

Etude des relations entre pauses et ponctuations pour la synthèse de la parole à partir de texte

Estelle Campione, Jean Véronis

Equipe DELIC, Université de Provence
29, Avenue Robert Schuman
13621 Aix-en-Provence Cedex 1
Estelle.Campione@up.univ-aix.fr

Résumé – Abstract

Nous présentons dans cette communication la première étude à grande échelle de la relation entre pauses et ponctuations, à l'aide de l'analyse de plusieurs milliers de pauses dans un corpus comportant près de 5 heures de parole lue en cinq langues, faisant intervenir 50 locuteurs des deux sexes. Nos résultats remettent en cause l'idée reçue de rapports bi-univoques entre pauses et ponctuations. Nous mettons en évidence une proportion importante de pauses hors ponctuation, qui délimitent des constituants, mais aussi un pourcentage élevé de ponctuations faibles réalisées sans pauses. Nous notons également une très grande variabilité inter-locuteur, ainsi que des différences importantes entre langues. Enfin, nous montrons que la durée des pauses est liée au sexe des locuteurs.

We present in this paper the first large-scale study of pause-punctuation relationships, based on the analysis of several thousand pauses in a corpus consisting of nearly 5 hours of read speech involving 50 male and female readers in five languages. Our results call into question the generally accepted idea of a one-to-one relationship between pauses and punctuation. We observe a large proportion of pauses outside punctuations (which mark phrases), but also a high percentage of weak punctuations with no pause. We also note a very high interspeaker variability, as well as important differences among languages.

Keywords – Mots Clés

Synthèse de la parole à partir de textes, pauses, ponctuation.

Text-to-speech synthesis, pauses, punctuation.

1 Introduction

La génération de pauses est absolument nécessaire à la synthèse de la parole, et le réalisme de leur durée et de leur position est indispensable à la qualité de la synthèse résultante. Pendant longtemps, les systèmes de synthèse à partir de texte ont pu, étant donné les qualités segmentale et intonative très moyennes que l'état de la technique permettait d'atteindre, se satisfaire d'heuristiques grossières pour la génération des pauses. On a vu ainsi de nombreux systèmes utiliser deux niveaux de pauses, brèves et longues, les premières étant associées aux ponctuations « faibles » (internes aux phrases), les secondes aux ponctuations « fortes » (entre phrases). Maintenant que la qualité segmentale et intonative s'est très largement améliorée, d'autres paramètres, comme les pauses (mais aussi le rythme, le débit, l'intensité), méritent une étude approfondie et des stratégies moins rustiques, permettant d'accroître le naturel et l'acceptabilité des systèmes. Bien sûr, d'autres informations que la simple ponctuation, et en particulier l'analyse syntaxique, sont utiles pour la génération des pauses, mais, dans l'état actuel de la technique, l'analyse syntaxique est coûteuse et peu fiable. Les ponctuations (d'ailleurs partiellement corrélées aux structures syntaxiques) restent donc des indices précieux pour la synthèse de la parole à partir de texte, à condition d'en connaître mieux le fonctionnement lors de la lecture.

Or, si les études sur les pauses sont nombreuses (Grosjean & Deschamps, 1972, 1973, 1975 ; Duez, 1991 ; Candéa, 2000 ; etc. — pour une revue étendue, voir Zellner, 1994), on dispose de très peu de données empiriques sur les pauses dans la parole lue et leur relation avec les ponctuations. On ne dispose guère par exemple pour le français que de l'étude faite par Vannier (1999), dans laquelle on peut constater qu'il ne renvoie lui-même à aucune étude antérieure. La question se complique encore dans un cadre de la synthèse multilingue, le peu de données dont on dispose (comme d'ailleurs l'ensemble des études sur les pauses en général) étant difficile à comparer entre les langues (différences de méthodologie, de types de parole, de techniques d'analyses, etc.). Ainsi, Vannier (1999) trouve (sur le français) que 64% des pauses correspondent à une ponctuation, alors que O'Connell & Kowal (1983) affirment (à propos de l'allemand) que ce sont 91% des pauses qui correspondent à une ponctuation. Ces différences proviennent-elles de la langue, ou bien des locuteurs particuliers¹, du type de texte et de lecture², ou bien encore de la méthodologie d'analyse³ ?

Nous proposons dans cette communication une étude à grande échelle, portant sur l'analyse de plusieurs milliers de pauses dans près de 5 heures de parole lue en cinq langues et faisant intervenir 50 locuteurs des deux sexes. Les données que nous publions permettent d'avoir pour la première fois une image précise de la relation entre pauses et ponctuations, qui bouscule d'ailleurs quelque peu les idées reçues, ainsi qu'un point de vue contrastif sur les différences de fonctionnement entre langues.

¹ L'étude de O'Connell & Kowal fait intervenir quatre locuteurs, celle de Vannier un seul.

² Les corpus sont très différents (sermons radiophoniques pour O'Connell & Kowal, lecture du Monde pour Vannier).

³ Voir par exemple la question des seuils ci-après.

2 Corpus

Le corpus que nous avons utilisé est le « Few Talker Set » de la base de données EUROM1 produite dans le cadre du projet Esprit SAM (Chan et al., 1995), totalisant 4h et demie de parole. Ce corpus est constitué de passages de cinq phrases chacun, lus dans différentes langues, parmi lesquelles nous avons sélectionné pour cette étude l'allemand, l'anglais, l'espagnol, le français et l'italien. Les passages dans les différentes langues sont la traduction de la version originale anglaise. Cette traduction a été réalisée de façon libre, puisqu'elle devait refléter non seulement une bonne adaptation de la version anglaise dans la langue de destination, mais aussi la culture propre à celle-ci (Figure 1).

Anglais

I have a problem with my water softener. The water-level is too high and the overflow keeps dripping. Could you arrange to send an engineer on Tuesday morning please? It's the only day I can manage this week. I'd be grateful if you could confirm the arrangement in writing.

Allemand

Ich habe da ein Problem mit meinem Boiler. Das Überdruckventil tropft ständig, weil der Wasserstand wohl zu hoch ist. Könnten sie mir bitte Dienstag früh einen Klempner schicken? Das ist leider der einzige Tag in dieser Woche, an dem ich Zeit habe. Ich wäre ihnen dankbar, wenn sie mir den Termin schriftlich bestätigen würden.

Espagnol

Tengo un problema con mi descalcificador. El nivel de agua es demasiado alto y el rebosadero gotea constantemente. ¿Por favor, podría enviarme un técnico el martes por la tarde? Es el único día que me va bien esta semana. ¿Me podrían llamar por teléfono antes de venir?

Français

J'ai des problèmes avec mon adoucisseur d'eau : le niveau d'eau est toujours trop haut, ce qui fait que le trop-plein coule sans arrêt. Vous pourriez pas m'envoyer un réparateur mardi matin ? C'est le seul jour de la semaine que j'ai de libre. Si vous êtes d'accord, soyez gentil de me confirmer le rendez-vous par écrit.

Italien

Il mio depuratore dell'acqua non funziona. Il livello dell'acqua è troppo alto e l'acqua comincia a traboccare. Può mandarmi un idraulico martedì mattina per favore? È l'unico giorno della settimana possibile per me. Può mandarmi una risposta scritta? Grazie.

Figure 1. Exemple de passage du corpus

Chaque passage a été lu par plusieurs locuteurs, 50 au total (25 hommes et 25 femmes). Les enregistrements sont de haute qualité acoustique (vitesse d'échantillonnage de 20 KHz, échantillonnage à 16 bits, enregistrement en chambre anéchoïque). Les lecteurs avaient pour instruction de lire les passages de la façon la plus naturelle possible, et le matériel enregistré a été contrôlé durant les acquisitions afin que les acquisitions de mauvaise qualité (bruit, ou phrases mal lues) soient directement annulées et répétées.

Nous avons marqué manuellement la totalité des pauses du corpus. Contrairement à la plupart des études sur les pauses (Grosjean & Deschamps, 1972, 1973, 1975 ; Duez, 1991 ; Candéa, 2000 ; etc.), aucun seuil, ni inférieur, ni supérieur, n'a été retenu, et toutes les pauses ont été comptabilisées et mesurées, quelle que soit leur durée⁴. En effet, nous avons montré ailleurs que l'utilisation de seuils peut avoir des effets dramatiques sur les résultats, conduisant à des conclusions parfois strictement inverses à la réalité (Campioni & Véronis, 2002). On voit par exemple sur la Figure 2 que le nombre de pauses très brèves (< 200 ms) est important (jusqu'à 17,9% en italien), et que les négliger ne serait pas sans conséquences.

⁴ Chaque pause a été minutieusement vérifiée auditivement pour éviter des confusions avec les occlusives.

Langue	(%)
<i>en</i>	3,9
<i>fr</i>	11,3
<i>ge</i>	4,8
<i>it</i>	17,9
<i>sp</i>	2,9
Total	8,2

Ponct.	Proportion
.	3128
?	440
!	78
...	2
Total	3648

Ponct.	Proportion
,	1820
:	105
;	23
(6
)	6
Total	1960

Figure 2. Pauses < 200 ms Figure 3. Ponctuations fortes Figure 4. Ponctuations faibles

La Figure 3 et la Figure 4 donnent la répartition des types de ponctuations dans le corpus.

3 Position des pauses

3.1 Pauses hors ponctuation

On note une proportion assez importante de pauses hors ponctuations, de l'ordre de 15,7% en moyenne sur l'ensemble des cinq langues, avec une variabilité importante entre celles-ci, puisque la proportion varie entre 7,5% pour l'anglais et 33,7% pour l'italien (Figure 5). S'il est à peu près du même ordre que celui publié par O'Connell & Kowal (1983) pour l'allemand (9%), ce chiffre est très inférieur à celui donné par Vannier (1999) pour le français (36%).

Langue	Absence de ponct.
<i>en</i>	6,4
<i>fr</i>	11,9
<i>ge</i>	11,0
<i>it</i>	33,7
<i>sp</i>	14,0
Moyenne	15,4

Figure 5. Pauses hors ponctuation (%)

La Figure 6 montre un exemple de passage en italien, où les pauses hors ponctuations sont particulièrement nombreuses. Celles-ci marquent, comme l'on peut s'y attendre, des ruptures syntaxiques importantes, mais on voit qu'elles sont à peu près toutes facultatives.

Texte	Locuteur			
	<i>ba</i>	<i>bf</i>	<i>bl</i>	<i>au</i>
Mia sorella	(+)	(+)		(+)
e terrorizzata dal buio	(+)	(+)	(+)	(+)
e rifiuta	(+)			
assolutamente	(+)			
di uscire sola	(+)			
la notte .	+	+	+	+
Vuole sempre qualcuno con se .	+		+	+
Mio padre	(+)	(+)		
l'ha consigliata	(+)			
di portare il cane	(+)	(+)	(+)	(+)
che almeno		(+)		
la proteggerebbe abbaiando	(+)			
se qualcuno la molestasse .	+	+	+	+

Figure 6. Pauses hors ponctuations (+)

Les différences entre langues ne peuvent être expliquées par des différences de densités de ponctuations dans les textes. En effet, la Figure 7 montre que le pourcentage de ponctuations par mots est à peu près constant entre les langues et de l'ordre de 13,5 %. Ceci ne peut donc rendre compte de la forte tendance de l'italien à placer des pauses hors ponctuation, et de la tendance moindre de l'anglais.

Langue	Total	Répartition	
		<i>Faibles</i>	<i>Fortes</i>
<i>en</i>	12,5	28,4	71,6
<i>fr</i>	13,8	50,6	49,4
<i>ge</i>	14,5	36,7	63,3
<i>it</i>	13,0	20,2	79,8
<i>sp</i>	13,6	36,6	63,4
Moyenne	13,5	34,5	65,5

Figure 7. Nombre de ponctuations par mot (%)

On peut également constater sur la Figure 6 que la variabilité inter-locuteur est grande, allant d'un style très scandé (locuteur *ba*) à un style très lié (locuteur *bl*). Cette variabilité est résumée par la Figure 8. On voit que l'italien se distingue des autres langues par une très grande variabilité inter-locuteurs, certains locuteurs allant jusqu'à placer 2/3 des pauses hors ponctuation. Pour le français, le chiffre de 36% donné par Vannier (1999) est très largement hors de l'intervalle de variation que nous observons (le maximum parmi nos locuteurs est de 16% ; le chiffre de Vannier se situe à 6,28 écarts-types de notre moyenne). Il se pourrait donc que le locuteur (unique) choisi par Vannier soit totalement atypique ou bien que le style de lecture soit totalement différent.

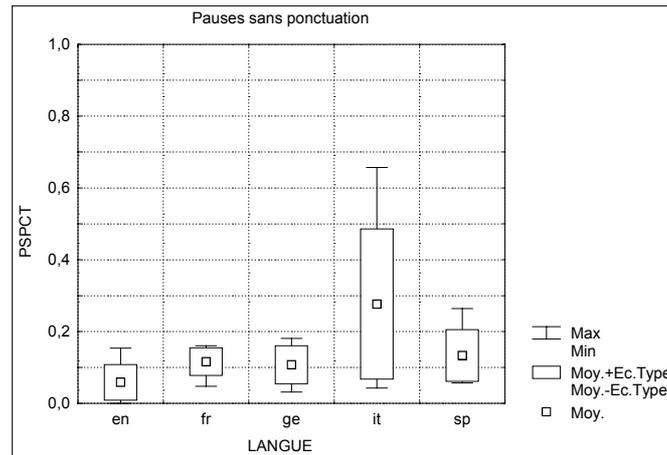


Figure 8. Pauses hors ponctuation : variabilité inter-locuteur

3.2 Ponctuations sans pauses

Inversement, de nombreuses ponctuations n'ont pas été réalisées avec une pause. En ce qui concerne les ponctuations fortes, ce phénomène est marginal. Absent en français et en espagnol, il concerne seulement quelques cas dans les autres langues (Figure 9).

Langue	Faible	Forte	Total
<i>en</i>	56,7	1,9	20,1
<i>fr</i>	33,1	0,0	18,7
<i>ge</i>	55,1	2,4	24,6
<i>it</i>	38,9	2,7	11,5
<i>sp</i>	46,4	0,0	19,6
Moyenne	46,0	1,6	19,0

Figure 9. Ponctuations sans pause (%)⁵

On observe des situations où les locuteurs ont enchaîné des phrases longues sans pause comme dans l'exemple ci-dessous, où un segment de plus de 9 secondes est prononcé sans aucune pause (le signe ø marque l'absence de pause) :

Schicken sie so schnell wie möglich einen krankenwagen in die großgarage in der hopfengasse. ø
Ein alter mann ist auf dem eis ausgerutscht, ø hat sich das bein gebrochen und kann nicht bewegt werden .

Ce type de prononciation n'est pas nécessairement un modèle à reproduire en synthèse, et est généralement contredit par les réalisations d'autres locuteurs sur le même passage.

En ce qui concerne les ponctuations faibles, la proportion de réalisations sans pause est presque de moitié (46,0%) ; elle atteint 56,7% en anglais. Ces cas concernent à peu près exclusivement les virgules (949 sur 962). Les démarcations prosodiques sont alors marquées par les mouvements mélodiques seuls :

⁵ Nous n'avons pas inclus les ponctuations finales des différents passages dans ce décompte.

She has a series of meetings, ø 9am to 5pm, ø in Paris, ø Bruges, ø Frankfurt, ø Rome, ø and Hamburg on consecutive days.

Quiere ponérselas todas de una vez , ø el mismo día.

Avec ce que je gagne, ø je peux tout juste me payer une paire de chaussures neuves.

Des études plus approfondies seront nécessaires pour déterminer dans quelles circonstances syntaxiques il est préférable d'omettre les pauses correspondant aux ponctuations faibles, mais il est certain que leur réalisation par des pauses systématiques aboutit à une sensation de parole hachée, trop scandée par rapport à la parole naturelle.

A nouveau, on observe une forte variabilité inter-locuteurs (Figure 10). Pour Vannier, 4,6 % des ponctuations (faibles et fortes confondues) ne donnent pas lieu à des pauses en français. Ce chiffre est beaucoup moins élevé que notre moyenne (19,0 %) et en dessous du minimum pour nos locuteurs français (6,4%). L'écart entre le chiffre de Vannier est toutefois seulement de 1,44 écart-types, ce qui est nettement moins important que dans le cas précédent.

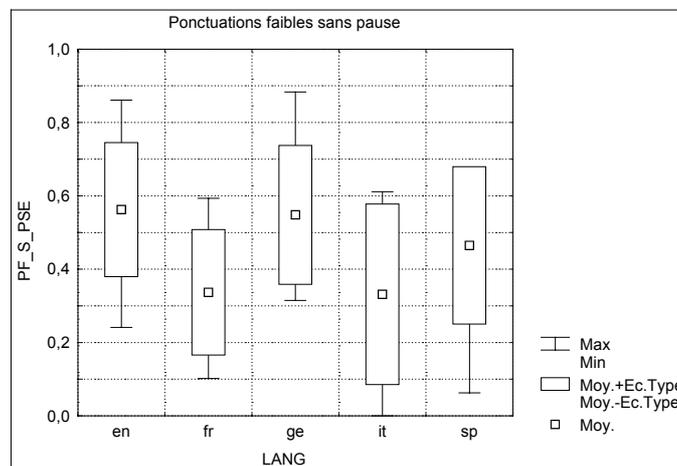


Figure 10. Ponctuations faibles sans pause : variabilité inter-locuteur

3.3 Durées

La distribution des durées de pause est fortement asymétrique, d'allure log-normale (Figure 11), comme il apparaît dans toutes les études sur les pauses. L'utilisation de la moyenne et de l'écart-type sur les données non transformées peut donner une représentation biaisée des données, et surtout rendre erronée l'application de tests comme l'analyse de variance, qui reposent sur des hypothèses de normalité et d'homogénéité des variances. Nous travaillons donc sur les logarithmes des durées, ou, ce qui revient au même, nous en exploitons les moyennes géométriques et non arithmétiques.

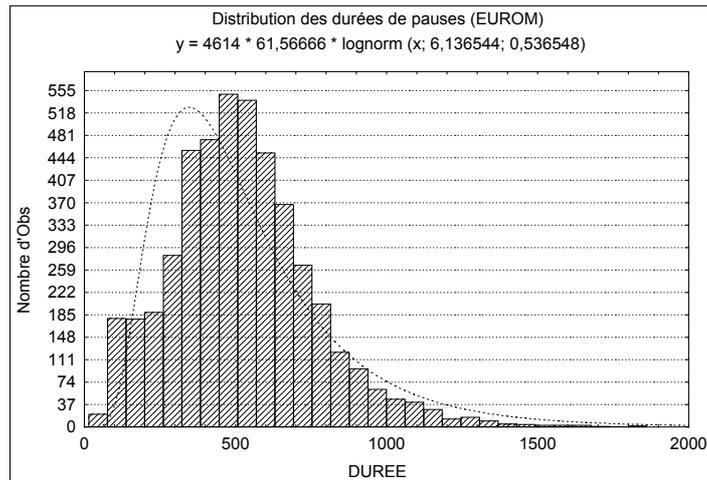


Figure 11. Distribution des durées de pause

L'analyse de variance sur les trois facteurs *langue*, *sexe* et *type de pause* révèle que la durée (logarithmique) des pauses dépend significativement de la langue⁶, mais aussi qu'elle dépend fortement du type de ponctuation où elles interviennent [$F(2, 4420)=1154,56; p<10^{-4}$]. En l'absence de ponctuation, la moyenne (géométrique) des pauses se situe à 289 ms. Elle passe à 336 ms pour les ponctuations faibles et à 598 ms pour les ponctuations fortes (Figure 13).

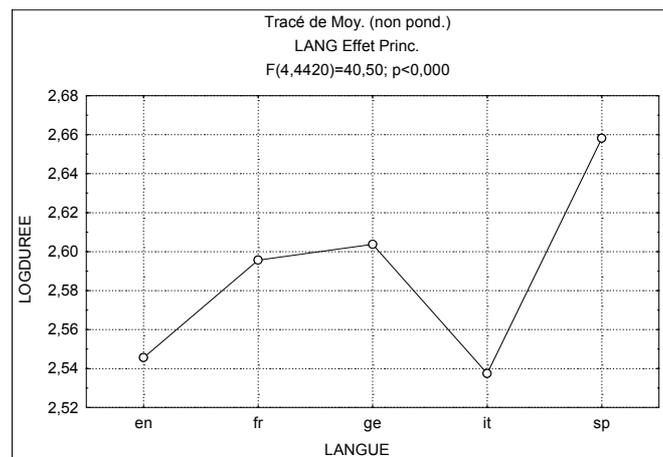


Figure 12. Durée moyenne des pauses en fonction de la langue

Ponctuation	log ₁₀ ms	ms (moy. géo)
<i>aucune</i>	2,46	289
<i>faible</i>	2,53	336
<i>forte</i>	2,78	598

Figure 13. Durée moyenne des pauses en fonction de la ponctuation associée

⁶ Cet effet peut être mis en relation avec des différences de débit dans chaque langue (Campione, 2001).

L'étude des interactions montre que les femmes et les hommes s'opposent de façon significative [$F(2, 803) = 247,06; p < 10^{-4}$]. Si les durées de pauses sont quasiment identiques pour les hommes et les femmes en présence de ponctuations fortes, les comparaisons post-hoc (test HSD de Tukey) montrent des différences très significatives ($p < 10^{-4}$) pour les durées de pauses en l'absence de ponctuation et en présence de ponctuation faible (Figure 14). Des études ultérieures seront nécessaires pour comprendre cet effet. Une hypothèse (à manier avec précaution) pourrait tenir à des raisons physiologiques (volume respiratoire) pouvant amener à une organisation différente des prises de souffle. En tous cas, cette différence pourrait être prise en compte dans la modélisation de locuteurs de sexes différents dans la synthèse de la parole.

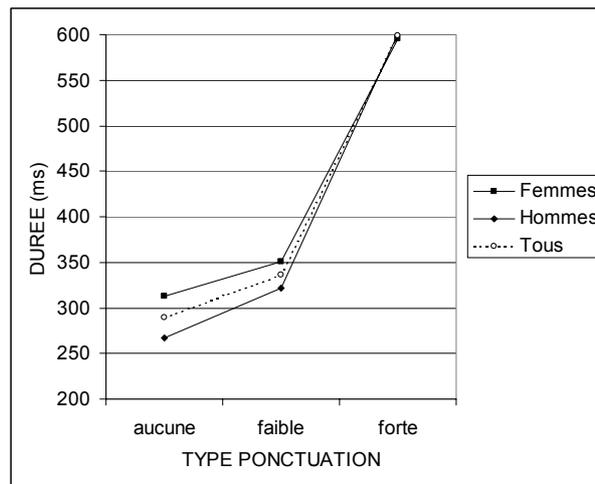


Figure 14. Durée des pauses en fonction du type de ponctuation

Les durées de pauses annoncées par Vannier (1999) sont plus élevées que celles que nous trouvons, puisque Vannier situe les durées moyennes (arithmétiques) à environ 450 ms pour les ponctuations faibles et 1600 ms pour les ponctuations fortes, alors que celles que nous observons sont respectivement de 391 ms et 708 ms⁷. Si le chiffre de Vannier se situe pour les ponctuations faibles dans l'intervalle de variation de nos locuteurs français (258 – 569 ms), il se situe largement au-delà pour les ponctuations fortes (567 – 1028 ms). Les textes de Vannier sont plus longs (500 mots), et il suffirait de quelques pauses très longues à des ruptures de paragraphes pour expliquer les durées moyennes plus élevées pour les ponctuations fortes. Nous n'avons pas le détail de la distribution des pauses qu'il observe.

4 Conclusion

Nous présentons dans cette communication la première étude à grande échelle de la relation entre pauses et ponctuations, à l'aide de l'analyse de plusieurs milliers de pauses dans un corpus multilingue comportant près de 5 heures de parole lue en cinq langues faisant intervenir 50 locuteurs des deux sexes. Nos résultats remettent en cause l'idée reçue de rapports bi-univoques entre pauses et ponctuations. Nous mettons en évidence une proportion

⁷ Nous avons ici calculé des moyennes arithmétiques pour rendre les chiffres comparables avec ceux de Vannier.

importante de pauses hors ponctuation, qui délimitent des constituants, mais aussi un pourcentage élevé de ponctuations faibles réalisées sans pauses. Nous notons également une très grande variabilité inter-locuteurs, ainsi que des différences importantes entre langues. Enfin, nous montrons que la durée des pauses est liée au sexe des locuteurs. Ces constatations pourront être utilisées pour améliorer la variabilité et le naturel des systèmes de synthèse de la parole à partir de texte. Il va de soi que cette étude devra être complétée par d'autres, et en particulier par un examen systématique des positions syntaxiques dans les différentes langues.

Références

Campione E., Véronis J. (2002), A large-scale multilingual study of silent pause duration, Actes de *Speech Prosody 2002*, Aix-en-Provence [à paraître].

Campione E. (2001), *Étiquetage automatique de la prosodie dans les corpus oraux : algorithmes et méthodologie*, Thèse de l'Université de Provence, Aix-en-Provence.

Candéa M. (2000). *Contribution à l'étude des pauses silencieuses et des phénomènes dits « d'hésitation » en français oral spontané*, Thèse de l'Université Paris III, Paris.

Chan D., Fourcin A., Gibbon D., Granström B., Huckvale M., Kokkinakis G., Kvale K., Lamel L., Lindberg B., Moreno A., Mouropoulos J., Senia F., Trancoso I., Veld C., Zeiliger J. (1995), Eurom - A Spoken Language Resource for the EU, *Proceedings of the 4th European Conference on Speech Communication and Speech Technology (Eurospeech'95)*, 867-870.

Duez D. (1991), *La pause dans la parole de l'homme politique*, Paris, Editions du Centre National de la Recherche Scientifique.

Grosjean F., Deschamps A. (1972), Analyse des variables temporelles du français spontané, *Phonetica*, vol. 26, pp. 130-156.

Grosjean F., Deschamps A. (1973), Analyse des variables temporelles du français spontané, *Phonetica*, vol. 28, pp. 191-226.

Grosjean F., Deschamps A. (1975), Analyse contrastive des variables temporelles de l'anglais et du français, *Phonetica*, vol. 31, pp. 144-184.

O'Connell D. C., Kowal, S. H. (1986), Use of punctuation for pausing: oral readings by german radio homilists, *Psychological Research*, vol. 48, pp. 93-98.

Vannier G. (1999), *Étude des contributions des structures textuelles et syntaxiques pour la prosodie : application à un système de synthèse vocale à partir de texte*, Thèse de l'Université de Caen/Basse-Normandie, Caen.

Zellner B. (1994), Pauses and the temporal structure of speech, In E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition* (pp. 41-62), Chichester: John Wiley.