Facteurs caractérisant les hésitations dans les grands corpus : langue, genre, style de parole et compétence linguistique

Ioana Vasilescu & Martine Adda-Decker

LIMSI-CNRS, Bât. 508, BP 133, 91403 Orsay cedex, France
Mél : ioana, madda@limsi.fr

ABSTRACT

This paper deals with the factors characterizing the autonomous vocalic filled pauses in large spontaneous speech corpora, namely language, gender, speaking style and language proficiency. Two corpora are analyzed: a corpus of broadcast news in French and American English and a corpus of short talks in a conference in English spoken by native and non-native speakers. Several acoustic and prosodic parameters are evaluated and correlated with each factor, namely timbre, pitch, duration and density. Results presented here show that the timbre is correlated with language and language proficiency, whereas the duration is linked both to gender and speaking style, the latter conditioning also the hesitation density in speech.

1. INTRODUCTION

Nous nous intéressons ici aux voyelles d'hésitation autonomes dans les grands corpus. Les hésitations représentent un des multiples phénomènes dits de « disfluence » recensés dans toutes les langues. Parmi eux notons l'allongement vocalique, les répétitions, les reformulations etc. Les hésitations vocaliques autonomes représentent un phénomène largement rencontré dans les langues qui consiste en l'insertion « à tout moment » dans le flux de parole spontanée d'un segment vocalique plutôt allongé. Ce segment vocalique peut être accompagné ou non de segments adjacents (coda nasal en anglais, diphtongaison etc.). Son rôle est « d'annoncer l'initiation de ce qui est attendu comme […] un délai dans la parole » (notre trad.) [1]. Les hésitations vocaliques n'ont pas de support lexical, ce qui les différencie de phénomènes similaires tels les allongements vocaliques d'un segment appartenant à un item lexical précis (généralement un mot grammatical).

La réalisation vocalique n'est pourtant pas la seule rencontrée parmi les langues du monde, d'autres phénomènes d'hésitations autonomes sont dénombrés, telles des consonnes nasales allongées (« mm » en chinois mandarin, par exemple), ou des demonstratives délexicalisés (« ano », « eto » en japonais, par exemple) [2,3]. Nous prenons en compte ici uniquement les hésitations vocaliques autonomes en français (« euh ») et en anglais (« uh », « um » en anglais américain / « er » en anglais du Royaume-Uni).

Antérieurement nous avons comparé les hésitations vocaliques autonomes en 8 langues : anglais américain, arabe, allemand, chinois mandarin, français, italien, espagnol sud-américain, portugais européen. Nous nous sommes intéressées à la voyelle support de chaque hésitation. Nous appelons *voyelle support*, la voyelle la plus longue et la plus stable de chaque occurrence. Cette voyelle est parfois diphtonguée à la fin ou suivie d'une coda nasale, comme en anglais. La voyelle support constitue en général l'élément principal d'une hésitation. Parmi les paramètres considérés (*durée*, *hauteur* et *timbre*) il s'est avéré que le *timbre* est le paramètre dépendant de la langue qui caractérise le mieux les hésitations. La *hauteur* et la *durée* permettent de différencier la voyelle d'hésitation des voyelles intra-lexicales de timbre similaire au sein d'une même langue. L'analyse inter-langue des paramètres hauteur et durée n'a pas révélé des différences majeures parmi les 8 langues considérées. Nos analyses ont confirmé des observations antérieures, i.e. l'hésitation vocalique autonome est significativement plus longue que les segments intra-lexicaux de timbre similaire et possède un contour F0 plat et stable [4]. Nous avons formulé l'hypothèse que le paramètre *timbre* est dépendant de la langue tandis que les paramètres *hauteur* et *durée* tendent à être des critères universels.

Nous considérons ici quatre facteurs susceptibles d'influer la production d'hésitations autonomes dans les grands corpus. Il s'agit des facteurs *langue* ; *genre* ; *style de parole* et *compétence linguistique* (langue maternelle, langue seconde).

2. CORPUS ET METHODOLOGIE

Deux corpus ont été utilisés dans cette étude : un corpus de journaux télévisés en anglais américain et en français et un corpus d'enregistrements d'interventions orales dans une conférence en anglais. Dans le corpus de journaux télévisés en anglais américain (désormais JTA) et le français (désormais JTF) sont parlés par des natifs, hommes et femmes. Nous comptons 6 sources en anglais américain (CNN, VOA, ABC etc.) et environ 150 locuteurs (dont 2 fois plus d'hommes) et 4 sources en français (France Inter, France Info, France2, France3) et environ 130 locuteurs (100 hommes, 30 femmes). Au total, environ 200 heures pour JTF et 100 heures pour JTE ont été utilisées. Le corpus d'interventions dans une conférence est « Terrible English Database » et consiste en des enregistrements de 10 minutes environ par des locuteurs natifs et non natifs d'anglais dans la conférence Eurospeech de 1993 [5]. Nous avons sélectionné 8 locuteurs français (désormais TEDF) et 3 locuteurs anglais (désormais TEDA). Ce sous-corpus préliminaire

ne contient pas d'intervenants femmes. Nous avons utilisé environ 1h30 d'enregistrements.

Les hésitations ont été automatiquement extraites et manuellement vérifiées afin d'éliminer des erreurs d'extraction potentielles ou des hésitations accompagnées de phénomènes non-verbaux pouvant jouer sur le calcul des paramètres (i.e. rires, bruits de bouche, bip de téléphone). Différemment de nos études précédentes, aucun critère de durée n'a été employé afin d'avoir une estimation réaliste de la durée et de la densité du phénomène par langue et par style de parole [6,7]. Le nombre d'occurrences par langue et corpus est montré ci-dessous.

Tableau 1 : Nombre d'hésitations pour les Français (h/f) et Anglais (h/f) parlant la langue maternelle (L1) ou une langue seconde (L2).

Corp./Loc.	**Français**	**Anglais**
JT	L1 : 1640 (h)/270 (f)	L1 :4455 (h)/491 (f)
TED	L2 :762 (h)	L1 :439 (h)

Les paramètres suivants ont été considérés : fréquence fondamentale (F0), timbre (F1/F2), durée et densité (durée totale hésitations/durée totale corpus).

3. LES FACTEURS *LANGUE* ET *GENRE*

Dans des études préliminaires antérieures, nous nous sommes intéressées aux particularités dépendantes de la langue *vs* universelles caractérisant les hésitations vocaliques autonomes. A cet effet, nous avons exploité des données de type « journaux télévisés » en huit langues. La présente étude prend en compte uniquement des hésitations en anglais et en français. Les données sont importantes quantitativement et permettent de rendre compte du phénomène d'hésitation à travers un nombre de paramètres susceptibles d'influencer ses caractéristiques acoustiques et prosodiques.

Les données analysées ici et issues des deux corpus de JT en français *vs* anglais américain confirment ces observations, notamment en ce qui concerne *le timbre*. Ainsi, la voyelle d'hésitation en anglais américain est significativement plus ouverte (F1) et plus antérieure (F2) que sa correspondante française (t-tests séries appariées, $p<0,0001$). Cette différence est indépendante de la variable genre (Figure 1).

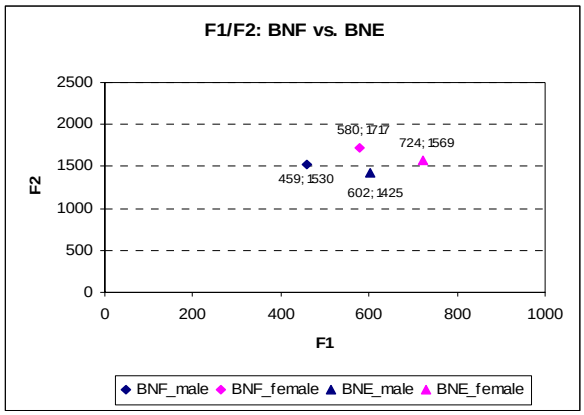

Figure 1 : Dispersion des valeurs moyennes des voyelles support dans un espace F1 *vs.* F2.

Le paramètre *hauteur* (F0) n'exhibe pas de différences notables entre les deux populations. En effet, la distribution des valeurs est équivalente, notamment pour les hésitations produites par les locuteurs hommes (Figure 2). En ce qui concerne les locutrices, il semblerait d'après la figure 2 que l'étendue pour les locutrices françaises est plus importante, avec plus de valeurs extrêmes. Notons toutefois que les données en français sont quantitativement moins importantes, ce résultat pouvant ainsi être un effet de corpus.

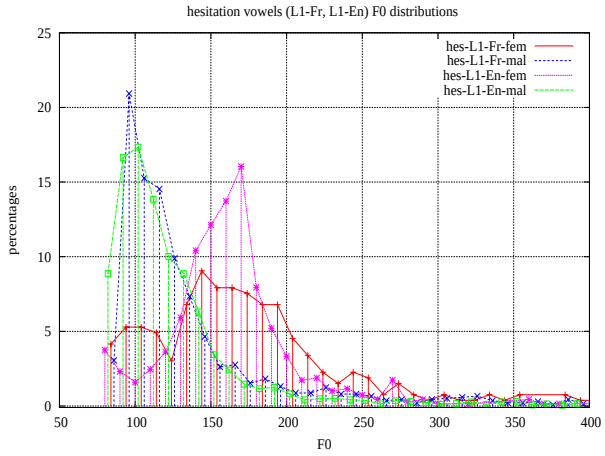

Figure 2 : Distribution des valeurs de F0 des voyelles support dans JT (anglais et français, hommes et femmes).

Le troisième paramètre analysé est la *durée*. Les données étudiées ici confirment la tendance observée dans d'autres langues, à savoir que la voyelle d'hésitation est significativement plus longue qu'une voyelle intra-lexicale. Elles mettent toutefois en évidence d'autres particularités (Tableau 2).

Tableau 2 : Durées moyennes des hésitations dans les corpus JTF et JTA (hommes et femmes)

Corpus/genre (ms)	**Hommes**	**Femmes**
JTF	343	262
JTA	267	266

Ainsi, en ce qui concerne la variable *langue*, il apparaît que les hésitations produites par les locuteurs hommes en français sont d'une durée significativement supérieures à celle des hésitations des locuteurs d'anglais américain (ANOVA, F=237,102, p<0,0001). Cette observation ne concerne pas les locutrices et, globalement, une différence significative entre les deux langues n'a pas été notée. Nous avançons l'hypothèse que la différence concernant les locuteurs serait due à un effet de corpus et non pas à un effet de langue.

Le dernier aspect considéré concerne la structure segmentale des hésitations. Alors qu'en français l'hésitation type est « euh », donc une voyelle centrale, l'anglais américain en possède deux. Comme observée plus haut, il s'agit d'une voyelle plus ouverte et plus antérieure que la correspondante française, suivie ou non d'un segment consonantique nasal. Le corpus JTE montre que les réalisations avec coda nasale sont minoritaires, à savoir 23% des hésitations seulement sont produites avec coda nasale. Plus encore, les réalisations avec coda nasale sont plus spécifiques aux femmes (45%) qu'aux hommes (19%).

4. LE FACTEUR *STYLE DE PAROLE*

Afin d'évaluer l'impact du facteur *style de parole* nous avons analysé et comparé les données de JT avec des données de TED. Deux conditions d'élocution sont ainsi mises en parallèle. Il s'agit d'une part du journalisme télévisé, impliquant une parole semi-préparée et des professionnels des interventions orales dans un temps limité, et d'autre part d'orateurs présentant leurs travaux à un public-juge. Ces derniers sont plus susceptibles de subir l'effet du stress, d'autant plus que pour 8 d'entre eux l'intervention se fait dans L2. Cette partie de l'étude prend en compte les productions des hommes dans le corpus JT, le corpus TED décrit ici comportant pour l'heure les productions à 11 locuteurs.

La *densité* est un paramètre qui s'avère dépendant du facteur *style de parole*. La durée totale des hésitations représente 0,7% du corpus JTA et 0,1% de JTF, tandis que dans TEDA il s'agit de 5,8% et dans TEDF de 5,7%. Cette différence semble conforter l'hypothèse que des facteurs tels que le stress, la parole non-planifiée (absence des prompteurs aidant les locuteurs des corpus JT) et le fait que les locuteurs soient non-professionnels, pourraient jouer sur le pourcentage de disfluences présentes dans le discours. Le statut de langue maternelle vs. seconde langue ne semble pas avoir influencé le paramètre *densité* : elle est comparable dans le corpus TEDF vs. TEDA.

En ce qui concerne la *hauteur* F0 nous n'avons pas noté des différences significatives entre les valeurs moyennes dans les deux corpus.

Tableau 3 : F0 moyenne des hésitations dans les corpus JT et TED (hommes).

Corpus/F0_Moy (Hz)	**F0_Moyen**	**Ecart-type**
JTF	142	86
JTA	129	71
TEDF	129	42
TEDA	141	67

Le corpus TED montre une variabilité inter-locuteurs importante. Une analyse ANOVA prenant en compte les données des locuteurs français et anglais respectivement met en évidence un effet de locuteur statistiquement significatif (TEDF : ANOVA, F=9,7111, p<0,0001 ; TEDA : ANOVA, F=23,151, p<0,0001). Par ailleurs, il semble que les locuteurs français s'exprimant en anglais (L2) présentent une étendue plus importante des valeurs de F0, notamment en ce qui concerne le registre aigu (Figure 3). L'hypothèse pourrait être formulée qu'à la variabilité liée aux contraintes temporelles de la prise de parole en public s'ajoute celle de l'expression dans une L2. Ces facteurs pourraient influencer le contrôle du flux de parole et notamment de l'usage d'un registre modal. Cependant, plus de données seraient nécessaires pour valider cette hypothèse.

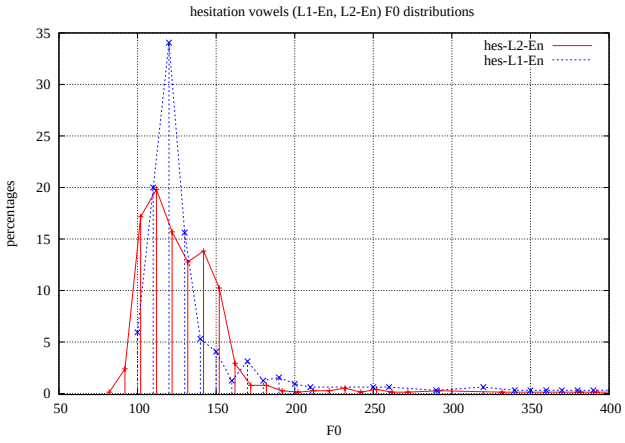

Figure 3 : Distribution des valeurs de F0 des voyelles support des hésitations dans le corpus TED.

La *durée* s'avère un paramètre sensible au facteur *style de parole*. Ainsi, les hésitations dans le corpus TED sont significativement plus longues que dans les corpus JT, et cela peu importe la langue (ANOVA, F=268,897, p<0,0001) (Tableau 3).

Tableau 4 : Durée moyenne des hésitations dans les corpus JT et TED (hommes).

Corpus/Durée moy. (ms)	**Durée moyenne**
JTF	342
JTE	266
TEDA	429
TEDF	415

Nous posons ici l'hypothèse que la différence de durée est liée au rôle des hésitations dans un corpus de type « journaux télévisés » *vs.* « conférence ». Dans une intervention de type conférence et, pour la plupart des locuteurs, menée dans une langue seconde, les hésitations marquent une vraie recherche du discours dans des conditions de stress. Dans les journaux télévisés les interventions sont souvent préparées et les hésitations pourraient avoir un caractère plus « canonique » que dans un contexte où la parole est moins contrôlée.

4. LE FACTEUR *COMPETENCE LINGUISTIQUE*

Nous appelons compétence linguistique le degré de maîtrise de l'anglais (L2) par les locuteurs français de TED. Ce facteur est évalué à travers le *timbre* des hésitations produites par les Français en anglais L2, par rapport au français L1, et à l'anglais L1.

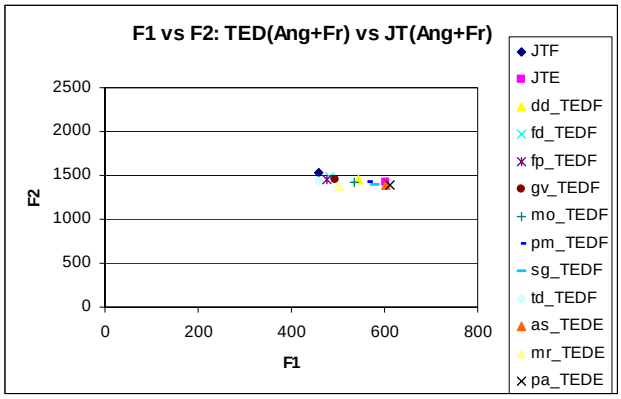

Figure 4 : Dispersion des voyelles support des locuteurs de JT et TED (JTF, JTA : moy./corpus ; dd, fd, fp, gv, mo, pm, sg, td : moy./loc. TEDF ; as, mr, pa : moy./loc. TEDA).

La figure 4 montre que les voyelles support en anglais (L2) se placent sur un continuum sur l'axe ouvert/fermé entre les valeurs des voyelles support en français et en anglais comme L1. Certains locuteurs français de TED produisent les hésitations de leur langue maternelle lorsqu'ils s'expriment en L2, tandis que d'autres produisent des voyelles support intermédiaires en termes d'ouverture, entre les valeurs de JTF et JTA.

5. DISCUSSION

Dans cette étude nous avons analysé plusieurs facteurs caractérisant les hésitations autonomes dans les grands corpus oraux. Les facteurs langue, genre, style de parole et compétence linguistique ont été évalués à travers les paramètres acoustiques et prosodiques traditionnellement mesurés pour décrire les hésitations vocaliques autonomes, i.e. timbre, durée, hauteur et densité. Le facteur *langue* a confirmé des observations antérieures, à savoir que le paramètre le plus distinctif est le timbre. La durée a été mise en lien à la fois avec le facteur *genre* et *style de parole*. Ce dernier doit être de plus mis en relation avec le paramètre densité. Enfin, la *compétence linguistique* se traduit par des productions intermédiaires en termes de timbre des voyelles support des hésitations en anglais (L2) par rapport au français et à l'anglais comme L1. Cette observation soulève des questions intéressantes liées au statut des disfluences lors de l'acquisition de L2.

BIBLIOGRAPHIE

[1] Clark H.H., Fox Tree J.E. 2002. Using uh and um in spontaneous speaking, *Cognition* 84, 73-111.

[2] Zhao, Y., Jurafsky, D., 2005, A preliminary study of Mandarin filled pauses, DISS05, Aix-en-Provence.

[3] Watanabe, M., Den, Y., Hirose, K. , Minematsu, N. (2005): "The effects of filled pauses on native and non-native listeners² speech processing", In *DiSS-2005*, 169-172.

[4] Shriberg, E., The 'errrr' is human: ecology and acoustics of speech disfluencies, *Journal of the International Phonetic Association*, 31/1, 2001.

[5] Lamel, L., Schiel, F., Fourcin, A., Mariani, J., Tillmann, H., "The translanguage english database ted", ICSLP 1994, Yokohama, Japon.

[6] Vasilescu, I. Candea, M., Adda-Decker, M., Hésitations autonomes dans 8 langues : une étude acoustique et perceptive, Workshop MIDL04, Paris France, 2004.

[7] Candea, M., Vasilescu, I., Adda-Decker, M., Inter- and intra-language acoustic analysis of autonomous fillers, DISS05, Aix-en-Provence, France.