assurer un équilibre diaphasique et diastratique entre (et à l'intérieur de) chacun des deux extraits ESLO1/2 dans l'objectif de diminuer les écarts entre eux et de donner la meilleure comparabilité

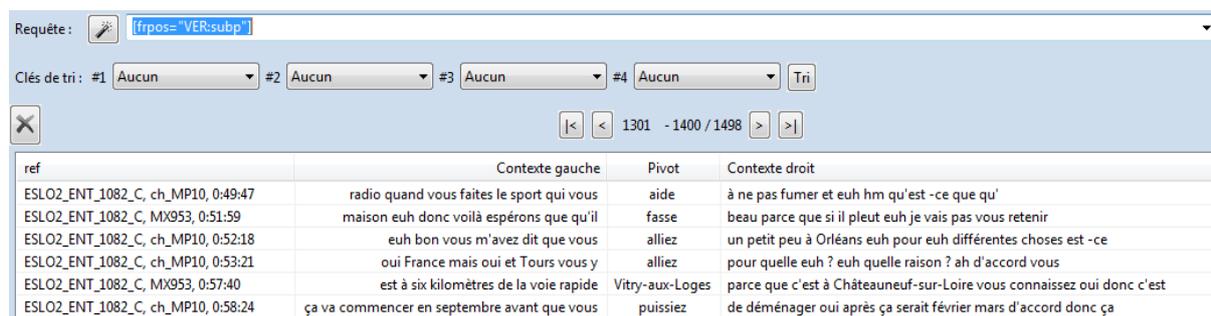
qualitative possible. Nous pouvons dire ainsi que, dans le cadre de notre étude sur le subjonctif, l'axe le plus important est celui qui met en balance les données sur le plan micro-diachronique.

## 2. Démarche d'annotation

Une fois le corpus élaboré, se pose la question de son annotation pour rendre possible son traitement. Inspirée d'Abouda & Skrovec (2015, 2018), nous avons entamé l'étape d'annotation de notre corpus qui a exigé, après l'extraction automatique des occurrences du subjonctif, notre intervention pour l'ajout des couches d'annotation syntaxiques et sémantiques.

### 2.1 Extraction du subjonctif

Pour exploiter le corpus constitué, nous avons utilisé un logiciel d'analyse textométrique, TXM<sup>5</sup>, qui permet, couplé à l'étiqueteur morphosyntaxique Treetagger<sup>6</sup>, d'importer les transcriptions des corpus oraux, faire plusieurs requêtes et calculer le vocabulaire d'ensemble<sup>7</sup> d'un corpus. L'étiquetage morphosyntaxique de notre corpus nous a permis d'extraire 1743 occurrences du subjonctif présent (subp) et 566 occurrences du subjonctif imparfait (subi) grâce aux requêtes CQL<sup>8</sup>.



ref	Contexte gauche	Pivot	Contexte droit
ESLO2_ENT_1082_C, ch_MP10, 0:49:47	radio quand vous faites le sport qui vous	aide	à ne pas fumer et euh hm qu'est -ce que qu'
ESLO2_ENT_1082_C, MX953, 0:51:59	maison euh donc voilà espérons que qu'il	fasse	beau parce que si il pleut euh je vais pas vous retenir
ESLO2_ENT_1082_C, ch_MP10, 0:52:18	euh bon vous m'avez dit que vous	alliez	un petit peu à Orléans euh pour euh différentes choses est -ce
ESLO2_ENT_1082_C, ch_MP10, 0:53:21	oui France mais oui et Tours vous y	alliez	pour quelle euh ? euh quelle raison ? ah d'accord vous
ESLO2_ENT_1082_C, MX953, 0:57:40	est à six kilomètres de la voie rapide	Vitry-aux-Loges	parce que c'est à Châteauneuf-sur-Loire vous connaissez oui donc c'est
ESLO2_ENT_1082_C, ch_MP10, 0:58:24	ça va commencer en septembre avant que vous	puissiez	de déménager oui après ça serait février mars d'accord donc ça

Figure 2. Extraction des formes subjonctives

Ensuite, nous avons exporté nos données sous forme de tableaux CSV pour les nettoyer manuellement (écarter les cas d'erreurs d'annotation et d'hypercorrection, étiquetés « erreur »).

<sup>5</sup> <http://textometrie.ens-lyon.fr/?lang=fr>, consulté le 15 août 2019.

<sup>6</sup> <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>, consulté le 15 août 2019.

<sup>7</sup> La plateforme TXM permet le décompte des mots et de leurs propriétés lexicales et syntaxiques (<http://textometrie.ens-lyon.fr/files/documentation/Manuel%20de%20TXM%200.7%20FR.pdf>, p. 5).

<sup>8</sup> CQL, acronyme pour *Corpus Query Language*, est le langage d'expression de requêtes utilisé pour interroger les corpus sous TXM. La syntaxe de la requête pour le SUB.p est la suivante : [frpos="VER:subp"] et pour le SUB.i : [frpos="VER:subi"].

Référence	ContexteGauche	Pivot	forme	Imm.vsub	sub	ElmR
ESLO1_CONF_503_C, 503CONF, 1:26:35	ils ont pas je crois pas qu'ils	aient	subps	comprendre	compl	elmtpr
ESLO1_CONF_503_C, 503LOC3, 1:26:43	se plaignent infiniment que les cours de langues	soient	subp	être	compl	elmtpr
ESLO1_CONF_504_C, 504CONF, 0:15:54	mécanisation l'informatique et cetera qu'il s'	agisse	subp	s'agir	non	subord
ESLO1_CONF_504_C, 504CONF, 0:19:55	pour le maître trop récente pour qu'il	puisse	subp	pouvoir	cir	subord
ESLO1_CONF_504_C, 504CONF, 0:24:32	collègue hé bien comment voulez -vous qu'il	fasse	subp	faire	compl	elmtpr
ESLO1_CONF_504_C, 504CONF, 0:28:44	qui achèvent leur licence par exemple hein euh	soit	conj	être		
ESLO1_CONF_504_C, 504CONF, 0:37:53	condition que les dimensions de cet espace ne	soient	subp	être	cir	subord
ESLO1_CONF_504_C, 504CONF, 0:43:09	le professeur qui accepte que les symboles mathématiques	soient	subp	être	compl	elmtpr
ESLO1_CONF_504_C, 504CONF, 0:45:12	cycliste enfin c'est le fait qu'on	puisse	subp	pouvoir	compl	elmtpr
ESLO1_CONF_504_C, 504CONF, 0:50:54	euh l'essentiel c'est que l'enfant	sache	subp	savoir	compl	elmtpr
ESLO1_CONF_504_C, 504CONF, 0:51:02	un raccourci dans certaines situations par exemple et	puisse	subp	pouvoir	compl	elmtpr
ESLO1_CONF_504_C, 504CONF, 0:51:05	transmettre cette connaissance à autrui voyez qu'il	puisse	subp	pouvoir	compl	elmtpr
ESLO1_CONF_504_C, 504CONF, 0:51:14	mais il n'est pas interdit qu'il	construise	subp	construire	compl	elmtpr
ESLO1_CONF_504_C, 504CONF, 0:55:45	que les deux aspects dont nous parlons ne	soient	subp	être	non	subord
ESLO1_CONF_504_C, 504CONF, 0:59:26	nombre d'idées mais je souhaite que vous	puissiez	subp	pouvoir	compl	elmtpr
ESLO1_CONF_504_C, 504CONF, 1:01:13	vous auriez évoquées avec vos conseillers pédagogiques euh	erreur				

Figure 3. Exportation des occurrences du subjonctif sous forme de tableau CSV

Ainsi, dans l'exemple ci-dessous, le mot *seconde* a été extrait comme une occurrence du subjonctif, venant du verbe *seconder* (1ère ou 3ème personne du singulier), alors qu'il s'agit plutôt d'un substantif :

- (1) ESLO1\_CONF\_504\_C, 504CONF, 1:28:34  
voquez c'est en seconde enfin c'est d'abord revenu en en terminale

Il faut noter aussi que quand le morphème *soit* est employé comme une conjonction disjonctive, nous l'avons étiqueté « conjd » :

- (2) ESLO2\_REPAS\_1254\_C, WZ384MAR, 0:37:39  
elle était soit ministre euh soit premier ministre

Comme nous nous intéressons uniquement aux formes verbales subjonctives, nous avons fait le choix d'écarter ce cas d'emploi lors des prochaines étapes d'annotation. Lors de cette phase d'observation, nous avons constaté que quelques formes du subjonctif passé ont été extraites en tant que formes de subjonctif présent. Nous avons donc procédé leur correction manuelle en les étiquetant « subps ».

Le tableau suivant présente le nombre d'occurrences du subjonctif obtenues après avoir nettoyé les bruits (erreurs de transcription, formes verbales du présent de l'indicatif) et éliminé les conjonctions disjonctives.

Après correction	ESLO1 + ESLO2 ESLOA-MD + ESLO2-MD
Subjonctif présent (subp)	1339
Subjonctif passé (subps)	21
Subjonctif imparfait (subi)	0

Tableau 2. Nombre d'occurrences du subjonctif

## 2.2 Annotation affinée des données

Les occurrences, une fois corrigées, ont fait l’objet d’une annotation affinée<sup>9</sup>. Lors de cette phase descriptive, afin de sous-spécifier chacune de nos occurrences par des valeurs et des attributs jugés pertinents (en veillant à éliminer tous types d’ambiguïté), nous avons consulté le contexte élargi de chaque occurrence, et écouté l’extrait d’enregistrement correspondant. Sur notre fichier CSV, nous avons inséré nos valeurs et attributs entre le pivot et le contexte droit sous forme de colonnes. Nous avons attribué à chaque couche d’annotation une étiquette (par exemple, forme pour préciser s’il s’agit d’un subjonctif présent, un subjonctif passé ou un subjonctif imparfait).

Pivot	forme	sub	ElmR	ImmSUB	ContexteDroit
vienne	subp	cir	subord	venir	pas euh on attaque sans lui hein attaquons attaquons attaquons euh
soient	subp	non	subord	être	des classes de ? classés niveaux soient séparés par niveau c'est
soient	subp	non	subord	être	séparés par niveau c'est ce qu'on a chez nous un
soient	subp	non	elmtpr	être	absolument indispensables au niveau de la seconde où les élèves arrivent euh
puisse	subp	rel.adj	ant	pouvoir	faire pour eux c'est de leur mettre constamment en situation la
soient	subp	compl	elmtpr	être	euh et alors euh ça introduit euh la notion d'humour et
soit	subp	compl	elmtpr	être	découpé en rondelles tant de minutes pour ça tant de minutes pour
fasse	subp	cir	subord	faire	une mi-temps hé mais oui bah d'ailleurs c'est pas l'
ait	subp	non	subord	avoir	des feuilles ou n'en ait pas que ses feuilles soient euh
ait	subp	non	subord	avoir	pas que ses feuilles soient euh trilobées ou quadrilobées que elles aient
soient	subp	non	subord	être	euh trilobées ou quadrilobées que elles aient euh un long pétiole ou
aient	subp	non	subord	être	euh un long pétiole ou pas tout ça n'a aucune importance
plaise	subp	compl	elmtpr	plaire	pas disons des conséquences d'ordre pédagogique en ce qui concerne les
puisse	subp	compl	elmtpr	pouvoir	mettre sur pied une méthode valable d'enseignement des langues valable pour

Figure 4. Insertion des étiquettes d’annotation

Notre méthodologie d’analyse, prise dans sa globalité, a été fondée sur corpus (*data based* dans la conception de Biber 2009) et conduite par les données (*data driven* pour Biber 2009). Autrement dit, notre idée de départ a consisté à fouiller dans un ensemble de données langagières pour vérifier les hypothèses classiques formulées sur le subjonctif selon lesquelles :

- il est considéré comme le mode de l’irréalité par opposition à l’indicatif, mode de la réalité ;
- la catégorie sémantique de son élément introducteur se rapporte à une notion de doute quant à la réalisation du procès ;
- il apparaît soit dans une proposition indépendante, soit dans une proposition subordonnée introduite par *que*.

Nonobstant, nous avons été confrontée à des exemples non standards et qui ont exigé une réadaptation de notre méthodologie de travail afin de les décrire. Regardons de près l’exemple suivant :

- (3) ESLO1\_ENT\_001\_C, OU, 0:59:30  
c’est dommage ça vienne à finir

<sup>9</sup> Les opérations de correction et d’annotation des occurrences ont été réalisées manuellement.

Si nous nous étions contentée d'effectuer uniquement une analyse sur corpus, cette occurrence aurait été éliminée parce qu'elle n'est pas conforme aux exemples standards qu'on aurait cherchés pour illustrer des classes d'emplois prédéfinies. Il nous a paru donc inévitable d'inventer une catégorie X dans laquelle nous avons classé les cas d'emplois atypiques du subjonctif. La création de cette catégorie permet de regarder de près les usages du subjonctif dans les corpus oraux qui diffèrent de ceux dans les corpus écrits afin de mieux comprendre le fonctionnement de ce mode dans les deux situations.

Nous pouvons dire ainsi que l'annotation de nos occurrences a été accomplie en plusieurs étapes au cours desquelles nous avons mobilisé un certain nombre d'étiquettes inspirées de nos hypothèses et guidées par nos données.

Le mode subjonctif est parmi les rares, voire le seul avec une telle ampleur, à être examiné sous l'angle de la syntaxe (voir par exemple Abouda 1997, 2002). Il nous semble donc intéressant de caractériser la structure syntaxique de la phrase où apparaît ce mode, pour vérifier les hypothèses classiques classant les contextes d'emplois du subjonctif en deux catégories principales, *i.e.* proposition indépendante et proposition subordonnée.

Le subjonctif dans une proposition non subordonnée a été étiqueté « - subor » :

- (4) ESLO1\_ENT\_001\_C, BA725, 0:12:03  
oui écoutez que je réfléchisse

Lorsqu'il apparaît dans une subordonnée (étiquetée « + subor »), nous avons spécifié son type :

- une complétive directe ou indirecte (étiquetée « compl »)

- (5) ESLO2\_DIA\_1227\_C, YT387\_DIA, 0:59:21  
faudrait que je le fasse

- (6) ESLO2\_ENT\_1002\_C, RL2, 0:11:59  
je tenais absolument à ce qu'ils fassent des études

- une circonstancielle (étiquetée « cir »)

- (7) ESLO1\_ENT\_008\_C, GJ131, 0:31:03  
on tourne jusqu'à temps que ça soit à peu près cuit quoi

- une relative (étiquetée « rel »)

- (8) ESLO1\_ENT\_058\_C, DH59, 1:14:17  
je me souviens plus plutôt des livres qui soient un peu scientifiques plutôt

La détermination du mode verbal emprunte des chemins syntaxiques de type « complémentation »<sup>10</sup>. L'étape suivante a consisté donc à informer un certain nombre d'attributs qui se différencient suivant le type de subordonnée. Le renseignement de ces éléments avait pour objectif de préciser l'élément déclencheur du subjonctif dans la subordonnée.

D'après les descriptions grammaticales, cet élément peut être soit :

- un élément recteur dans la principale (étiqueté « elmt.r »)
- (9) ESLO2\_DIA\_1225\_C,RF211\_DIA,0:29:06  
j'aimerais bien que euh nos jeunes lycéens aient un petit peu plus une vision de euh de l'histoire
- une conjonctive (étiquetée « conj »)
- (10) ESLO1\_ENT\_029\_C, 1268, 0:26:03  
je ferais n'importe quoi pour qu'ils fassent des études euh le plus longtemps possible
- un antécédent (étiqueté « ± ant<sup>11</sup> »)
- (11) ESLO1\_ENT\_132\_C, BV, 0:55:0  
vous ne connaissez pas de gens qui aient des enfants ?

Outre ces exemples typiques appartenant aux classes d'emplois prédéterminées, nous avons rencontré quelques exemples difficiles à étiqueter vu l'ambiguïté de leurs contextes d'emplois :

- (12) ESLO1\_ENT\_124\_C, SD272, 0:49:10  
RC: et lorsque vous votez pour un député euh qu'est-ce que vous attendez de lui ?  
SD 272: ben qu'il nous défende qu'il défende tous nos droits

Si nous nous reposons sur la notion de phrase, catégorie classique d'analyse syntaxique, pour examiner cet exemple, nous nous arrêtons au niveau de la ponctuation forte et, dans ce cas-là, nous pouvons nous interroger sur les contraintes syntaxiques imposant l'emploi du « subp » par le locuteur SD 272. Mais, dans le cadre de notre analyse, nous prenons aussi en compte les spécificités des données interactionnelles qui nous pousse à observer toute la séquence conversationnelle (incluant les tours de parole précédents). Cette démarche a révélé que le verbe *attendre*

<sup>10</sup> Abouda (1998 : 294) explique que « les modes, tous les modes, ont un lien étroit avec les caractéristiques syntaxiques de la proposition dans laquelle ils entrent. Ces caractéristiques les rendent syntaxiquement visibles. »

<sup>11</sup> Nous avons choisi les étiquettes « + ant » pour les relatives adjectives et « - ant » pour les relatives substantives.

utilisé dans l'interrogation de RC est l'élément qui a entraîné l'emploi du subp dans le tour de SD 272. En d'autres termes, le locuteur SD 272 a restitué mentalement cet élément pour répondre à la question de son interlocuteur, ce qui explique l'ellipse de la principale dans son tour de parole.

- (13) ESLO2\_ENT\_1085\_C, RN488, 0:41:43  
 ch\_MP 10: donc ça c'est ce qui vous intéresse le débat  
 RN 488: oui  
 ch\_MP 10 : ouais  
 RN 488: oui oui que chacun puisse euh échanger donner son avis

L'énoncé de ch\_MP 10 a permis au RN 488 la conception du schéma constructionnel complet (Langacker, 1987). Nous dépassons ici la perspective énonciative qui considère le locuteur comme un sujet solitaire pour arriver au phénomène de coénonciation qui met en avant la mise en place d'une stratégie d'anticipation de la configuration syntaxique de l'énoncé par RN 488 pour l'achever à la place de ch\_MP10.

L'analyse de ces deux exemples montre que, pour pouvoir décrire syntaxiquement les cas d'emploi du subjonctif à l'oral, il faut prendre en considération tous les aspects conversationnels. Les éléments interactionnels s'avèrent donc indispensables à l'étude des phénomènes qui échappent aux descriptions syntaxiques standards. Ainsi, comme le note Mondada (2001 : 5) :

---

La hiérarchisation des niveaux d'analyse est ainsi bouleversée: on ne peut plus pratiquer une analyse phonologique et morpho-syntaxique indépendante de considérations interactionnelles, puisque les principes de l'organisation conversationnelle pénètrent tous les niveaux d'analyse.

---

Passons maintenant aux emplois du subjonctif dans les subordonnées complétives. Pour mieux catégoriser le constituant-noyau déclenchant le subjonctif, nous l'avons examiné de près en identifiant sa nature grammaticale (nom étiqueté « N », adjectif étiqueté « Adj » ou verbe étiqueté « V ») et sa valeur sémantique (étiquetée « cs ».) Le lemme a été précisé seulement si l'élément principal est un verbe<sup>12</sup> :

- (14) ESLO1\_ENT\_149\_C, YR399, 0:47:01  
 il faut quand même que l'on connaisse euh l'histoire et la géographie
- « V » = faut
  - « Imm.pr » = falloir
  - « cs » = obligation

- (15) ESLO2\_ENT\_1009\_C, LX10, 0:53:50  
 mes craintes hm qu'il soit moins intéressant

---

<sup>12</sup> Nous envisageons une classification sémantique de ces verbes.

- « N » = craintes
- « cs » = sentiment

Afin de mieux décrire le mécanisme syntaxico-sémantique conditionnant le choix du subjonctif dans les subordinées circonstancielles, nous avons identifié la locution conjonctive (étiquetée « conj ») et la valeur sémantique exprimée par la circonstancielle (étiquetée « v.cir »).

(16) ESLO1\_REPAS\_278\_C, XZ501, 0:32:15  
avant que je sorte il faut que je fasse mon boulot

- « conj » = avant que
- « v.cir » = temps

Quant aux emplois du subjonctif dans les relatives adjectives, nous avons choisi de mentionner le type d'antécédent (étiqueté « typ.ant ») pour mettre en avant les cas les plus fréquents qui sélectionnent le subjonctif dans ce contexte d'alternance modale.

(17) ESLO1\_ENT\_001\_C, BA725, 0:39:36  
c'est la seule chose que je puisse vous dire à ce sujet-là

- « + ant » = la seule chose (cf. Frontier 1997)
- « typ.ant » = superlatif

Enfin, une fois la contextualisation des occurrences subjonctives faite, nous avons identifié les lemmes des verbes subordinés conjugués au subjonctif et avons vérifié leur éventuelle commutation avec l'indicatif (étiquetée « comm »).

(18) ESLO1\_CONF\_503\_C, 503CONF, 1:26:35  
je crois pas qu'ils aient compris

- « lmm » = comprendre
- L'emploi du subjonctif dans cet exemple a été régi par la présence du verbe *croire* à la forme négative dans la proposition principale. Cependant, selon les descriptions grammaticales, l'alternance modale est ici possible. On note donc « comm » = possible (étiquetée « poss »).

Le schéma global de l'annotation des occurrences du subjonctif se présente comme suit :

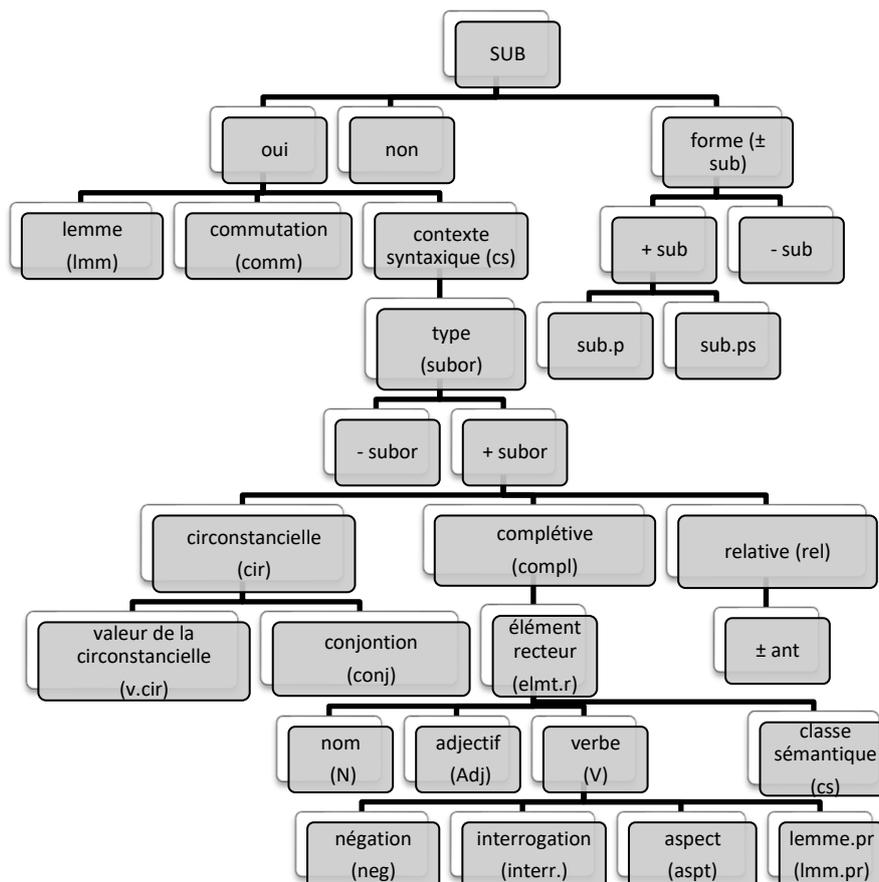


Figure 5. Schéma d'annotation

Après avoir examiné minutieusement nos données, nous les avons réinjectées dans TXM. Les différentes propriétés annotées ont constitué des variables pouvant faire l'objet de plusieurs requêtes possibles en rapport avec d'autres variables distributionnelles disponibles sur TXM (catégorie grammaticale, personne, etc.), ainsi qu'avec les métadonnées des ESLO (genres interactionnels, tranche d'âge, catégorie socio-professionnelle, etc.)

Les premières tendances qui se dégagent sur le plan microdiachronique sont les suivantes :

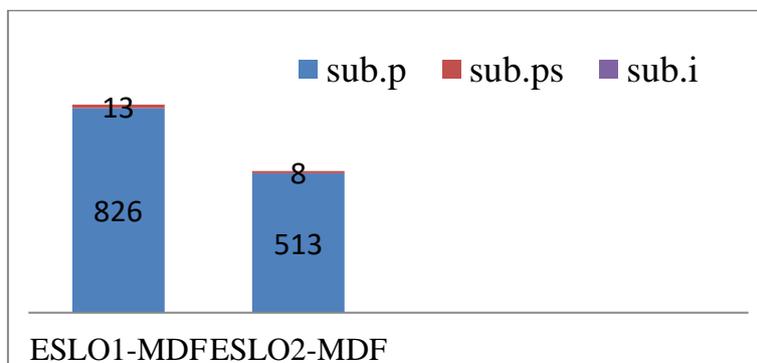


Figure 6. Tendances en micro-diachronie : SUB.P / SUB.PS / SUB.IMP

Nos données quantitatives montrent une chute de 17 % de l'emploi du subjonctif

présent (SUB.P), une quasi-stabilité du subjonctif passé (SUB.PS) et une absence du subjonctif imparfait (SUB.I).

## Conclusion

Dans cet article, nous avons présenté les différentes étapes de la construction et de l'exploitation d'un corpus micro-diachronique de français oral en vue de l'étude du subjonctif.

Lors de cette phase de travail méthodologique, nous avons tout d'abord constitué un corpus échantillonné et équilibré contenant des données comparables prélevées à parts égales dans les extraits d'ESLO. Ensuite, nous avons procédé à l'enrichissement de ces données par l'ajout des informations syntaxiques et sémantiques pour faciliter l'exploration de notre corpus sur les plans quantitatif et qualitatif.

Nous accordons priorité au jeu des corrélations qui peuvent être établies entre les variables mesurables sous observation. Nous nous focalisons aussi sur les représentations sociales en prenant en considération leur aspect dynamique et leur rôle important dans les processus de production et de réception.

Il sera question, dans les prochaines étapes, de croiser les données linguistiques avec les informations dont nous disposons sur les locuteurs, concernant non seulement les informations sociodémographiques (âge, sexe, catégorie socio-professionnelle), mais aussi d'autres données plus fines qui permettent de mieux catégoriser les locuteurs (discours obtenus par le biais des enregistrements) afin d'établir des liens entre les statistiques que nous obtiendrons et les contenus de nos données.

Nous entendons par la constitution de ce corpus l'examen de l'évolution des usages du subjonctif en français contemporain afin d'apporter des informations concrètes sur la validité des descriptions grammaticales classiques de ce mode et de créer une cartographie de son usage en français parlé à Orléans.

## Références bibliographiques

- Abouda, Lotfi. 1997. « Le conditionnel : temps ou mode ? Arguments syntaxiques ». *Revue Romane*, 32 (2), 179-198.
- Abouda, Lotfi. 1998. « Vers une localisation syntaxique des modes verbaux : Cas de la phrase indépendante ». *Cahiers Chronos*, Rodopi, Variations sur la référence verbale, 3, 293-322.
- Abouda, Lotfi. 2002. « Négation, Interrogation et Alternance Indicatif-Subjonctif ». *Journal of French Language Studies*, 12, 1-22.

- Abouda, Lotfi. 2008. « De la visibilité syntaxique des modes, de l'invisibilité syntaxique des temps », VIIIe Congrès International de Linguistique Française, Oviedo, Espagne, 319-333.
- Abouda, Lotfi. & Skrovec, Marie. 2015. « Du rapport entre formes synthétique et analytique du futur. Étude de la variable modale dans un corpus oral micro-diachronique », *Revue de Sémantique et Pragmatique*, 38, 35-57.
- Abouda, Lotfi. & Skrovec, Marie. 2018. « Pour une micro-diachronie de l'oral : le corpus ESLO-MD », CMLF 2018. Laboratoire Ligérien de Linguistique - UMR 7270 (LLL) (2018). ESLO-MD : Enquêtes Socio-Linguistiques à Orléans : Corpus Micro-Diachronie [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage) - [www.ortolang.fr](http://www.ortolang.fr), <https://hdl.handle.net/11403/eslo-md/v1>.
- Baude, Olivier & Dugua, Céline. 2015. « Corpus et savoirs : des liaisons heureuses ? », Colloque SHESL-HTL – Corpus et constitution des savoirs linguistiques, 9-10.
- Biber, Douglas. 2009. « Corpus-Based and Corpus-driven Analyses of Language Variation and Use » in Bernd Heine and Heiko Narrog (ed.), *The Oxford Handbook of Linguistic Analysis*, Oxford : Oxford University Press.
- Charaudeau, Patrick & Maingueneau, Dominique. 2002. *Dictionnaire d'Analyse du Discours*. Paris : Seuil.
- Frontier, Alain. 1997. *La Grammaire du français*. Paris : Belin.
- Habert, Benoît, Nazarenko, Adeline & Salem, André. 1997. *Les linguistiques de corpus*, Paris : Armand Colin.
- Heiden, Serge, Jean-Philippe Magué & Bénédicte Pincemin. 2010. TXM : Une plateforme logicielle open-source pour la textométrie-conception et développement. *10th International Conference on the Statistical Analysis of Textual Data-JADT 2010*, vol. 2, 1021–1032. Edizioni Universitarie di Lettere Economia Diritto.
- Kastronic, Laura. 2016. *A Comparative Variationist Approach to Morphosyntactic Variation in Hexagonal and Quebec French*, Thèse de doctorat en linguistique, Université Ottawa, Canada.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar, vol. 1: Theoretical Prerequisites*. Stanford: Stanford University Press.
- Laurier, Michel. 1989. « Le subjonctif dans le parler franco-ontarien : un mode en voie de disparition ? », Raymond Mougéon et Edouard Béniak (dirs.), *Le français canadien parlé hors Québec*. Québec : Les presses de l'Université Laval.
- Mondada, Lorenza. 2001. « Pour une linguistique interactionnelle », *Marges Linguistiques*, 1, mai, 142-162 (archive accessible sur Internet : <http://www.revue-texto.net/Archives/Archives.htm> ) (republié dans M. Santacrose, éd., 2001, *Faits de langue, faits de discours*, 2, Paris : L'Harmattan, 95-136) (trad. esp. Por una lingüística interaccional, *Discurso y Sociedad*, 3/3, 2001, 61-90).
- Nølke, Henning. 1985. Le subjonctif, fragments d'une théorie énonciative. *Langages* 20(80). 55–70. doi:10.3406/lgge.1985.1513.

Nordahl, Helge. 1969. *Les systèmes du subjonctif corrélatif. Etude sur l'emploi des modes dans la subordonnée complétive en français moderne*, Universitetsforlaget, Bergen-Oslo.

Sinclair, John. 1996. *Preliminary recommendations on Corpus Typology*, Rap.Tech., EAGLES (Expert Advisory Group on Language Engineering Standards) May 1996, CEE.

Siouffi, Gilles, Agnès Steuckardt & Chantal Wionet. 2012. Comment enquêter sur les diachronies courtes et contemporaines? *SHS Web of Conferences* 1. 215–226. doi:10.1051/shsconf/20120100214.

Soutet, Olivier. 2002. *Le subjonctif en français*. Paris : Ophrys.