

Répétitions et pauses (silencieuses et remplies) en français spontané

Sandrine Henry, Estelle Campione, Jean Véronis

Équipe DELIC

Université de Provence, 29 av. R. Schuman – 13621 Aix-en-Provence Cedex 1, France

Tél.: ++33 (0)4 42 95 31 37 - Fax: ++33 (0)4 42 95 34 95

Mél: Sandrine.Henry@up.univ-aix.fr - <http://www.up.univ-mrs.fr/delic/>

ABSTRACT

In this study, we show on a corpus of ca. one hour of spontaneous speech, that repetitions are strongly associated with pauses, and more specifically with filled pauses. In the latter case, syllabic lengthening is more frequent than the filler *eah*, which occurs mainly as a backup after a lengthening occurred. Our results indicate a gradation among the various marks, syllabic lengthening, fillers and silent pauses offering in an increasing order of disruptivity. Our results also show that in a large majority of cases, pauses (silent or filled) occur at the interruption point, which is coherent with previous experiments showing that listeners can detect repetitions very soon after the interruption point, and that this detection relies on acoustic cues rather than syntactic or semantic ones.

1. INTRODUCTION

Les disfluences ont fait l'objet de relativement peu de travaux, sans doute parce que, comme le font remarquer [5], [6], etc., l'oral spontané a été quelque peu négligé dans le passé dans les études sur la parole. Parmi ces disfluences, les *répétitions* ont reçu assez peu d'attention, particulièrement en français. Or, leur étude est intéressante d'un point de vue linguistique et phonétique, car elles peuvent apporter des indices précieux pour la compréhension des structures et modes de fonctionnement du langage (tout ne se répète pas, et surtout, ne se répète pas n'importe comment). De plus, des données précises sur les répétitions pourraient être très utiles pour des applications comme la reconnaissance de la parole, dans laquelle elles créent de très gros problèmes, d'autant que, pour des raisons pratiques les « modèles de langage » probabilistes utilisés sont le plus souvent entraînés sur de la parole lue non disfluente (journal *Le Monde*, etc.). Enfin, et c'est un point de vue qui intéresse tout particulièrement notre équipe, l'étude de la parole spontanée passe par le développement de grands corpus transcrits (cf. [7]). L'exploitation de ces corpus par des linguistes serait grandement facilitée si l'on pouvait les analyser du point de vue morpho-syntaxique, et marquer (semi)-automatiquement un certain nombre de phénomènes tels que les répétitions.

Pour cette dernière tâche, la difficulté provient du fait que toutes les séquences qui sont des répétitions du point de vue formel, et qui peuvent être repérées (plus ou moins aisément) par des algorithmes, ne sont pas toutes disfluentes. D'autres répétitions sont des répétitions de *langue*, prévues par le système grammatical, comme le montrent ces exemples tirés du *Corpus de référence de français parlé* (CRFP) [7] :

- emphase : *une hospitalité très très très grande, elle est belle belle belle*

- clitiques homographes : *nous nous sommes interrompus*

- rencontres de syntagmes : *banquier c'est un métier de relationnel relationnel au sens humain du terme*

Les répétitions disfluentes (e.g., *la la répartition du travail pose problème*) sont inhérentes à la parole spontanée et généralement produites en contiguïté, sans rupture intonative. Elles ne possèdent pas de valeur syntaxique ou sémantique.

Dans de précédents travaux [9], nous avons corroboré de façon quantitative, sur un corpus d'un million de mots de français spontané transcrit, l'observation de Blanche-Benveniste [1] selon laquelle les répétitions disfluentes ont lieu préférentiellement sur les débuts de syntagmes et que les répétitions font intervenir très majoritairement des mots grammaticaux (mots qui participent de la structuration de la langue) monosyllabiques. De plus, leur disposition reste stable : le cadre syntaxique est posé dans un premier temps et le remplissage lexical effectué dans un second. On ne rencontre jamais de cas où se serait l'inverse.

Nous nous intéressons ici à l'association des répétitions avec d'autres marques de disfluences, les pauses silencieuses et les pauses remplies (nous englobons sous ce dernier terme les *eah* et les allongements syllabiques, qui ont des caractéristiques comparables : [3] [4] [8]). Les deux types de pauses sont détectables par des moyens automatiques avec des taux de succès honorables (cf. respectivement [2], [8]), et peuvent donc fournir des indices importants pour la détection des répétitions disfluentes. L'étude des marques associées aux

répétitions est intéressante également du point de vue de la perception de la parole, puisque le locuteur doit lui aussi, tout comme un système de reconnaissance automatique, traiter de façon adéquate les répétitions et donc traiter les signaux lui permettant de les détecter efficacement [10].

Nous basons notre étude sur un corpus spontané tiré du CRFP, composé de 8 500 mots et 54 minutes de parole, faisant intervenir 10 locuteurs différents (5 hommes et 5 femmes), décrit plus en détail dans [2]. Les passages choisis correspondent à des parties monologiques où le locuteur parle de son métier, raconte des souvenirs, etc., en réponse aux questions d'un enquêteur (qui n'intervient pas dans les parties sélectionnées).

2. RÉSULTATS GÉNÉRAUX

Nous recensons 141 répétitions disfluentes dans notre corpus, soit environ 17 pour 1 000 mots, ou une occurrence toutes les 23 secondes. Ces valeurs corroborent celles que nous avons trouvées sur un corpus d'un million de mots [9]. Elles confirment le fait que les répétitions disfluentes sont un phénomène très fréquent en oral spontané, qui mérite donc étude détaillée.

L'analyse des répétitions sera facilitée par l'introduction de deux termes, *répétable* et *répété*, conformément à l'emploi qui en est fait par Candéa [3] : « Toute répétition forme un bloc dans la parole qui comporte au minimum deux éléments : un premier élément que nous appellerons le 'répétable' et un deuxième élément, identique au premier, que nous appellerons le 'répété' ». Pour notre étude, nous avons considéré plusieurs positions : avant le répétable, entre le répétable et le ou les répétés, et enfin après le dernier répété. Dans l'exemple :

le: + le parc euh régional

l'allongement d'hésitation (noté à l'aide des deux points) et la pause silencieuse (notée par le signe +) sont tous deux situés entre le répétable et le répété.

2.1. Répétitions simples vs répétitions multiples

Le premier point abordé concerne les traits structuraux de la répétition. Une dichotomie a été opérée entre les répétitions dites *simples* qui comprennent un unique répété :

il n'y a pas par exemple les: les mêmes productions pédagogiques

et les répétitions dites *multiples* qui comprennent plusieurs répétés :

il y a des des des sujets qui ont attiré au monde rural

Nous dénombrons 124 répétitions simples (87,9%) et 17 multiples (12,1%). Les répétitions multiples, bien qu'évidemment moins nombreuses que les simples, sont tout de même très fréquentes. Nous observons sur ce corpus des répétitions faisant intervenir jusqu'à 4 répétés (5 éléments en tout) :

des fiches euh - euh sur euh sur la: la la la la botanique

2.2. Taille du répétable

Nous avons également pris en considération le nombre d'éléments contenus dans le répétable. Le répétable peut être composé d'un seul élément :

euh si vous pouvez le faire un: un Chambolle-Musigny

ou de plusieurs éléments :

et: c'est: + le coq au vin euh est un de ces: est un de ces plats

Les répétables composés d'un mot unique sont majoritaires puisqu'ils représentent 116 cas de répétitions sur 141 (82,3%), mais les répétables complexes (beaucoup plus difficiles à détecter dans les transcriptions d'un point de vue algorithmique) représentent tout de même une fréquence non négligeable (17,7%).

Il est intéressant de noter que, à part le cas suivant :

donc je pense que c'est quand même assez: - assez bien

tous les mots impliqués dans les répétitions sont monosyllabiques.

3. PAUSES ASSOCIÉES

Le corpus comporte 318 pauses silencieuses¹ non démarcatives (PS) [2] et 669 pauses remplies (PR). Parmi elles, 55 PS et 131 PR sont en contact avec une répétition, soit respectivement 18,3% et 20,5%.

Inversement, sur 141 répétitions, 109 sont associées à (au moins) une pause, silencieuse ou remplie, soit 77,3%. Le caractère simple ou complexe de la répétition n'a pas d'influence significative sur cette répartition (test de χ^2 à 1 d.d.l. : $p > 0,9$).

On note donc le fort degré d'association entre les deux types de phénomènes, répétitions et pauses, qui justifie pleinement l'étude de leur interaction.

¹ Nous n'avons retenu que les pauses silencieuses de durée supérieure à 200 ms, car nous considérons que les pauses plus brèves n'ont pas de valeur iconique précise et qu'elles sont souvent de nature respiratoire : [2], [11].

Sur les 109 répétitions associées, 52 le sont à au moins une PS (47,7%), 98 à au moins une PR (89,9%) et 41 répétitions (37,6%) sont associées aux deux marques (Tableau 1). La proportion de répétitions associées à un PS ou à une PR ne dépend pas du caractère simple ou complexe de la répétition (test de χ^2 à 1 d.d.l. : respectivement $p = 0,7$, $p = 0,6$)

Tableau 1. Nombre de répétitions associées à une pause.

	PR-	PR+	Total
PS-	32	57	89
PS+	11	41	52
Total	43	98	141

L'association avec une PR est particulièrement notable ; seules 11 répétitions (7,8 %) sont associées à une pause silencieuse sans pause remplie. Ce résultat est en totale contradiction avec [10] pour qui les PR ne sont que marginalement associées aux répétitions dans un corpus de dialogues en anglais. Il est possible que la différence de type de parole influe sur la répartition des marques associées aux répétitions : en situation de monologue le locuteur pourrait avoir plus recours à des stratégies de « remplissage » destinées à « occuper le terrain » puisqu'il doit produire un discours suivi et ne peut transférer la parole à l'interlocuteur. Cette hypothèse demande à être testée sur un corpus diversifié.

La figure 1 donne la répartition des répétitions par nombre de marques associées.

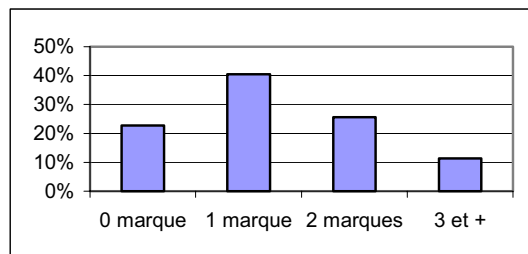


Figure 1 : Répartition des répétitions en fonction du nombre de marques associées.

Il est intéressant d'examiner plus en détail le cas des pauses remplies, dont on a vu qu'il représentait à peu près 90% des cas d'association : sur les 98 cas de répétitions associées à une PR, 89 contiennent un allongement (90,8%), 22 un *euh* (22,4%), mais seulement 9 (9,2%) un *euh* sans allongement. On note que la tendance s'inverse lorsque le nombre de marques augmente (figure 2), mais la proportion des allongements et des *euh* ne dépend pas du caractère simple ou complexe de la répétition (test de χ^2 à 1 d.d.l. : respectivement $p = 0,9$, $p = 0,3$).

Ceci indique que l'allongement est la pause remplie préférentiellement utilisée par le locuteur lorsqu'il doit répéter un fragment. Le *euh* n'intervient qu'en

deuxième recours, la plupart du temps lorsqu'un allongement s'est déjà produit. Les différents résultats de cette section indiquent donc une « gradation » dans la disfluence. On peut imaginer aisément que face à une difficulté momentanée (accès lexical, etc.), le locuteur essaie de « réparer » à l'aide d'une répétition sans marque aucune si c'est possible. Si ça n'est pas le cas il préfère, du point de vue de la « fluence » du discours utiliser une pause remplie plutôt qu'une pause silencieuse, plus disruptive, et parmi les pauses remplies, l'allongement est lui-même moins disruptif que les *euh*. L'absence d'interaction avec le caractère simple ou multiple de la répétition semble indiquer que les stratégies restent analogues pendant tout le temps du processus de « réparation ». Des études de perception pourraient mettre ces hypothèses à l'épreuve.

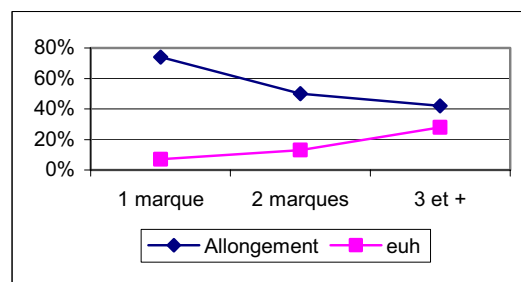


Figure 2 : Proportion des allongements et des *euh* en fonction du nombre de marques.

4. POSITIONS DES PAUSES

Les pauses, silencieuses ou remplies, peuvent apparaître à différentes positions dans la répétition : (#0) avant le répétable, (#1) entre le répétable et le premier répété (ou *point d'interruption*), (#2) entre les différents répétés (dans les cas de répétitions multiples) et enfin (#3) sur le dernier répété. Ces positions sont notées sur l'exemple suivant :

*et c'est toujours euh #0 en: #1 en #2 en #2 en #3
vigueur ça hein*

Dans cette séquence, un *euh* intervient avant le répétable, et un allongement entre le répétable et le premier répété (#3) ; les autres positions ne portent pas de marque associée.

Nous avons examiné plus en détail les phénomènes qui entourent le répétable (positions #0 et #1). En effet, des éléments expérimentaux [10] indiquent que les locuteurs perçoivent les répétitions après le point d'interruption (#1), sans qu'ils aient nécessairement recours à des informations syntaxiques ou sémantiques.

On note que sur 109 répétitions associées, 98 (89,9%) ont une marque au point d'interruption (#1), et c'est l'allongement qui est alors le plus souvent présent (84 cas, soit 85,7%). Dans 8 cas, la marque se trouve *avant* le répétable (en #0). Ces résultats

sont cohérents avec les expériences perceptives de [10] rappelées plus haut. Exemple :

on a quand même un laps de: de repos moi je suis du matin ma collègue est du soir

Tableau 2. Marque au point d'interruption

	Nb
Allongement	84
euh	7
PS	7
Non	11
Total	109

Sur les 98 cas de répétition avec marque(s) associée(s) juste après le répétable, on recense 64 cas (65,3%) où il n'y a qu'une marque (dont allongement pour 54 cas) :

les quatre départements de: de la région

28 cas font intervenir deux marques, et la combinaison de loin la plus fréquente est l'allongement suivi de la pause silencieuse (25 cas) :

tu as quand même des : + des tissus qu'il faut particulièrement connaître

Seuls 6 cas font intervenir plus de deux marques :

difficilement: + euh prévisibles: et: euh - et quantifiables

Aucun cas ne fait intervenir plus de trois marques consécutives : peut-être ce nombre constitue-t-il un seuil à partir duquel les disfluences sont perçues comme pathologiques (bégaiement, aphasies). La comparaison avec des corpus de parole pathologique, selon les critères méthodologiques et quantitatifs que nous avons esquissés dans ce papier, pourrait, de ce point de vue, être instructive.

5. CONCLUSION

Nous avons montré dans cette étude, à l'aide d'un corpus de près d'une heure de parole spontanée, que les répétitions sont fortement associées aux pauses, et plus particulièrement aux pauses remplies. Dans ce cas, c'est l'allongement syllabique qui est la marque préférentiellement associée à la répétition, le *euh* intervenant plutôt en deuxième recours, la plupart du temps lorsqu'un allongement s'est déjà produit. Ces résultats semblent indiquer une gradation dans les marques associées aux répétitions, allongement, *euh* et pause silencieuse présentant un degré croissant de disruptivité. Enfin, nos résultats montrent que les pauses silencieuses ou remplies affectent le point d'interruption dans une très grande majorité de cas, ce qui est cohérent avec les études antérieures sur la perception des répétitions par les locuteurs, qui est très rapide après ce point d'interruption et semble reposer sur des indices acoustiques plutôt que syntaxiques ou sémantiques.

BIBLIOGRAPHIE

- [1] C. Blanche-Benveniste. La naissance des syntagmes dans les hésitations et répétitions du parler. In J.L. Araoui (Éd.), *Le sens et la mesure. Hommages à Benoît de Cornulier* (pp. 40-55). Paris : Éditions Honoré Champion. 2003.
- [2] E. Campione. *Étiquetage semi-automatique de la prosodie dans les corpus oraux: algorithmes et méthodologie*. Thèse d'état, Université de Provence, Aix-en-Provence, France. 2001.
- [3] M. Candéa. *Contribution à l'étude des pauses silencieuses et des phénomènes dits d'« hésitation » en français oral spontané*. Thèse d'état, Université Paris III, Paris, France. 2000.
- [4] M. Candéa. Les euh et les allongements dits « d'hésitation » : deux phénomènes soumis à certaines contraintes en français oral non lu. *XXIIIèmes Journées d'étude sur la Parole* (pp. 73-76). Aussois, 19-23 juin. 2000.
- [5] A. Cutler. The recognition of spoken words with variable representations, *Proceedings of the ESCA Workshop on the Sound Patterns of Spontaneous Speech* (pp. 83-92). Aix-en-Provence, France, 1998.
- [6] D., Duez. The aim of SPoSS, *Proceedings of the ESCA Workshop on the Sound Patterns of Spontaneous Speech* (pp.VII-IX). Aix-en-Provence, France. 1998.
- [7] Équipe DELIC. Présentation du corpus de référence du français parlé. *Recherches sur le français parlé, 18*, à paraître.
- [8] M. Goto, K. Itou, & S. A. Hayamizu. A Real-time Filled Pause Detection System for Spontaneous Speech Recognition. *Proceedings of the 6th European Conference on Speech Communication and Technology (Eurospeech '99)* (pp. 227-230). Budapest. 1999.
- [9] S. Henry. Étude des répétitions en français parlé spontané pour les technologies de la parole. *TALN/RECITAL 2002* (pp. 467-476), ATALA : Nancy. 2002.
- [10] R.J. Lickley. *Detecting Disfluency in Spontaneous Speech*. PhD Thesis, University of Edinburgh. 1994. [<http://www.ling.ed.ac.uk/~robin>]
- [11] M. A. Morel & L. Danon-Boileau. *Grammaire de l'intonation. L'exemple du français*. Paris : Ophrys. 1998.