

## 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

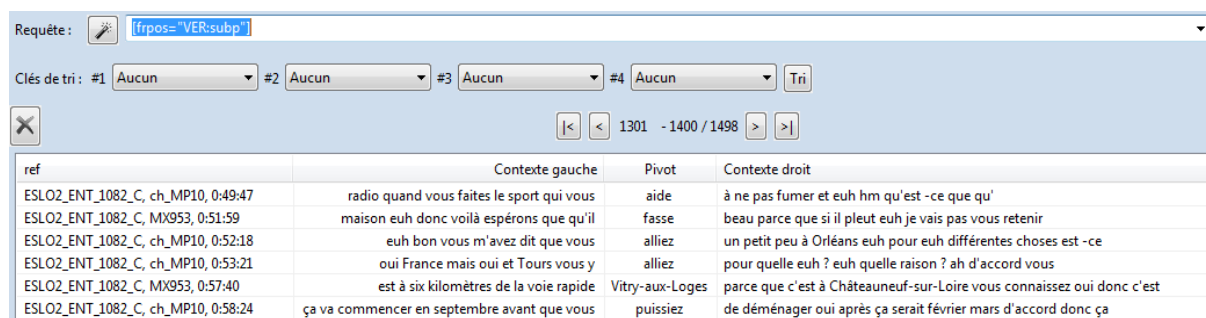
qualitative possible. Nous pouvons dire ainsi que, dans le cadre de notre étude sur le subjonctif, l'axe le plus important est celui qui met en balance les données sur le plan micro-diachronique.

## 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  
 voyez 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 :

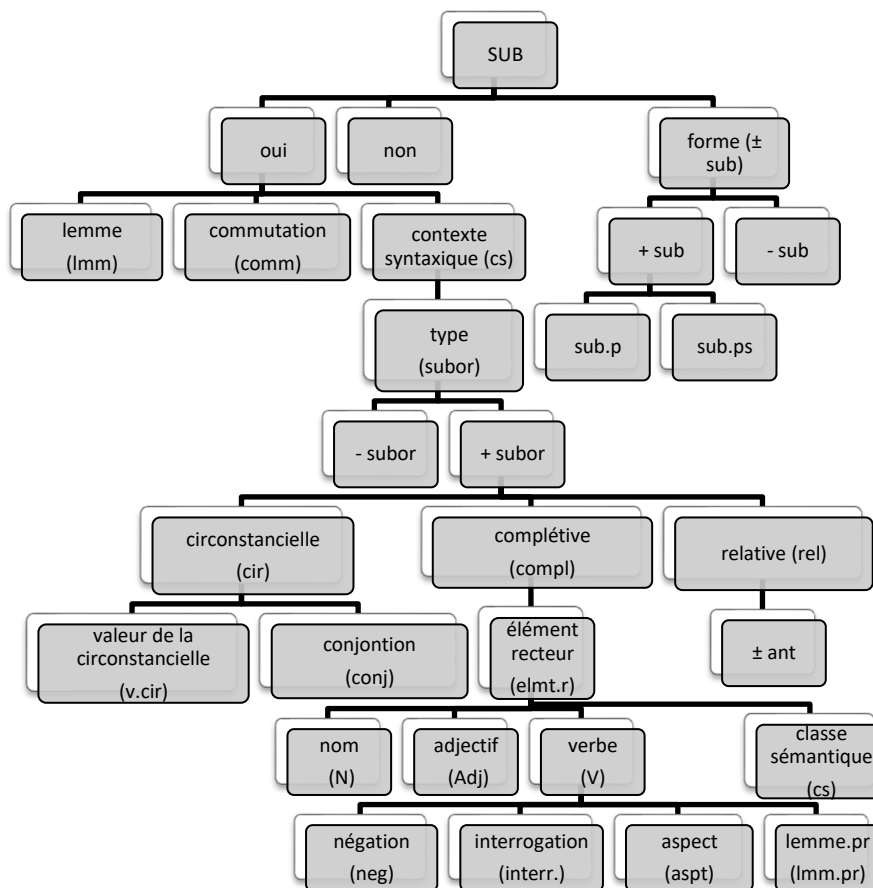


Figure 5. Schéma d'annotation

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 :

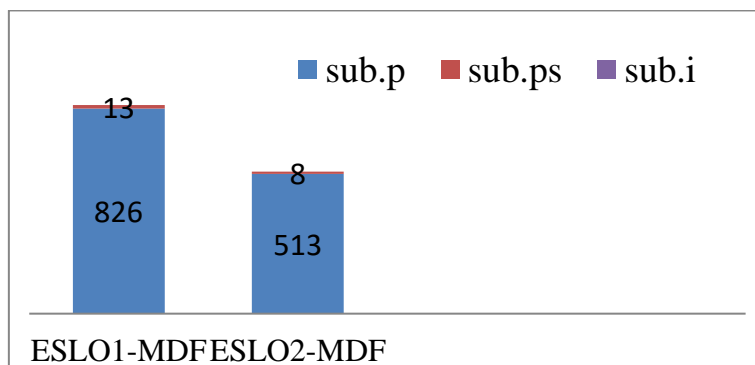


Figure 6. Tendances en micro-diachronie : SUB.P / SUB.PS / SUB.IMP

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.