



HAL
open science

Transcription et annotation de données orales pour étudier la prosodie en FLE : enjeux méthodologiques

Fabian Santiago

► To cite this version:

Fabian Santiago. Transcription et annotation de données orales pour étudier la prosodie en FLE : enjeux méthodologiques. Giuliano Bernini; Ada Valentini; Lorenzo Spreafico; Jacopo Saturno. *Superare l'evanescenza del parlato. Un vademecum per il trattamento digitale di dati*, Bergamo : Sestante, pp.167-199, 2021. halshs-03252621

HAL Id: halshs-03252621

<https://shs.hal.science/halshs-03252621>

Submitted on 22 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SUPERARE L'EVANESCENZA DEL PARLATO

Un vademecum per il trattamento digitale di dati linguistici

a cura di

Giuliano Bernini - Ada Valentini
Jacopo Saturno - Lorenzo Spreafico



BERGAMO UNIVERSITY PRESS

sestante edizioni



**UNIVERSITÀ
DEGLI STUDI
DI BERGAMO**

Dipartimento
di Lingue, Letterature
e Culture Straniere

Comitato scientifico
Giuliano Bernini
Maria Grazia Cammarota
Ada Valentini
Università di Bergamo
Régine Delamotte
Université de Rouen
Klaus Düwel
Universität Göttingen
Edgar Radtke
Universität Heidelberg

© 2021, Bergamo University Press
Sestante Edizioni - Bergamo
www.sestanteedizioni.it

SUPERARE L'EVANESCENZA DEL PARLATO

Un vademecum per il trattamento digitale di dati linguistici

Giuliano Bernini - Ada Valentini - Jacopo Saturno - Lorenzo Spreafico (A cura di)

p. 262 cm. 15,5x22,0

ISBN: 978-88-6642-369-0

Printed in Italy
by Sestanteinc - Bergamo

In copertina: “Evangelista, or letter-writer, and his clients”. Immagine tratta da Brown, Robert. 1894. *The Countries of the World: being a popular description of the various continents, islands, rivers, seas, and peoples of the globe [with plates]*. Londra: Cassell, Petter & Galpin. <https://www.flickr.com/photos/britishlibrary/11226480883/>

Superare l'evanescenza del parlato: lo sforzo può comportare lo sgomento riflesso nel volto dello scriba di fronte ai modi di parlare di personaggi tanto diversi.

SUPERARE L'EVANESCENZA DEL PARLATO

Un vademecum per il trattamento
digitale di dati linguistici

a cura di

Giuliano Bernini - Ada Valentini
Jacopo Saturno - Lorenzo Spreafico



BERGAMO UNIVERSITY PRESS

sestante edizioni

Direttore responsabile
Prof. Giuliano Bernini

Biblioteca di Linguistica e Filologia

6.

Superare l'evanescenza del parlato
Un vademecum per il trattamento digitale di dati linguistici

a cura di
Giuliano Bernini - Ada Valentini
Jacopo Saturno - Lorenzo Spreafico

Questo volume è stato stampato con il contributo del Dipartimento di Lingue, Letterature e Culture Straniere dell'Università degli Studi di Bergamo.

Contributi rivisti dai curatori.

Licenza *Creative Commons*:

This journal is published in Open Access under a Creative Commons License Attribution-Noncommercial-No Derivative Works (CC BY-NC-SA 3.0).

You are free to share – copy, distribute and transmit –
the work under the following conditions:

You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

You may not use this work for commercial purposes.

You may not alter, transform, or build upon this work.



Indice

<i>Introduzione</i>	p.	7
LORENZO SPREAFICO <i>La trascrizione strumentale del significante: dalle origini alle digital humanities</i>	»	11
ALESSANDRO VIETTI <i>Il ruolo della variabilità acustica nella costruzione del dato linguistico</i>	»	45
CINZIA AVESANI, BARBARA GILI FIVELA <i>Analysing Prosody: Methods, issues, and hints on crosslinguistic comparison and L2 learning</i>	»	71
SANDRA BENAZZO, MARZENA WATOREK <i>Transcription de corpus oraux d'apprenants débutants en français L2 : quelques enjeux théoriques</i>	»	127
FABIAN SANTIAGO <i>Transcription et annotation de données orales pour étudier la prosodie en FLE : enjeux méthodologiques</i>	»	167
LUCIANO ROMITO <i>La trascrizione in ambito forense</i>	»	201
JACOPO SATURNO <i>La trascrizione di dati linguistici – istruzioni di base</i>	»	231

FABIAN SANTIAGO

(Université de Paris 8)

Transcription et annotation de données orales pour étudier la prosodie en FLE : enjeux méthodologiques

1. *Introduction*

Les études en linguistique, en général, et en phonologie post-lexicale en particulier, sont souvent censées confronter les modèles théoriques d'analyse aux faits observables, que ceux-ci proviennent de données orales ou de productions écrites. De fait, toute étude voulant dégager le fonctionnement intonatif, accentuel ou segmental d'une langue, qu'elle soit L1 ou L2, repose essentiellement sur la description des événements sonores quantifiables extraits des données qui sont rassemblées dans des corpus. Dans le domaine des études de l'acquisition de la phonologie et la phonétique en L2, le chercheur voulant étudier l'interlangue des apprenants doit passer, dans la plupart des cas, par une étape de description fondée sur l'observation des faits. En observant une grande variété de données orales offrant une bonne représentativité de la langue cible, le chercheur peut expliquer la façon dont un système prosodique et phonologique d'une langue, qu'elle soit déjà acquise (L1) ou en cours d'acquisition (L2), fonctionne et se manifeste dans le flux de parole.

De surcroît, comparer des données produites par des locuteurs natifs et non natifs permet de dresser l'inventaire des formes prosodiques et de dire en quoi un système prosodique de la L2 s'écarte/se rapproche de celui de la L1, cela permettant à terme d'évaluer le poids du transfert de la L1 (ou d'autres L2), ainsi que les effets d'un processus universel d'acquisition en L2, entre autres phénomènes. Même si l'étude des corpus oraux permet d'atteindre de nombreux objectifs, toute approche par corpus nécessite de s'interroger sur leur nature, sur la manière de les compiler, de les annoter, etc. Pour garantir une bonne représentativité, plusieurs étapes nous paraissent essentielles (Delais-Roussarie 2003) :

- a. effectuer un travail préparatoire préalable à la constitution du corpus ainsi qu'établir une méthodologie appropriée pour enregistrer les données (choix du matériel et conditions d'enregistrements) ;
- b. mettre en forme les données pour les rendre exploitables (documentation du corpus, transcription et annotation) ;
- c. employer des métriques/mesures appropriées pour analyser des données riches, variées et/ou de grande échelle selon le phénomène étudié.

Dans cette contribution, notre objectif est double. D'une part, nous essayons de discuter des enjeux méthodologiques qu'une telle entreprise impose au chercheur voulant étudier certains phénomènes prosodiques en français langue étrangère (FLE) : Quelles données collecter en L2 : production orale contrôlée ou spontanée ? Comment les transcrire ? Comment annoter les phénomènes segmentaux et prosodiques afin que les données en L2 et en L1 soient comparables ? D'autre part, nous exposons les différentes métriques que les chercheurs peuvent employer pour étudier la prosodie en FLE.

Notre contribution est organisée comme suit. La section deux porte une attention particulière à la méthodologie sur la constitution d'un corpus en français L2 pour l'étude de la prosodie, la phonologie et la phonétique. Dans cette section, nous soulevons les problèmes méthodologiques préalables à l'analyse des données : transcription et annotation. La section trois présente de manière plus détaillée les enjeux de l'annotation prosodique en FLE. Enfin, dans la dernière section, nous dressons un bilan sur les perspectives que ces enjeux méthodologiques soulèvent en phonologie post-lexicale et segmentale des langues secondes.

2. Corpus oraux en L2 : collecte de données, transcription et annotation

Le premier point auquel tout chercheur est confronté est de savoir en quoi consiste un corpus, d'autant que collecter des données à l'ère d'internet est aisé. Sinclair (1996 : 4) propose la définition suivante : « A corpus is a collection of *pieces of language* that are selected according to explicit linguistic criteria in order to be used as a sample of the language ».

Constituer un corpus (qu'il soit écrit ou oral, en L1 ou en L2) ne se réduit aucunement à une simple collecte arbitraire d'échantillons de langage, ni à un simple dépôt des données orales numérisées. La constitution d'un corpus doit être systématiquement organisée et contrôlée conformément à un protocole élaboré par le chercheur en fonction des hypothèses ou des objectifs fixés.

Construire un corpus d'apprenants nécessite de se poser des questions d'ordre méthodologique et théorique, d'autant que le nombre de variables qui peuvent affecter le processus d'acquisition de la L2 est considérable :

- Quel type de données doit être rassemblé et étudié pour étudier l'acquisition de la prosodie et de la phonologie d'une L2 ?
- Les tâches employées pour collecter ces données doivent-elles être conformes à un protocole très contrôlé ?
- Les tâches réalisées par les participants en L2 représentent-elles des situations de communication authentiques ?
- Comment transcrire et annoter des données en L1 et L2 afin d'étudier les phénomènes prosodiques ?

Nous considérons que répondre à ces questions et justifier les réponses en fonction des présupposés théoriques et méthodologiques retenus est une étape indispensable pour garantir la validité des observations faites lors des études fondées sur une approche par corpus. Un point important pour garantir la validité des résultats concerne la représentativité des données compilées. En effet, quelle que soit la méthodologie employée pour compiler le corpus, tout chercheur veut que les données collectées soient représentatives des situations de communication réelles ; autrement dit, que les analyses qu'il extrait des données soient faites à partir de structures langagières observables. Nous allons présenter quel type de protocole nous avons adopté pour les données que nous examinerons plus tard. Au préalable, nous allons présenter différentes approches méthodologiques utilisées dans des études sur l'acquisition des L2, et plus particulièrement, de la prosodie et de la phonétique.

2.1 Collecte et type de données

La plupart des études en prosodie de L2 a privilégié les données expérimentales (ou construites) et se centrent sur l'étude d'un seul aspect de la prosodie de la L2 (*cf.* la trentaine d'articles cités et examinés par

Colantoni & Steele & Escudero 2015). A cela s'ajoutent les effets que peut avoir le choix des structures linguistiques contrôlées : les énoncés produits par les apprenants dans les interactions orales en salle de classe ne sont pas toujours comparables aux formes langagières canoniques utilisées dans les protocoles expérimentaux très contrôlés. Ce biais affecte, sans aucun doute, les éléments prosodiques produits par les apprenants avec un tel protocole. En outre, plusieurs études ont délaissé l'étude des facteurs extralinguistiques qui peuvent affecter également l'acquisition de la prosodie d'une L2 (l'âge, le style de parole, le type de tâche demandée, etc.). Les approches par corpus peuvent contribuer à l'analyse de tous ces facteurs d'une manière plus intégrale.

En se fondant sur l'observation d'une large base de données, l'approche par corpus permet de faire des analyses d'ordre quantitatif et qualitatif. Le corpus devient une source valide pour évaluer la fréquence et le type d'erreurs chez les apprenants (Gut 2007). Un corpus d'apprenants ayant une bonne annotation linguistique et contenant des informations non linguistiques est en mesure de fournir au chercheur une description exhaustive de plusieurs aspects de l'interlangue des apprenants. Pour garantir tout cela, une réflexion sur la collecte et type de données s'avère indispensable. Dans les travaux de phonologie/phonétique, deux grands types de données orales peuvent être compilées et analysées lors de la constitution de corpus :

- les données construites : obtenues dans des conditions expérimentales très contrôlées par l'expérimentateur afin de faire émerger des structures linguistiques particulières.
- les données authentiques : obtenues dans des conditions expérimentales peu contrôlées par l'expérimentateur où toute sorte de structures langagières peut émerger.

Les données construites sont collectées dans un contexte artificiel moyennant un protocole très précis. Elles peuvent être fabriquées par le chercheur (logatomes, pseudo-mots, phrases, énoncés ou passages de textes) ou pas (emploi de textes journalistiques, histoires, extraits de romans). Ces données sont obtenues avec des tâches de lecture oralisée, mais elles peuvent également être compilées à partir de procédés expérimentaux très précis comme les techniques visant à faire répéter selon un modèle sonore (imitation, répétition). Elles peuvent être obtenues

aussi à partir d'exercices où les participants doivent décrire des images où ils sont amenés à produire des structures linguistiques particulières.

Les données authentiques, contrairement aux données construites, sont collectées à partir de la réalisation de tâches dans des situations de communications non contrôlées. Dans ces tâches, les participants n'ont aucune contrainte sur les structures langagières qu'ils doivent employer. Les techniques employées pour éliciter ce type de données demandent la réalisation de tâches où aucune contrainte n'est demandée : une interview, un échange oral, un débat, raconter une histoire. Enfin, des exercices comme les *Map Tasks* permettent au chercheur de collecter des dialogues spontanés tout en contrôlant certains contextes discursifs ou linguistiques de son intérêt.

Lorsque l'expérimentateur a une hypothèse *a priori* à vérifier, le recours à des données construites s'impose, en tout cas dans un premier temps. Un protocole de collecte plus strict permet de mieux contrôler les structures linguistiques qui seront produites par les locuteurs et ensuite étudiées. Dans la plupart des recherches consacrées à la prosodie en L2, le recours à des données construites se révèle nécessaire puisque cela permet d'évaluer certaines hypothèses de départ (les effets du transfert de la L1 surtout, mais également les hypothèses concernant la validation d'un certain ordre acquisitionnel, entre autres). Ce type de données permet au chercheur de contrôler les contextes linguistiques et discursifs dont il a besoin, et, en conséquence, d'obtenir des données d'une manière plus souple et plus économique (en temps et en investissement). Le chercheur peut opter pour contrôler certaines informations lexicales (type de lexies, taille des mots en termes du nombre de syllabes), métriques (position de l'accent lexical, absence/présence d'une force métrique motivée par une valeur pragmatique comme le focus correctif), syntaxiques (type de construction, taille de constituants syntaxiques), ou sémantiques (contrastes de sens des mots/énoncés) afin de valider certaines hypothèses. En un mot, ce type de données fournit des informations plus claires sur le sens et la manière dont un facteur en particulier peut être corrélé au processus d'acquisition d'un aspect phonologique, phonétique ou prosodique d'une L2.

Bien que nécessaires pour tester certaines hypothèses de départ, les données construites connaissent certaines limites, surtout sur le plan qualitatif. D'une part, ces données sont limitées non seulement du fait des structures prosodiques ou segmentales qu'elles illustrent, mais aussi parce

qu'elles sont déjà orientées par l'expérimentateur lors de la collecte. De fait, les données collectées avec de tels protocoles ne peuvent mettre en évidence que des phénomènes limités et imposés au départ, en l'occurrence, la validation/rejet de l'effet d'un facteur *x* ou *y* sous les conditions *a* ou *b*. En conséquence, ces données ne permettraient pas d'évaluer facilement si d'autres facteurs peuvent motiver les patrons phonétiques observés en L2.

Un autre inconvénient de ce type de données concerne leur généralisation. Les résultats et analyses qui sont extraits de l'observation de données construites en L2 ne fournissent pas une description qualitative et quantitative de leur représentativité dans l'usage effectif de la langue (Vaguer 2007). Ainsi, ce type de données n'est pas nécessairement transposable à d'autres styles de parole : « observations from artificial speech tasks cannot always be extrapolated to natural conditions » (Leather 1999 : 32).

Les données authentiques présentent certains avantages vis-à-vis des données construites. Les exemples collectés sont le résultat d'un usage spontané d'une certaine forme ou structure sans que ces énoncés soient biaisés par une hypothèse établie *a priori*. Les données authentiques connaissent aussi des limites. D'une part, la fréquence et la densité d'une forme/structure qui intéresse le chercheur ne sont pas garanties. D'autre part, même si une forme *x* ou structure *y* apparaissent avec une fréquence et une relative densité dans le corpus, leur distribution par rapport au contexte linguistique dans lequel celles-ci apparaissent ne sont pas nécessairement pertinentes pour les objectifs de la recherche. Enfin, si les données construites ne garantissent pas une description neutre de par l'hypothèse que le chercheur leur a imposée, les données authentiques ne la garantissent pas totalement non plus. En effet, on peut faire le recensement d'une certaine forme/structure dans le corpus, une tâche relativement neutre en elle-même ; en revanche, « Ce qui n'est pas neutre, c'est ce que l'on fait de ce recensement [...] on élimine ce qui paraît redondant, du même type ; on garde ce qui semble le plus propre à illustrer ce que l'on veut dire, mais on ne signale pas ce sur quoi on n'a rien de particulier à observer [...] » (Vaguer 2003 : 212)

Nous considérons que les données authentiques et construites en L2 ne s'opposent pas mais se complémentent. Passer par l'analyse de deux types de données est nécessaire pour la construction/validation de certaines hypothèses. La compilation des données authentiques ou des

données construites font partie de ce qu'on entend par *corpus*. Il faut noter cependant que, selon Delais-Roussarie (2003), le terme *approche sur corpus* est réservé surtout à des études utilisant des données authentiques : des échantillons de langue non fabriqués élicités selon des protocoles de collecte peu contrôlés. Cependant, une collection de données construites peut également constituer un corpus. Nous considérons que toute *approche sur corpus* repose sur l'analyse de données authentiques à laquelle peuvent s'ajouter ou non des données contrôlées.

Rares sont les corpus destinés à l'étude de l'acquisition de la phonétique et de la phonologie du français L2. De fait, ce n'est que très récemment que ce genre de corpus commence à être constitué. Parmi ces corpus, nous pouvons citer le corpus IPFC (Interphonologie du Français Contemporain) consacré à l'étude de la prononciation du français L2 (Racine *et al.* 2012 ; Durand *et al.* 2009). Ce corpus utilise des tâches similaires à celles du corpus PFC (Phonologie du Français Contemporain), lequel est destiné à l'étude phonologique des variétés du français contemporain (Detay *et al.* 2010 ; Durand & TARRIER 2006). Le corpus IPFC comprend des tâches de répétition d'une liste de mots, des tâches de lecture, un entretien avec un locuteur natif et une interaction semi-contrainte entre deux apprenants. Ce corpus a été conçu pour l'étude de la prononciation au niveau segmental, et n'est, de ce fait, pas toujours adapté pour l'étude de la prosodie.

La constitution du corpus COREIL (Delais-Roussarie & Yoo, 2011) a tenté de répondre au manque de corpus disponibles de nos jours pour l'étude de la prosodie du français L2. Ce corpus repose sur certains principes :

- Représentativité des données : le corpus a été compilé à partir d'extraits de parole comprenant autour de 10 mille mots en français L2 (30 participants), 2,7 mille mots en français L1 (dix locuteurs) et 4 mille mots dans différentes L1.
- Profil des locuteurs : ont été enregistrés des apprenants adultes du français L2 en contexte d'apprentissage au milieu universitaire ayant différentes L1 (espagnol, anglais, grec, coréen, allemand) ainsi que des locuteurs natifs monolingues du français et des locuteurs natifs de toutes les L1 des apprenants.
- Diversification des tâches : le corpus a été constitué à partir de l'application de tâches diverses, à savoir, production orale

monologuée (description d'une image), production orale semi-directive (interview), interaction orale (jeu de rôles), lecture de scripts.

- Niveau de maîtrise de la L2 : dans le but de voir s'il y a un ordre dans l'acquisition des formes prosodiques, deux niveaux de maîtrise dans la langue cible ont été pris en compte pour les apprenants (A2 vs B1).

Le protocole du corpus COREIL permet également de traiter une partie des données collectées, même si le processus de collection n'est pas encore terminé. Il est aussi possible d'ajouter des données supplémentaires ou bien des tâches sans perdre l'homogénéité de l'ensemble du corpus. En tout état de cause, nous avons proposé que la collecte du corpus oral d'apprenants englobe nécessairement des données construites et artificielles (enregistrées à partir de la résolution des plusieurs tâches de production orale) et des données authentiques. La compilation d'un corpus suffisamment grand et varié pourrait ainsi permettre de surmonter plusieurs problèmes méthodologiques dans les études d'acquisition de la phonologie d'une L2, en favorisant l'émergence d'une plus grande variété de structures langagières. Dans les sections qui suivent, nous discuterons de l'importance des tâches impliquées dans l'analyse de la prosodie en FLE et son annotation.

2.2 Transcription orthographique

Un corpus oral regroupe des documents qui correspondent à des transcriptions de productions orales alignées ou non avec le signal acoustique. Toute transcription orthographique de données orales en vue d'analyses phonético-phonologiques ou prosodiques doit fournir une image fidèle de ce qui a été dit. Transcrire des données sonores consiste à fournir une représentation symbolique du signal, mais cette représentation n'est pas équivalente au signal, dans la mesure où elle est le résultat d'une analyse, ou plutôt d'une abstraction, des données réelles (Delais-Roussarie 2003). Ainsi, la transcription orthographique peut être vue comme un lien entre les productions orales effectives et ce qui en est compris par l'annotateur.

Transcrire orthographiquement le discours oral nécessite une réflexion théorique et méthodologique à la fois (Bilger 2007). Selon Blanche-Benveniste & Jeanjean (1987), toute représentation écrite de l'oral exige certaines conventions d'écriture. D'autres auteurs consentent à cette idée

mais proposent également que ces conventions doivent refléter le caractère oral du corpus (Durand & Tarrier 2006). Durant ce processus de transcription, de nombreux problèmes apparaissent lorsque l'on veut représenter orthographiquement l'oralité (Blanche-Benveniste & Jeanjean 1987) :

- reconstruction de certains éléments inexistantes sur le signal (transcrire *il y a* à la place de *i y a*, *il pleut* à la place de *i pleut*) ;
- choisir une représentation écrite nécessaire de lever des ambiguïtés dues à l'oral – la séquence [ilpɑvɛlfvɑsɛ] peut correspondre aux suites orthographiques *ils parlent français* ou *il parle français*.

Dans le cas des corpus d'apprenants, s'ajoutent d'autres difficultés : le transcripteur peut se voir influencé par le crible phonologique de sa propre L1 et faire des interprétations erronées lorsqu'il ne connaît pas le système de la L1 de ces derniers (Racine *et al.* 2011). La tâche de transcription n'est pas du tout aisée, surtout lorsqu'il s'agit d'annoter la parole non native contenant « des réalisations déviantes ». Ainsi, par exemple, la suite [jenetraβajpa] en français L2 produit par un hispanophone dans une tâche de production orale spontanée correspondrait, très certainement à *je ne travaille pas* et non à *je né travaille pas*, ou *j'ai n'est travaille pas*.

Pour essayer de surmonter ces problèmes d'ordre méthodologique et théorique, le chercheur peut transcrire orthographiquement les éléments contenus dans le signal à partir des unités existantes en L1 et éviter au maximum de faire des interprétations. Pour cela, le chercheur peut effectuer la transcription de l'ensemble du corpus moyennant une adaptation des recommandations faites par le *Text Encoding Initiative – & Eagles* (TEI Consortium) et des conventions suggérées par le système CHAT de CLAN (MacWhinney 2000). Pour le corpus COREIL, ces conventions peuvent se résumer comme suit :

- Emploi de la ponctuation forte : seuls le point « . », le point d'interrogation « ? » et d'exclamation « ! » ont été retenus, et on a évité l'emploi des virgules, parenthèses et tout autre symbole employé dans le style scriptural conventionnel.
- Emploi de l'orthographe standard, même si le locuteur n'a pas réalisé tous les segments : [ʃsɛpa] est transcrit *je (ne) sais pas* les parenthèses indiquant les segments canoniques non réalisés.

- L'insertion d'une pause a été codée avec le symbole « (.) ».
- Le nombre de répétitions des mots ou des suites de mots a été encodé avec « [x n] », où « n » représente le nombre de répétitions de la suite en question : [zəʒəməpɛl] a été transcrit *je [x 2] m'appelle*.
- Les faux départs ont été indiqués avec le symbole « [/ ?] » : la suite [zəsɥiazəsɥizale] a été transcrite *je suis a [/?] je suis allé*.
- Les fragments inintelligibles ont été transcrits avec « xxx ».

En plus des conventions ci-dessus, la transcription des erreurs des apprenants mérite d'être discutée dans le corpus COREIL. Pour les hispanophones, si la morphologie a été identifiée comme correcte mais sa réalisation phonétique déviante, nous avons utilisé la représentation orthographique standard : [zəβɛosinema] est transcrit *je vais au cinéma* et non *je bé au cinéma*. Si la morphologie est identifiée comme incorrecte mais sa réalisation est existante dans la L2, nous avons assumé une certaine connaissance morphologique de l'apprenant dans la L2 et avons fait une adaptation orthographique en fonction de ce qui reflète au mieux son interlangue : [mõpɛvaneameksiko] est transcrit *mon père a né à Mexico*. On peut supposer que l'apprenant fait une surgénéralisation de l'auxiliaire *avoir* lors de la construction du passé composé. Si l'apprenant emploie sa L1, les mots sont transcrits dans le système orthographique conventionnel de sa L1 : [jetudibjolo'xia] est transcrit *j'étudie biología@s:sp* (l'étiquette *@s:spa* signale que le mot est produit en espagnol, L1 de l'apprenant). Si l'apprenant a rajouté des phonèmes non spécifiés par le système morphosyntaxique de la L2, nous avons encodé ces éléments avec une adaptation de la norme orthographique de la L2 : [ləɔm] est transcrit *le homme*.

La Figure 1 représente une copie d'écran de l'interface de CLAN montrant un exemple de transcription orthographique dans les lignes principales (*S04 indique le code associé à l'apprenant) et les lignes secondaires recevant un codage linguistique, en l'occurrence ici le type de phrase (cf. Santiago & Delais-Roussarie 2015 pour plus de détails). Ces exemples sont tirés de la tâche de lecture. Les surlignés jaunes indiquent que l'extrait est aligné avec la portion du signal sonore (oscillogramme en bas).

L'alignement entre la transcription orthographique et le signal de parole mérite une attention particulière. Chaque ligne dans l'éditeur correspond à ce qu'on appelle ici un « énoncé », c'est-à-dire une séquence

correspondant à une clause (phrase verbale entourée de son sujet et de tous les éléments disjoints à sa gauche/droite qui en dépendent). Cette définition de l'énoncé a permis d'aligner aisément la transcription orthographique avec le signal de parole dans les tâches de lecture oralisée d'autant plus que la ponctuation forte coïncidait avec une frontière syntaxique majeure. En revanche, cette notion d'énoncé se révèle très problématique lorsqu'il s'agit d'aligner la transcription orthographique et le signal de parole dans les tâches de production orale non contrôlées (descriptions d'images, interview, etc.). En effet, découper en séquences le signal sonore afin d'apparier ces extraits à ce qu'on entend par « énoncés » ou « clauses » est extrêmement difficile. Nous avons donc opté d'aligner la transcription orthographique avec la séquence formée par une clause, même si à l'intérieur de celle-ci des pauses y figuraient. Par exemple, dans la séquence « on est allé en Normandie (.) euh donc j' ai essentiellement vécu en Normandie (.) Alançon (.) Falaise (.) Saint+lô +... » produit par un locuteur français natif en parole spontanée, nous avons

1 @Begin
 2 @Languages: fra
 3 @Participants: S04 Student
 4 @ID: fra|FrenchL1|S04|19;|male|French_Native_Speakers||Participant|High_School|Bachelor_Student|
 5 @Birth of S04: 21-OCT-1992
 6 @Date: 18-MAR-2011
 7 @Location: Paris, LLF-Paris7.
 8 @Situation: Reading task 01
 9 @Media: S04_FREL1_READ_1 audio
 10 *S04: il faut que je sois à Grenoble samedi vers quinze heures. •
 11 %spa: asser
 12 *S04: pourriez vous m' indiquer à quelle heure est la correspondance
 13 à Valence? •
 14 %spa: quest_tot_inv
 15 *S04: si je dois partir avant midi de Marseille j' aimerais savoir
 16 s' il+y+aura un wagon restaurant. •
 17 %spa: asser
 18 *S04: il paraît que sur cette ligne la S_N_C_F prévoit un wagon de
 19 jeux pour les enfants. •
 20 %spa: asser
 21 *S04: est+ce+que c' est vrai? •
 22 %gra: quest_tot_est
 23 *S04: vous prenez les réservations par téléphone? •
 24 %gra: quest_tot_dec
 25 @End

20mar12[E|CHAT] * 17

Fig. 1. Copie d'écran de l'éditeur de texte sous CLAN montrant l'alignement de la transcription orthographique et le signal de parole en français L2.

distingué deux clauses, les « (.) » représentant des pauses : *on est allé en Normandie et euh donc j'ai essentiellement vécu en Normandie Alençon Falaise Saint Lô.*

Bien que la question de la segmentation en unités fonctionnelles pour la transcription orthographique gagnerait à être approfondie et discutée (cf. Benazzo & Wątorrek dans ce volume), nous ne nous attarderons pas beaucoup sur ce point. Dans notre cas, la transcription orthographique et son alignement avec le signal sonore ne sont là que pour faciliter l'analyse des données *a posteriori* pour les études prosodiques et phonétiques. Il faut en tout cas remarquer que la transcription orthographique passe nécessairement par une étape d'analyse et d'interprétation du signal sonore. Cette interprétation se révèle moins problématique lorsqu'il s'agit de transcrire les tâches de lecture. Dans ce cas, en effet, les participants sont censés reproduire ce qui est déjà écrit. Dans le cas de la parole spontanée, reconstruire ce qui a été dit dans le signal de parole est une tâche plus compliquée, surtout lorsqu'il s'agit de la parole des apprenants.

2.3 Annotations phonologiques

Même si les données ont été transcrites au niveau orthographique, il peut être utile d'avoir accès à des informations relatives à la forme sonore. Se pose alors la question de savoir quelles informations il faut extraire du signal de parole : les réalisations phonétiques concrètes ou plutôt les formes phonologiques ? Lorsque le chercheur opte pour la transcription phonologique, cela suppose qu'il a déjà découvert le système phonologique de la variété sur laquelle il travaille (Durand & TARRIER 2006 : 141). Si cela n'est pas forcément un problème en L1, ce n'est pas la même chose lorsqu'on analyse la parole des apprenants, le système phonologique qui est sous-entendu est en effet émergent et méconnu par le chercheur. Puisque nous travaillons sur des données d'apprenants, nous ne pouvons pas à proprement parler de faire une transcription phonologique. Nous avons plutôt opté d'adapter une transcription phonétique de la L1 et avons rejeté une transcription phonétique authentique qui caractérise la prononciation des étudiants.

Il faut noter, cependant, qu'une transcription phonétique authentique de l'ensemble des données procure plusieurs avantages. D'une part, elles fournissent des représentations symboliques de ce qui a effectivement été prononcé par les locuteurs non natifs, et non une abstraction phonologique

qui serait forcément erronée puisque le système des apprenants n'est pas connu. D'autre part, elles rendent mieux compte de ce qui a été dit par les apprenants que la transcription orthographique. En outre, avoir une transcription phonétique authentique permettrait au chercheur de croiser des informations plus fines : les compétences prosodiques en L2 sont corrélées au degré de maîtrise de la prononciation des consonnes et des voyelles ?

Cependant, s'engager dans une transcription phonétique des données non natives à partir de l'oreille du transcripteur entraîne une grande subjectivité, d'autant quand le transcripteur ne connaît pas les possibles formes déviantes en L2. Entreprendre la même chose à partir des informations acoustiques enlève la subjectivité, mais cette tâche est assez lourde et coûteuse en temps.

L'annotateur est confronté au choix du degré de finesse qu'il veut atteindre dans la transcription phonétique. Pour s'en convaincre, voilà quelques exemples. Observons les spectrogrammes des suites *m'indiquer* et *si je dois partir* par deux apprenants hispanophones du niveau A2 et B1 respectivement (tâche de lecture oralisée).

Dans la Figure 2, une analyse portant sur la première syllabe de la suite *m'indiquer* nous permet de constater que la production canonique [médike] est remplacée par [maɲdike]. Cette figure illustre bel et bien l'absence d'anti-formants nasals caractéristiques du [ɛ̃] et montre plutôt

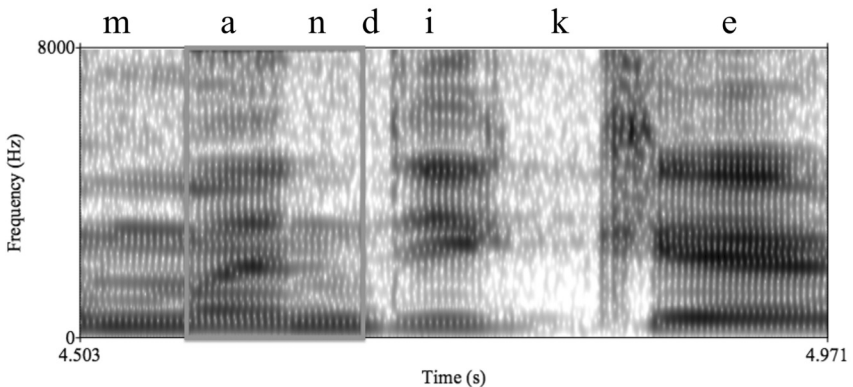


Fig. 2. Spectrogramme de la suite transcrite comme *m'indiquer* en français L2 produite par une locutrice FL2 niveau A2, le carré enfermant les phonèmes [a] et [n].

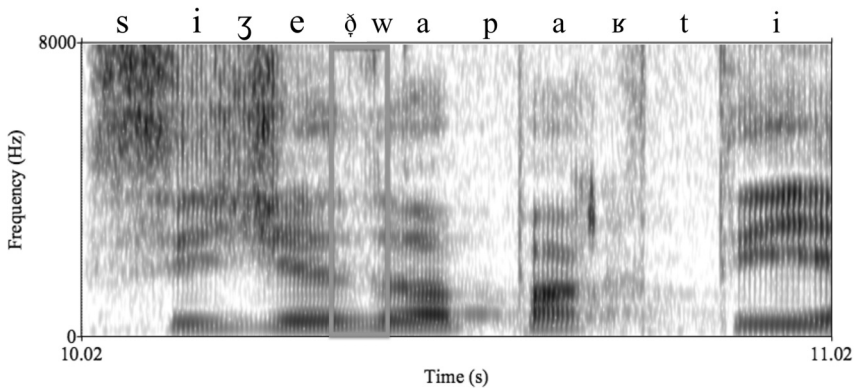


Fig. 3. Spectrogramme de la suite transcritte comme si je dois partir en français L2 produite par un locuteur FL2 niveau B2, le carré enfermant le phone [ø] typique de l'espagnol.

l'articulation de la voyelle [a] qui propage ses transitions formantiques de F1, F2 et F3 vers la nasale [ɲ] qui, à son tour, est dentalisée avec le contact du [d]. Dans la Figure 3, l'articulation du mot *dois* de la suite *si je dois partir* montre l'absence de l'occlusion dentale, celle-ci étant remplacée par une aperture plus ou moins importante des organes articulatoires, en l'occurrence, l'apex de la langue et les dents supérieures. La présence des transitions formantiques de la voyelle [e] précédente nous conduit à affirmer que cette consonne a été articulée comme une approximante dentale [ø] et non comme une occlusive [d] (sans doute à cause d'un transfert de l'espagnol).

Transcrire avec fidélité les deux segments ci-dessus devient une tâche extrêmement lourde lorsqu'il s'agit d'un large corpus comme celui qui est traité ici où plus de 100 mille phonèmes sont à encoder manuellement avec ce degré de finesse ! Face à ce problème, les réalisations phonétiques des apprenants peuvent être encodées à partir des informations orthographiques, c'est-à-dire à partir d'une phonétisation reposant sur la phonologie du système. Ainsi, les segments peuvent être transcrits en n'utilisant que les symboles phonétiques du système français. Ainsi, la transcription donnée pour la suite illustrée dans la Figure 2 a été [mẽdike] et non [maɲdike], et pour la Figure 3 [siʒədwapaɾti] et non [siʒeøwapaɾti].

Cette méthode présente un avantage lorsque la transcription doit se faire de façon semi-automatique. A partir de la tire contenant la transcription

orthographique sous CLAN, on peut générer des tiers sous le logiciel *Praat* (Boersma & Weenink 2018) afin de générer des transcriptions phonétiques automatiques. Par exemple, une segmentation en mots, syllabes et phones en employant le script *Easyalign* développé par Goldman (2011) peut être menée dans l'ensemble de données en L2. La Figure 4 illustre donc les 5 tires générées sur *Praat* à partir d'*Easyalign*. Nous tenons à faire remarquer que la transcription des mots *parti* et *Marcelle* est ici un exemple des formes morphologiques et prononciations erronées en L2 des mots *partir* et *Marseille* respectivement. Or, selon les conventions sur la transcription de l'orthographe adoptées dans la section 2.2, nous fournissons ici une transcription orthographique la plus proche de ce que l'apprenant a prononcé, mais en utilisant des mots existants en français L1. Les transcriptions phonétiques correspondent donc aux mots en français L1 *parti* et *Marcelle*.

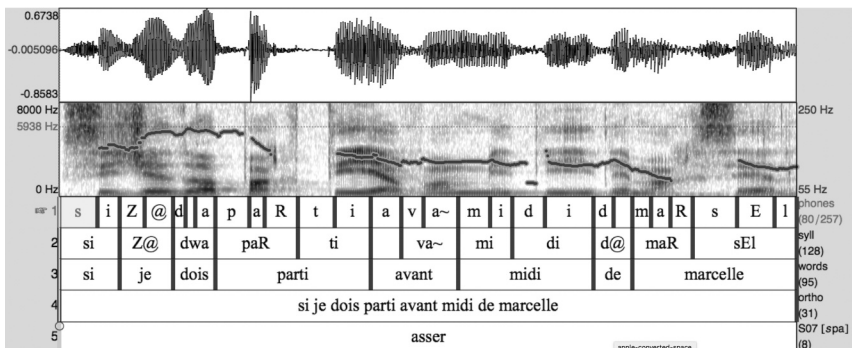


Fig. 4. Copie d'écran de Praat montrant différentes tires encodant les annotations syntaxiques (5), l'orthographe (4), les mots (3), les syllabes (2) et les phones (1) d'un extrait de parole en français L2.

3. Transcrire la prosodie en FLE

Une annotation prosodique consiste à fournir une représentation symbolique et discrète des phénomènes prosodiques observés dans le flux de parole. Lors de la transcription, une grande importance est accordée aux événements qui ont des fonctions linguistiques. Ainsi, l'annotation prosodique doit fournir un encodage symbolique d'une grande variété d'événements tels que les accents, les contours intonatifs, les pauses, etc.

Ces derniers se manifestent généralement à des endroits clés de la structure prosodique. Sur le plan phonétique, ces événements sont indiqués par des variations de la fréquence fondamentale (f_0), dont la force et l'ampleur jouent un rôle, et par des modifications de paramètres temporels comme la durée syllabique.

Pour les données d'apprenants, la transcription envisagée doit permettre d'analyser les systèmes accentuels et intonatifs sous-jacents qui évoluent durant l'apprentissage d'une L2 (Delais-Roussarie & Yoo 2011). Autrement dit, cette annotation doit rendre compte des connaissances de la grammaire de la L2 acquises, des éventuelles règles de la L1 se transposant dans la L2, et pour finir, des propres règles développées dans l'interlangue des apprenants.

Afin d'obtenir une annotation prosodique fournissant un codage symbolique d'une grande variété d'événements prosodiques pour chaque énoncé analysé, le chercheur doit faire face à plusieurs problèmes. Ceux-ci sont dus en grande partie au fait que la transcription des phénomènes prosodiques est très souvent un point épineux largement débattu parmi la communauté scientifique. Cela peut s'expliquer par plusieurs raisons :

- l'étude de l'accentuation ou l'intonation est souvent compliquée par le fait que ces phénomènes sont continus, et donc difficiles à représenter de façon discrète (Post & Delais-Roussarie, & Simon 2006) ;
- les caractéristiques prosodiques varient en fonction d'éléments comme le débit ou l'origine sociogéographique des locuteurs (Post & Delais-Roussarie 2006) ;
- lors d'une transcription prosodique, il faut déterminer quels paramètres acoustiques (f_0 , intensité et durée) y sont représentés (Delais-Roussarie & Yoo 2011).

Transcrire les événements prosodiques est souvent plus compliqué que coder les segments. Cela résulte sans doute du fait que pour la transcription des segments, qu'il s'agisse d'une analyse phonétique (au niveau des phones) ou phonologique (au niveau des phonèmes), il existe un système de transcription qui semble être accepté par la communauté scientifique : l'API (Alphabet Phonétique International). En revanche, pour les transcriptions prosodiques, aucun système ne fait consensus.

3.1 Quels systèmes de transcription choisir pour la prosodie en L2 ?

L'API comprend un volet qui permet d'encoder différents phénomènes, tels que certains contours mélodiques ou les patrons accentuels. Mais employer le système de transcription prosodique de l'API suppose en général deux choses : (i) le système prosodique à transcrire est connu par l'annotateur ; (ii) les phénomènes prosodiques à annoter sont linguistiquement pertinents. Ces deux caractéristiques font penser que l'emploi de ce système pour l'annotation des événements prosodiques en L2 n'est pas opérationnel : le système prosodique des apprenants est méconnu par l'annotateur et il est difficile de distinguer les phénomènes d'ordre phonologique *vs* phonétique.

D'autres systèmes comme ToBI (*Tones and Break Indices*) ont connu une grande popularité dans la communauté scientifique. Ce système a été développé initialement pour coder l'intonation de l'anglais américain standard (Beckman & Hirschberg 1994) en distinguant deux types d'informations :

- Le codage des tons selon la théorie métrique et autosegmentale : les contours mélodiques sont représentés comme une séquence linéaire des cibles tonales distinctives représentées avec les symboles L (*Low*)/H (*High*) et leurs combinaisons (HL, LH, etc.).
- Le codage des *Break Index* reflètent la structure prosodique des langues : les unités prosodiques, tels que les groupes accentuels (AP), les syntagmes intermédiaires (ip) ou les syntagmes intonatifs (IP) sont indiqués par des indices numériques apparaissant aux frontières. En outre, ils sont considérés comme se manifestant dans des événements intonatifs qui portent des diacritiques relatifs à leur position dans la hiérarchie prosodique : « * » pour les accents mélodiques ou *pitch accents* (T*), « - » pour les accents de phrases indiquant la frontière des ip (T-), et « % » pour les tons de frontières liés aux IP (T%) (Beckman & Pierrehumbert, 1986).

Ce système est à l'heure actuelle le plus utilisé pour le codage de l'intonation et de la structuration prosodique dans plusieurs langues, dont le français (Delais-Rousarie *et al.* 2015, Jun & Fougeron, 2000), et l'espagnol (de la Mota, Martín Butragueño & Prieto, 2010 principalement) entre d'autres dizaines de langues (*cf.* Jun 2014). Utiliser ce type de

système suppose plusieurs éléments, notamment l'acceptation des concepts centraux de la théorie métrique et autosegmentale que nous n'allons pas détailler ici. En outre, il s'agit d'un système dit phonologique dont l'emploi repose sur quelques conventions clairement établies. Ainsi, les conventions d'annotation doivent témoigner de :

- l'exactitude : quelle que soit la langue annotée, les annotations doivent être fondées sur une analyse rigoureuse de la phonologie intonative, une analyse phonologique de la langue en question étant donc un prérequis.
- l'efficacité : l'annotateur ne devrait transcrire/encoder que les contours mélodiques non distinctifs, qu'ils soient montants ou descendants.

Bien que dans certaines études le système ToBI ait été employé pour coder les événements prosodiques de plusieurs L2 afin de les comparer à la langue cible (Jilka 2000 pour l'allemand L2, Ueyama & Jun 1996 pour l'anglais L2, Gabriel & Kireva 2014 pour l'espagnol L2, Mennen 2004 pour le grec L2.), plusieurs problèmes sont à noter. Le système ToBI repose sur une analyse abstraite. De fait, seuls les événements intonatifs ayant une fonction contrastive sont encodés. En tous cas, cela suppose

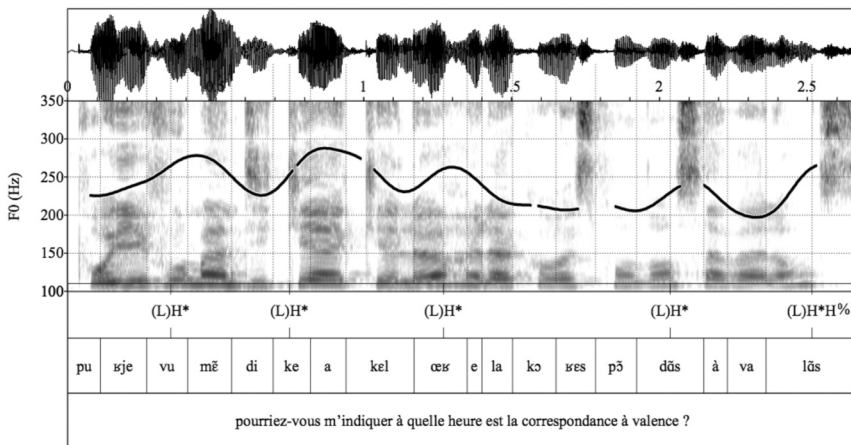


Fig. 5. Exemple d'annotation de parole en français L1 avec le système ToBI French (Delais-Roussarie et al. 2015) d'un énoncé interrogatif.

que le système intonatif de la langue de l'apprenant soit connu avant que l'annotation prosodique se fasse.

Des exemples de difficultés sont discutés dans les figures qui suivent. Dans la Figure 5, le découpage en groupes accentuels en français L1 est indiqué par la présence des accents mélodiques finals. Les symboles (L)H* représentent un mouvement mélodique dynamique montant à la fin de groupes accentuels (pourriez-vous)_{GA} (m'indiquer)_{GA} (à quelle heure)_{GA} (est la correspondance)_{GA} (à Valence ?)_{GA}, et le mouvement H% réalisé à la fin de l'énoncé indique la frontière du syntagme intonatif (en l'occurrence, une question).

Ce codage, et surtout son analyse en unités prosodiques, découle de la phonologie du français et des connaissances connues par l'annotateur : (i) en français, il y a des accents mélodiques (T*) qui marquent la frontière des groupes accentuels, (ii) le phrasé prosodique contraint le choix du contour intonatif (la force du mouvement mélodique dépend de son statut dans la hiérarchie prosodique, selon qu'il s'agisse d'un groupe accentuel ou un syntagme intermédiaire/intonatif, et (iii) en position position prénucléaire, les mouvements mélodiques sont généralement montants (H*, H- ou H%). Ils sont encodés par les symboles LH ou H (Delais-Roussarie, et al. 2015, Di Cristo 2016).

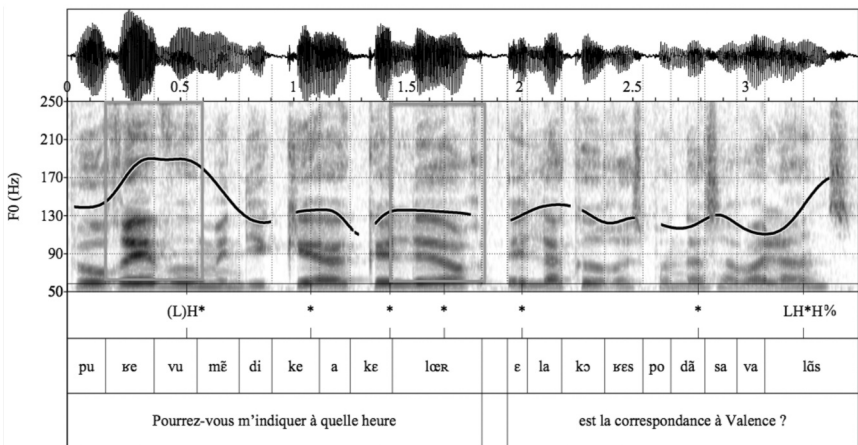


Fig. 6. Exemple d'annotation de parole en français L2 avec le système ToBI French d'un énoncé interrogatif produit par un apprenant hispanophone.

Que faire en revanche pour encoder les productions des apprenants qui ne suivent pas nécessairement les règles phonologiques de la langue cible ? Si on compare la réalisation d'un locuteur natif illustré dans la Figure 5 à celle d'un apprenant hispanophone dans la Figure 6, on voit au moins deux différences importantes. Dans la Figure 6, il y a un mouvement mélodique sur la syllabe finale du mot *pourriez*, mais également un ton haut sur la syllabe *vous* (premier rectangle rouge).

Comment traiter cette réalisation ? Comme une séquence de deux accents mélodiques sur deux syllabes adjacentes (LH* suivi de H*) ou comme un accent initial suivi d'un ton haut ? S'agit-il d'un transfert concernant un positionnement erroné de l'accent de groupe qui est sur la pénultième en espagnol L1 ? Les choix ne sont pas simples, d'autant qu'ils ont des implications sur l'analyse des faits. De la même façon, comment doit-on encoder l'absence de mouvement mélodique montant sur la syllabe finale de « indiquer » et de « heure » (deuxième rectangle dans la Figure 6) : par une absence d'accent mélodique qui va de pair avec une absence de réalisation des groupes accentuels ? Cela refléterait en partie les patrons mélodiques en vigueur dans la L1 de l'apprenant, et pourrait résulter d'un transfert (*cf.* Santiago & Delais-Roussarie 2015 où l'absence d'accents mélodiques en position interne dans l'énoncé interrogatif en espagnol est fréquente). Mais on peut aussi penser que les apprenants réalisent un accent à l'aide de la durée et non pas avec un mouvement mélodique, et que les découpages des groupes accentuels seraient alors satisfaisants.

Cela nous amène au deuxième problème avec le système ToBI. Les étiquettes et événements retenus sont limités surtout aux variations mélodiques de f0. Cela écarte par exemple le codage des phénomènes liés à des variations de durée ou d'intensité, paramètres acoustiques pouvant être importants pour le codage de différents contours mélodiques en français L1 (Martin 2012) ou exploités différemment par les apprenants d'une L2. Sur ce dernier cas, par exemple, à l'écoute de l'exemple illustré dans la Figure 6, l'apprenant semble marquer le phrasé prosodique (deuxième rectangle) du groupe accentuel à *quelle heure* par un allongement de la voyelle de la syllabe finale du mot *heure* à la place de la modulation de la voix, celle-ci étant la plus exploitée en français L1. De fait, l'emploi de l'allongement syllabique pour marquer les groupes accentuels peut être vu comme licite, car l'allongement des syllabes est relié au marquage prosodique des phénomènes d'ordre métriques, et non pas intonatifs.

Le même cas peut être vu dans la Figure 7, où un allongement de la syllabe finale du groupe accentuel *nous avons mangé* est observé. Ces paramètres prosodiques doivent être considérés dans une transcription phonétique en L2 afin d'évaluer dans quelle mesure les locuteurs non natifs exploitent les paramètres acoustiques de la même façon que les natifs. Ces exemples montrent bel et bien que le marquage des unités prosodiques, tels que les groupes accentuels, peut émerger exclusivement avec les variations temporelles en L2. Or, le système ToBI pour le français ne considère pas ces facteurs.

Pour finir, comme dans ce système d'annotation la définition des unités prosodiques repose tantôt sur des informations morphosyntaxiques tantôt sur les formes tonales des mouvements mélodiques, cela peut compliquer les annotations prosodiques de certaines réalisations observées chez les apprenants. En effet, il est généralement admis que la force d'un mouvement mélodique montant associé à la frontière droite d'un groupe accentuel en français a une importance mineure comparée à celle des contours montants bornant les énoncés interrogatifs ou les syntagmes intonatifs en position non finale associés à ladite « continuation majeure ». En d'autres termes, la force de la pente mélodique des contours montants bornant un syntagme intonatif est généralement plus ample que celle d'un accent mélodique bornant un groupe accentuel (Di Cristo 2016). Prendre ces critères pour encoder la prosodie pose un problème lorsqu'il s'agit d'encoder la parole des apprenants.

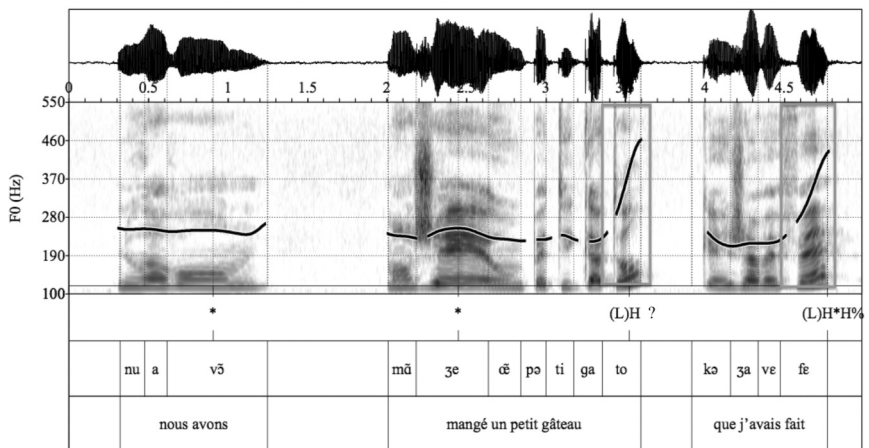


Fig. 7. Exemple d'annotation de parole en français L2 avec le système ToBI French Delais-Roussarie et al. (2015) d'un énoncé assertif.

Par exemple, dans la Figure 7, la force de la montée mélodique du groupe accentuel (un petit gâteau)_{GA} est plus ample que celle bornant la frontière droite d'un syntagme intonatif non final [(nous avons mangé)_{GA} (un petit gâteau)_{GA} (que j'avais fait)_{GA}]_{SI}. Or, il est clair qu'en français une montée exagérée bornant un groupe accentuel dans une position pré-nucléaire n'est pas communément réalisée comme telle dans les données des natifs. L'annotateur se voit donc dans l'incertitude d'encoder cette montée exagérée comme la réalisation d'un groupe accentuel ou celle d'un syntagme intonatif. Cela aurait des implications sur l'analyse : erreur d'implémentation phonétique des contours montants des groupes accentuels ou erreur phonologique relative aux découpages prosodiques ?

Pour toutes ces raisons, nous avons considéré qu'utiliser un système ToBI adapté au français n'était pas convenable pour le codage de nos données en français L2. En revanche, nous considérons qu'une adaptation de ce système doit être envisagée si l'on veut comparer les réalisations natives et non natives.

En plus des deux systèmes de transcription mentionnés ci-dessus, phonologiques dans leur essence, d'autres systèmes de transcription sont utilisés pour encoder les événements prosodiques, dont des systèmes automatiques, notamment le système Momel-INTSINT (*International Transcription System for INTonation*) développé par Hirst (2007). Faute de place, nous renvoyons le lecteur à l'ouvrage de Santiago (2014) afin de réfléchir aux avantages et inconvénients de ce système de transcription.

Les problèmes d'utiliser des systèmes d'annotation tels que l'API ou ToBI en français L2 posent également un problème de par le fait que la précision des techniques d'annotation automatique utilisant ces systèmes d'annotation n'est encore pas satisfaisante, et que le travail d'annotation est donc nécessairement manuel. Bien que certains travaux essaient d'explorer l'automatisation de ces systèmes de transcription (Rosenberg 2010), à notre connaissance, ces systèmes n'ont pas encore été testés dans la parole non native. En revanche, d'autres outils d'annotation prosodique, mais non des systèmes d'annotation à proprement parler, peuvent pallier ces problèmes.

3.2 Vers un outil semi-automatique pour transcrire la prosodie en FLE

Un outil d'annotation prosodique pour des données non-standards a comme objectif de fournir un étiquetage automatique des patrons prosodiques du signal de parole lorsque les systèmes accentuels et intonatifs

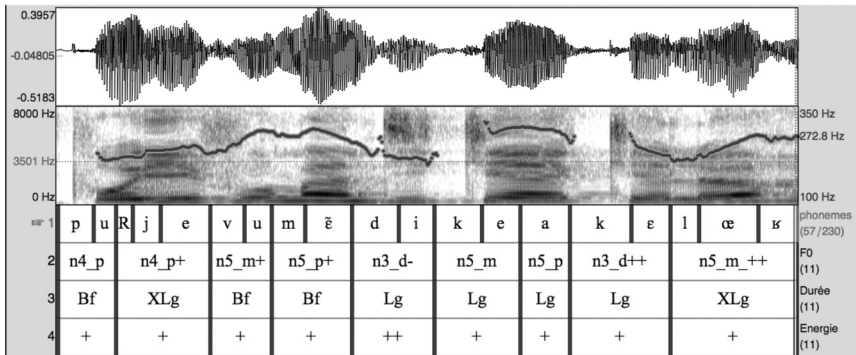


Fig. 8. Codage fournit par PROSOTRAN de l'énoncé Pourriez-vous m'indiquer à quelle heure en français L1.

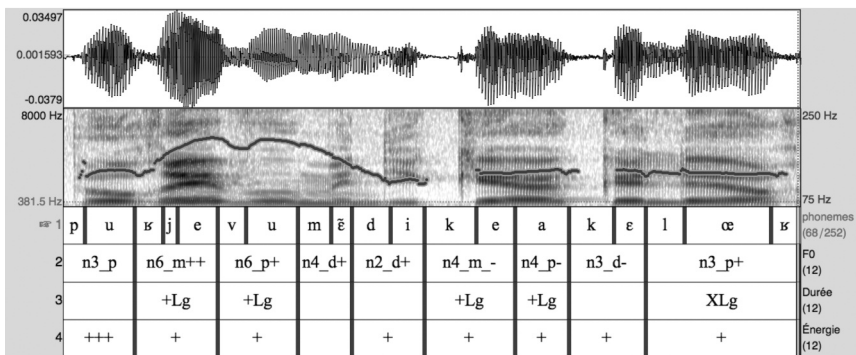


Fig. 9. Codage fournit par PROSOTRAN de l'énoncé Pourriez-vous m'indiquer à quelle heure en français L2.

des langues à annoter sont méconnus par le transcritteur. L'un de ces exemples et l'outil PROSOTRAN (Bartkova & Delais-Roussarie & Santiago 2012). Il a été conçu pour faire des comparaisons entre le codage prosodique des données en L2 et celui des données en L1. Cet outil est encore en phase de développement. PROSOTRAN fournit un codage de trois paramètres acoustiques : (i) les variations mélodiques de f0 (en termes de hauteur tonale, de direction du mouvement mélodique par rapport à la syllabe précédente et d'ampleur du mouvement selon le seuil du glissando) ;

(ii) les variations temporelles et (iii) les variations d'intensité. Un exemple de l'annotation que fournit PROSOTRAN pour le codage des énoncés est illustré dans les Figures 8 et 9. Dans ces figures, dans la tire « F0 », nous trouvons les informations concernant les variations mélodiques, en l'occurrence, le niveau atteint (codées comme n3, n6... , n= niveau tonal), la direction du mouvement (p = plateau, m=montant, d=descendant) et le seuil du glissando (+, ++...). Dans la tire « Durée », nous trouvons l'étiquetage du rallongement vocalique : bref (Bf), Long (Lg) ou extralong (XLg). Enfin, dans la tire 4, nous trouvons le codage concernant l'énergie (+, ++...).

Le codage fourni par PROSOTRAN permet de faire émerger des événements prosodiques se réalisant par des variations affectant plusieurs paramètres conjointement. Cela permet au chercheur d'analyser les indices acoustiques qui peuvent être exploités différemment par les locuteurs venant de différents profils linguistiques, et ce, de manière indépendante. C'est l'exemple du marquage prosodique du groupe accentuel à *quelle heure* : en français L1, cette syllabe est encodée avec un mouvement mélodique montant (seuil de glissando) et un allongement important, tandis qu'en français L2, la même syllabe ne reçoit aucun étiquetage concernant un mouvement de la direction de f0 (plateau) mais seulement celui de la variation temporelle (syllabe extra-allongée). Comme nous pouvons le constater, un outil de transcription comme celui qui est présenté ici permettrait de pallier certaines insuffisances que nous avons rencontrées dans beaucoup de systèmes de transcription prosodiques. Un avantage du PROSOTRAN est qu'il ne demande pas certaines connaissances linguistiques comme la segmentation et l'annotation de la parole et permet d'encoder plusieurs dimensions prosodiques de manière indépendante. En outre, PROSOTRAN pouvant générer un étiquetage automatique, il serait aisé de comparer une quantité relative des énoncés (en l'occurrence, de la tâche de lecture) afin de pouvoir faire émerger des patrons communs chez les apprenants.

Comme le calcul conduisant au codage se fait à partir des valeurs associées aux paramètres physiques, le système peut être utilisé aussi bien pour annoter des données dans des langues dont le fonctionnement prosodique est connu (L1) que des données dont on ne connaît pas la grammaire sous-jacente, comme c'est le cas de la parole en L2. Enfin, un tel outil fournit une annotation prosodique même si aucune connaissance linguistique de la langue à annoter n'est disponible.

En revanche, le calcul de la durée moyenne de toutes les voyelles ne permet pas de distinguer les différents niveaux de frontières prosodiques en français L1. En outre, l'outil n'a pas été testé pour le codage de la parole spontanée en L2 ou les variations temporelles sont beaucoup plus instables que dans la parole lue.

Un dernier outil qui mérite notre attention est le système Prosogramme (Mertens, 2004) qui fournit une stylisation automatique du tracé de f_0 en rapport avec le seuil de perception. Cette stylisation a l'avantage de donner une représentation des événements mélodiques complètement indépendante du système langagier à encoder, et cela que le système intonatif sous-jacent soit méconnu ou non. Le Prosogramme est donc un système de transcription qui se veut neutre. A la différence du système INTSINT, le Prosogramme tient compte des impacts auditifs et perceptifs des paramètres acoustiques analysés aux noyaux syllabiques. En d'autres termes, le Prosogramme offre *une représentation de l'intonation perçue* (Mertens, 2004 : 111).

Les objectifs du Prosogramme sont les suivants (cf. Mertens, 2004 : 112) : (i) fournir une représentation objective et fiable de la prosodie facile à interpréter, et ce, de manière automatique utilisable dans de grands corpus, (ii) considérer l'évolution de la hauteur de f_0 en termes de déclinaison, d'attaque, de registre et de changement de registre, et (iii) analyser les facteurs temporels en tenant compte des pauses, hésitations, de déterminer le débit et les aspects rythmiques comme les accélérations ou les ralentissements.

Ce système de transcription est fondé sur une simulation de la perception de la hauteur mélodique tout en considérant le noyau syllabique comme unité de base. Il prend en considération le seuil de Glissando Différentiel ou DG (Alessandro & Mertens 1995). Ce seuil établit que les variations de la fondamentale doivent présenter une ampleur minimale qui varie en fonction de la fréquence de départ et de la durée du stimulus pour qu'elles soient perçues. Le système fournit ainsi une stylisation de la courbe mélodique en fonction des phénomènes fonctionnels ou audibles. Dans ce système, tout changement de pente est comparé au seuil de DG, et s'il est inférieur au seuil, le changement de pente est considéré comme inaudible.

Cette stylisation est très proche de la notation manuelle, au moins pour coder les événements de *glissando*. La validation de la stylisation obtenue par le Prosogramme peut se réaliser avec la resynthèse du signal de parole utilisée. À cette fin, le Prosogramme reprend toutes les caractéristiques

du signal original, sauf la f_0 , pour laquelle la stylisation calculée est remplacée. Cela permet à l'annotateur de vérifier à l'écoute du signal resynthétisé l'évaluation de la correspondance de la représentation auditive obtenue par le Prosogramme et une transcription manuelle de la part de l'annotateur. Dans la figure 10, nous illustrons une copie d'écran du mode interactif de la stylisation obtenue du Prosogramme de la phrase *Pourriez-vous m'indiquer à quelle heure est la correspondance à Valence ?* en français L2 (donnée préalablement). Cette image illustre la fenêtre interactive montrant la resynthèse du contour mélodique stylisé. Dans le Prosogramme, la ligne continue verte représente l'intensité, le trace fin bleu discontinu la f_0 et les traits épais noirs superposés sa stylisation. Les valeurs de f_0 sont données en demi-tons (relatifs à 1 Hz). Les lignes horizontales pointillées représentent les changements tonals à 2 demi-tons.

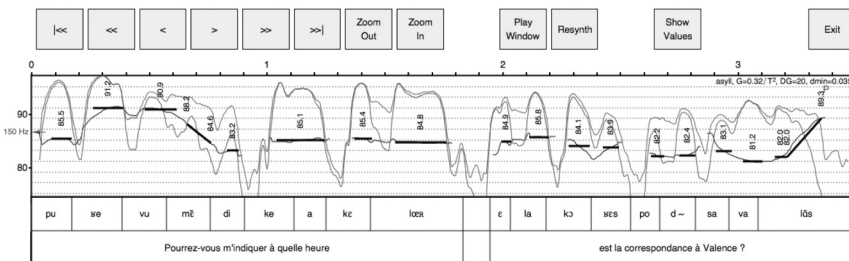


Fig. 10. Exemple de resynthèse avec le Prosogramme appliqué à un énoncé en français L2.

Le Prosogramme a en revanche certaines limites. D'une part, il ne fournit que des valeurs de la stylisation de la courbe mélodique et ne fournit aucune représentation symbolique de la nature de ces mouvements. Cependant, ces valeurs peuvent permettre au chercheur de comparer, de façon semi-automatique, où se trouvent les mouvements mélodiques importants dans l'énoncé, leur forme et leur ampleur. C'est en analysant ces patrons de manière récurrente que le chercheur peut ensuite tirer des conclusions sur la nature de ces mouvements. Les études de Santiago & Delais-Roussarie (2015) et Santiago & Mariano (2019) ont montré que cet outil s'avère efficace pour comparer de manière semi-automatique les productions en français par des hispanophones et des

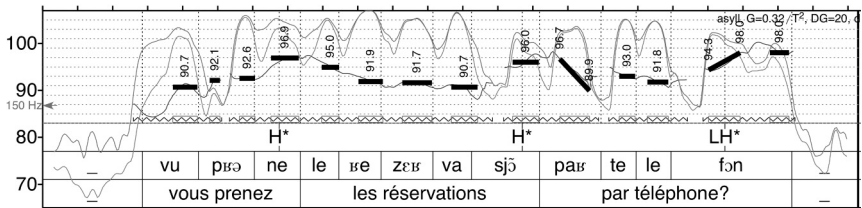


Fig. 11. Tracé de la stylisation de f_0 (ligne noire) et valeurs exprimées en demi-tons obtenues par le Prosogramme d'un énoncé interrogatif produit par une locutrice francophone.

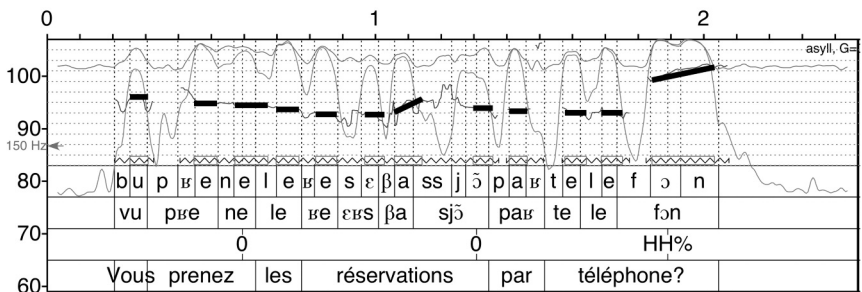


Fig. 12. Tracé de la stylisation de f_0 (ligne noire) et valeurs exprimées en demi-tons obtenues par le Prosogramme d'un énoncé interrogatif produit par une étudiante hispanophone de FLE.

anglophones. De surcroît, cet outil est parfaitement compatible avec des annotations manuelles qui peuvent relever d'autres systèmes comme ToBI.

C'est l'exemple de l'étude de Santiago & Delais-Roussaire (2015) portant sur les groupes accentuels en français L2. L'énoncé *Vous prenez les réservations par téléphone ?* contient trois GA potentiels : [Vous prenez]_{GA} [les réservations]_{GA} [par téléphone]_{GA} (cf. Di Cristo 2016; Jun & Fougeron, 2002). Le codage des GA apparaît dans la Figure 11, selon la stylisation obtenue par le Prosogramme et l'annotation manuelle par les auteurs pour le cas d'un francophone natif. Dans la Figure 12, le Prosogramme et l'annotation ToBI illustrent le même énoncé produit par une étudiante hispanophone. Notons que les étiquettes H* et LH*/HH% sont associés à des contours mélodiques hauts (non dynamiques) et montants (dynamiques)

respectivement. A partir de ces exemples, nous pouvons constater clairement que, dans le cas d'un apprenant hispanophone, les GA ne sont pas marqués prosodiquement : cet énoncé a comme caractéristique générale un contour mélodique final très montant sans avoir aucune variation tonale perçue en position interne (indiqué par 0). Enfin, le Prosogramme fournit également les valeurs temporelles des toutes les syllabes analysées afin de pousser les analyses temporelles.

4. Bilan et conclusion

Cette contribution avait pour objectif d'exposer les critères théoriques et méthodologiques pour étudier la prosodie en français L2. Nous avons insisté sur le fait que la constitution d'un corpus ne se réduit aucunement à un simple recueil de données numérisées. En effet, plusieurs étapes d'ordre méthodologique ont été menées afin de documenter et d'annoter le corpus. Ces étapes comprennent : le type de données collectées, la transcription orthographique du corpus et les annotations phonologiques et prosodiques. La réalisation de ces étapes a soulevé plusieurs questions d'ordre théorique et méthodologique.

Plus particulièrement, la transcription orthographique et les annotations phonologiques/phonétiques ont mérité un point de réflexion important. Les codages orthographique et phono-phonétique obéissent à des contraintes d'ordre méthodologique pour l'exploitation des données *a posteriori*. Sans trop s'attarder sur ce point, le chercheur doit être conscient que cette approche pragmatique permet de se focaliser sur la prosodie en L2. En revanche, les choix pris ne permettent pas faire des analyses plus fines entre les compétences phonologiques (consonnes et voyelles) et prosodiques.

Pour étudier la prosodie d'une L2, une annotation prosodique doit être effectuée lors de l'étude du corpus. Afin que l'annotation des patrons prosodiques des apprenants puisse être comparée à celles des productions des natifs, un système de codage prosodique neutre ne tenant pas compte du statut phonologique/phonétique des événements prosodiques doit être conçu. Il ressort de l'analyse portant sur plusieurs systèmes de transcription prosodique actuels, tels que l'API ou ToBI, qu'aucun de ces systèmes ne sont pas adaptés pour coder les phénomènes prosodiques en L2. Plusieurs inconvénients émergent lorsqu'on veut adopter ces systèmes pour le codage de la prosodie en L2. Des systèmes comme l'API ou ToBI

ont été essentiellement conçus pour annoter des événements contrastifs qui impliquent, nécessairement, la connaissance du système langagier à annoter. Or, l'interlangue des apprenants est un système méconnu par le chercheur. En outre, des systèmes comme ToBI ont été conçus pour la modélisation de la variation mélodique sans tenir compte des variations temporelles. Lors de l'analyse de nos données, nous avons souligné que certains paramètres acoustiques, comme la durée, peuvent être utilisés chez les apprenants comme des stratégies compensatoires pour marquer la structure prosodique dans la L2. Il a donc été évident que fonder une analyse exclusivement sur l'observation des mouvements mélodiques pourrait réduire l'analyse de la compétence prosodique en L2.

Au vu des contraintes observées dans les systèmes de transcription existants, nous avons discuté des outils d'annotation prosodique automatiques basés sur des critères perceptifs ou acoustiques : Prosotran et le Prosogramme. Ce dernier a l'avantage de fournir une stylisation de la courbe mélodique à partir du seuil de *glissando* en offrant une image acoustique de la perception des mouvements mélodiques du signal sans tenir compte des connaissances phonologiques de la langue à annoter. Nous avons expliqué comment ce système permet d'évaluer les productions des apprenants et des natifs sans pour autant faire référence à une théorie linguistique ou phonologique en particulier.

Le codage exemplifié ici s'est inspiré des symboles employés dans l'approche métrique et autosegmentale, et plus particulièrement dans les systèmes ToBI appliqués au français et à l'espagnol. En nous appuyant sur les étiquettes du système ToBI et en les adaptant, nous avons présenté les avantages d'un système semi-automatique pour représenter les phénomènes phonologiques et leurs manifestations phonétiques observés dans les données des apprenants, notre but étant de faire émerger les catégories phonologiques de l'interlangue. Lors de cette adaptation, nous avons clarifié ses limites. Nous avons ainsi mentionné que l'adaptation de tels systèmes ne reflète pas non plus les variations temporelles. Bien que les valeurs temporelles fournies par le Prosogramme puissent compléter cette analyse, les annotations prosodiques sont pour l'instant semi-automatiques. Enfin, des outils comme Prosotran doivent être pilotés et améliorés afin de comparer les représentations des deux logiciels. Enfin, il est clair que, pour rendre compte des règles prosodiques des apprenants, il faut que ces analyses soient réalisées dans différents styles de parole et avec des données construites et authentiques.

Bibliographie

- (d')Alessandro, Christophe & Mertens, Piet. 1995. Automatic pitch contour stylization using a model of tonal perception. *Computer Speech and Language* 9 (3). 257-288.
- Bartkova, Katarina & Delais-Roussarie, Élisabeth & Santiago, Fabian. 2012. PROSOTRAN: a tool to annotate prosodically non-standard data. In Ma, Qiuwu & Ding, Hongwei & Hirst, Daniel J. (eds.), *Proceedings of 6th International Conference on Speech Prosody*, 55-58. Shanghai : Tongji University Press.
- Beckman, Mary & Hirschberg, Julia. 1994. The ToBI Annotation Conventions. Disponible sur http://www.ling.ohiostate.edu/~tobi/ame_tobi/annotation_conventions.html.
- Bilger, Mireille. 2007. Réflexions sur un obscur objet de désir : le corpus. *Cahiers de l'Association for French Language Studies (e-journal)*. 13 (1). 2-17.
- Blanche-Benveniste, Claire & Jeanjean, Colette. 1987. *Le français parlé*. Paris : Didier Erudition.
- Boersma, Paul & Weenink, David. 2018. *Praat: doing phonetics by computer* [Computer program], Version 6.1.09, retrieved 01 January 2018 from <http://www.praat.org/>
- Colantoni, Laura & Steele, Jeffrey & Escudero, Paola. 2015. *Second Language Speech. Theory and Practice*. Cambridge: Cambridge University Press.
- De la Mota, Carme & Martín Butragueño, Pedro & Prieto Pilar. 2010. Mexican Spanish Intonation. In Prieto, Pilar & Roseano, Paolo (eds.), *Transcription of Intonation of the Spanish Language*, 319-350. München: Lincom Europa.
- Delais-Roussarie, Élisabeth. 2003. Constitution et annotation de corpus : méthodes et recommandations. In Delais-Roussarie, Élisabeth & J. Durand, Jacques (éds), *Corpus et variation en phonologie du français : méthodes et analyse*, 127-157. Toulouse : Presses Universitaires du Mirail.
- Delais-Roussarie, Élisabeth & Yoo, Hi-Yon. 2011. Learner corpora and prosody : from the COREIL corpus to principles on data collection and corpus design. *Poznań Studies in Contemporary Linguistics* 41 (1). 26-39.
- Delais-Roussarie, Élisabeth *et al.* 2015. Intonational phonology of French: Developing a ToBI system for French. In Frota, Sonia & Prieto, Pilar (eds.), *Intonation in Romance*, 63-100. Oxford : Oxford University Press.
- Detay, Sylvain & Durand, Jacques & Laks, Bernard & Lyché, Chantal (éds.). 2010. *Les variétés du français parlé dans l'espace francophone*. Paris: Ophrys.

- Di Cristo, Albert. 2016. *Les musiques du français parlé : Essais sur l'accentuation, la métrique, le rythme, le phrasé prosodique et l'intonation du français contemporain*. Berlin : De Gruyter.
- Durand, Jacques & Laks, Bernard & Lych, Chantal. 2009. Le projet PFC : une source de données primaires structurées. In Durand, Jacques & Laks, Bernard & Lyche, Chantal (éds), *Phonologie, variation et accents du français*, 19-61. Paris : Hermès.
- Durand, Jacques & Tarrier, Jean-Michel. 2006. PFC, corpus et systèmes de transcription. *Cahiers de Grammaire*. 30. 139-158.
- Gabriel, Christoph. & Kireva, Elena. 2014. Prosodic transfer in learner and contact varieties. *Studies in Second Language Acquisition*. 36. 257-281.
- Goldman, Jean-Philippe. 2011. EasyAlign: an automatic phonetic alignment tool under Praat. *INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association*. Florence, Italy, August 27-31, 2011.
- Gut, Ulrike. 2007. Learner corpora in second language prosody research and teaching. In Trouvain, Jürgen & Gut, Ulrike (eds), *Non-Native Prosody. Phonetic Description and Teaching Practice*, 145-167. Berlin: Mouton de Gruyter.
- Hirst, Daniel. 2007. A Praat plugin for Momel and INTSINT with improved algorithms for modelling and coding intonation. In Barry, William G. & Trouvain, Jürgen (eds.). *Proceedings of the XVIth International Conference in Phonetic Sciences*, Saarbrücken, Germany, 6-10 August 2007, 1233-1236. Saarbrücken : Universität des Saarlandes.
- Jilka, Matthias. 2000. *The contribution of intonation to the perception of foreign accent*. Thèse de doctorat, Universität Stuttgart.
- Jun, Sun-Ah (ed.). 2014. *Prosodic Typology II. The Phonology of Intonation and Phrasing*, Oxford: Oxford University Press.
- Jun, Sun-Ah & Fougeron, Cécile. 2002. Realizations of accentual phrase in French intonation. *Probus* 14. 147-172.
- Leather, Jonathan. 1999. Second-Language Speech Research: An Introduction. *Language Learning* 49. 1-56.
- MacWhinney, Brian. 2000. *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition, Mahwah, NJ: Lawrence Erlbaum Associates.
- Martin, Philippe. 2012. The Autosegmental-Metrical Prosodic Structure: not fit for French? In Ma, Qiuwu & Ding, Hongwei & Hirst, Daniel J. (eds.), *Proceedings of 6th International Conference on Speech Prosody*, 131-134. Shanghai : Tongji University Press.

- Mennen, Ineke. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, 32. 543-563.
- Mertens, Piet. 2004. The Prosogram : Semi-Automatic Transcription of Prosody based on a Tonal Perception Model. In Bel, Bernard & Marlien, Isabelle (ed.), *Proceedings of the International Conference on Speech Prosody 2004, Nara, Japan, March 23-26*.
- Post, Brechtje. & Delais-Roussarie, Élisabeth. 2006. Transcribing intonational variation at different levels of analysis. In Hoffmann, Rüdiger & H. Mixdorff, Hansjörg (ed.), *Proceedings of the International Conference on Speech Prosody, Dresden, May 2-5 2006*, Dresden : TUDpress.
- Post, Brechtje & Delais-Roussarie, Élisabeth & Simon, Anne-Catherine. 2006. Développer un système de transcription des phénomènes prosodiques. In Simon, Anne-Catherine & Caelen-Haumont, Geneviève & Pagliano, Claudine (éds), *Bulletin PFC*. Vol. 6. 51-68.
- Racine, Isabelle & Detey, Sylvain & Zay, Françoise & Kawaguchi, Yuji. 2012. Des atouts d'un corpus multitâches pour l'étude de la phonologie en L2: l'exemple du projet « Interphonologie du français contemporain » (IPFC). In Kamber, Alain & Skupien Dekens, Carine (éds.), *Recherches récentes en FLE*, 1-19. Bern : Peter Lang.
- Racine, Isabelle & Zay, Françoise & Detey, Sylvain & Kawaguchi, Yuji. 2011. De la transcription de corpus à l'analyse interphonologique: enjeux méthodologiques en FLE. In Col, Gilles & Osu, Sylvester N. (éds.), *Transcrire, écrire, formaliser*, 13-30. Rennes : PUR. (Travaux Linguistiques du CerLiCO, 1).
- Rosenberg, Andrew. 2010. AuToBI - a tool for automatic ToBi Annotation, *INTERSPEECH 2010*, Makuhari, Chiba, Japan, September 26-30, 2010.
- Santiago, Fabian. 2014. *Systèmes prosodiques et acquisition d'une L2 : production et perception des mouvements mélodiques en français et en espagnol*. PhD Thesis. Université de Paris (Sorbonne Paris Cité).
- Santiago, Fabian & Delais-Roussarie, Élisabeth. 2015. The acquisition of Question Intonation by Mexican Spanish Learners of French. In Delais-Roussarie, Élisabeth & Avanzi, Mathieu & Herment, Sophie (eds), *Prosody and Language in Contact: L2 Acquisition, Attrition and Languages in Multilingual Situations*, 243-270. Heidelberg: Springer.
- Santiago, Fabian & Mairano, Paolo. 2019. Prosodic effects on L2 French vowels: a corpus-based investigation. In Calhoun, Sasha & Escudero, Paola & Tabain, Marija & Warren, Paul (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS 2019), Melbourne, Australia, 5-9 August 2019*, 1084-1088.

- Sinclair, John. 1996. *Preliminary recommendations on Corpus Typology*. Rapport Technique. EAGLES (Expert Advisory Groups on Language Engineering Standards). Pisa : Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale.
- TEI Consortium, eds. *Guidelines for Electronic Text Encoding and Interchange*. [24.06.2020]. <http://www.tei-c.org/P5/>.
- Ueyama, Motoko & Jun, Sun-Ah. 1996. Focus realization of Japanese English and Korean English intonation. *UCLA Working Papers in Phonetics* 94. 110-125.
- Vaguer, Céline. 2003. Corpus, vous avez dit corpus ! De la notion de corpus à la création d'un "corpus informatisé. In Williams, Geoffrey Clive (éd.), *Corpus, Langues et Linguistique, Actes des 3es Journées de la linguistique de corpus*, 207-223.



Questa pubblicazione è stata realizzata utilizzando carta fabbricata nel pieno rispetto dell'ambiente senza l'utilizzo di sostanze nocive e con l'impiego di prodotti ecocompatibili nella fase di stampa e confezione.

Finito di stampare
nel mese di luglio 2021
sestanteinc - Bergamo



9 788866 423690