COMPLEXITY CUES OR ATTENTION TRIGGERS? REPETITIONS AND EDITING TERMS FOR NATIVE SPEAKERS OF FRENCH

Iulia Grosman

Université catholique de Louvain
iulia.grosman@uclouvain.be

ABSTRACT

A growing stream of research shows evidence of the metalinguistic information that disfluencies (*silent and filled pauses, repetitions, false-starts, repairs, etc.*) can display to listeners. As a result, disfluencies may work as fluent devices. By means of a decision task latencies, this study investigates whether lexical repetition co-occurring with an editing term affects the perception of native speakers of French. There is a lack of consensus in the literature: do disfluencies trigger conceptual priming of complex entity or act simply as attention cues? Results from multiple analysis of variance and linear mixed-effect modelling show that the presence of a disfluency triggers a faster response from the participant, however complex the following noun-phrase might be, supporting the hypothesis that repetition and co-occurring editing terms act as cognitive signposts rather than as cues of complexity of an upcoming event.

Keywords: Disfluencies, reaction time, perception, prosody, repetitions, French.

1. INTRODUCTION

Literature qualifies oral speech production as notoriously disfluent (i.e. containing silent and filled pauses, false-start, prolongations, repetitions, substitutions, deletions, insertions, etc.): disfluency rate range from 1 up to 6% of speech in conversation [9], varying according to multiple parameters as cognitive, social and situational factors that may interact to affect speech production (age, gender, familiarity of the subject, etc.) [15, 4]. In perception, most of these occurrences do not lead to comprehension issue. Research agrees that disfluencies may work as fluent devices as they depict information about the speaker's confidence, his planning difficulties, and are even used as tools to coordinate conversational interaction [18].

The goal of a cognitive approach to disfluencies would be to determine on the one hand which factors are responsible for disfluencies' production in everyday speech and on the other hand to understand how speech containing disfluencies would affect the processes of the listeners in terms of comprehension, prediction, memory or attention [5]. Accordingly, in Segalowitz' terminology, the hypothesis underling this contribution works towards a better understanding of the *perceived fluency* by native French speakers or "the inferences hearers make about speakers' cognitive fluency based on their perceptions of the utterance fluency" [17]. Within this framework, the utterance fluency is accoustically measured on the generated audio stimuli while the experimental design connects it to the cognitive fluency (or the "efficiency of operation of the underlying processes responsible for the production of utterances") [17] by measuring latency of responses (RT) to a decision task with sentences including a disfluency, namely a lexical repetition and a co-occuring editing term.

1.1. The present study

Repetitions are the second most frequent marks of (dis)fluency, after filled and silent pauses: they represent an average 25% of all disfluency types in various discourse genres [12]. Disfluency have been sometimes depict as a tool to capture attention and/or a cue for conceptual priming of a complex entity [2, 19]. As priming facilitates perceptual processing, some disfluencies would make utterances more fluent. Literature already attests for the structuring function of silent and filled pauses, especially in the hearers' expectations about the complexity of an upcoming event [7, 19] and its relevance [2]. The field of linguistics tends to assume that all disfluencies types display metalinguistic information [6] but neurological work attested that "not all disfluencies are equal": they do not affect underlying cognitive processes of speech perception identically in terms of prediction, attention, comprehension or memorization [13]. However, apart from works by [9], [16] and [19] little experimental evidence exists on the hypothetical coherence and fluency function of disfluencies other than filled pauses (FP), such as repetitions, false starts or reformulations.

1.2. Hypothesis

According to the fact that, as the simplest exemplars of a target category, visually non compound stimuli (a sphere without *vs* with arrows) are inherently more fluent (i.e. easier to process) [1], we predict that RTs to simple shapes would be shorter than to compound ones. Secondly, we hypothesize that RTs to compound shapes would be shorter when preceded by a repetition, than when there is none, as they might give a cue to the speaker's planning difficulty and allow some time to predict the content of the upcoming utterance [19]. Thirdly, due to FPs' relative salience in the speech flow, we predict an additional result of disfluencies: repetitions containing an editing term will result in quicker response than simple repetitions.

2. METHOD

2.1. Participants and Design

Inspired by [19], the within repeated measures design invited 62 native speakers of French to listen once to 96 sentences describing both a simple and a compound shape (*complexity factor* – as verbally, linguistic complexity in surface generally "includes more words"). Their task was to listen once to each sentence and press the left or right key as soon as, from a displayed pair on a computer screen, they identified the shape corresponding to the description.

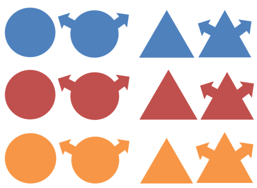


Figure 1: Example of 6 paired stimuli

The paired shapes (Figure 1) were always of the same type and colour so that these features would not affect the task. The simple shape was always on the left side and the compound one on the right. The actual measurement of the exhaustion of the free capacity (*log*(RT)) is used as a concrete definition of "retrieval disfluency", the speed with which objects are accessed. They were automatically extracted by subtracting the word onset-time (disambiguating syllable) from the RT measured from the start of the sound files.

2.2. Material

In summary, the experiment consisted of 96 items (24 similar items in 4 different conditions). The shapes in the sentences were either preceded by a repetition including (1) a FP (*original filler repetition*), (2) a silent pause of same duration (*repetition with silent pauses*), (3) no pause (*contiguous repetition*) or (4) no repetition at all (*no repetition*) (*fluency factor*). In addition, each utterance contained a phrase describing a colour and a shape in this order (e.g. "Bring me the uhm the red circular paper from the living room" (simple shape with filler repetition), "Bring me the the orange circular paper with arrows from office" (compound shape with contiguous repetition)). All sentences were transcribed and phonetically aligned under Praat [3] using EasyAlign [11].

A thorough prosodic analysis of the 24 original sentences including a filler and a repetition were pronounced twice by a French native speaker, allowing the selection of stimuli with very little prosodic variation. The original utterances including a filler repetition were used to generate the 3 other sets of conditions by editing the appropriate audio segment. Using Prosogram v2.9 [14], the 24 stimuli showing the least prosodic variation were selected according to the smallest relative distance from the mean and median of : the speech rate, the pitch range in semitones, the top and bottom pitch, the pitch mean and median, the mean proportion of glissando, rises and falls of each audio-sample, the duration of the schwa in "le" (the first "le" in "Apporte-moi le euh le papier bleu triangulaire qui est dans le bureau") and the relative duration of the filler "euh".

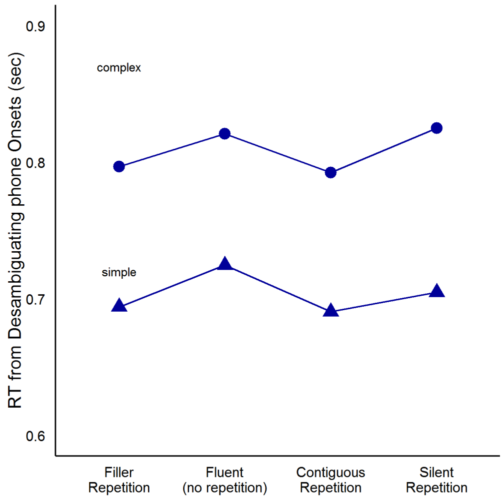
T-tests for the two levels of complexity were separately performed for each prosodic feature and none of them was proven significant apart from the absolute speech time of the stimuli between the two complexity conditions. On average compound shape stimuli are 600 ms. longer than simple stimuli. Filled pauses in the simple condition were not longer than in the compound shape condition. Schwas before the repetition were uttered with a homogeneous lengthening across conditions ($M = 350$ ms., $SD = 38$ ms.), in accordance with the tendency observed on the first part of repetitions and segments preceding a FP [18].

3. MAIN RESULTS

3.1. Error Rates and Types

Overall, the mean rate of correct responses was 98.147% ($SD = 1.73$), results comparable with [19]. Precocious responses (i.e. anticipated responses made before the disambiguating sound) were discarded and treated as errors. Error rates were significantly different in the simple and compound conditions ($t(4.92) = 4.92, p < 0.005$), supporting the fact that the compound shapes were indeed of higher complexity.

Figure 2: RT across conditions



Roughly, the participants pressed a key between 0.55 and 0.70 sec. after the onset of the disambiguating syllable (Fig. 2). As in [19], a two way repeated measures analysis of variance (ANOVA) revealed an expected main effect of the complexity factor ($F(1,61) = 26.76, p < 0.001$). As the first hypothesis predicted, latency is indeed shorter for simple NPs than for complex ones. Moreover, contrary to the homologue Japanese study, we did find a significance of the main effect of the fluency factor as post-hoc comparisons using t-test with Bonferroni correction indicated that the mean scores within fluency conditions were statistically significant ($F(3,183) = 3.87, p < 0.02$).

3.2. Linear mixed-effect Modeling

A linear-mixed effect model was selected based on AIC and log-likelihood ($pseudo - R^2 = 57\%$). Aside from random items and subjects, the model included complexity and fluency factors as well as the position of the item in the experiment and the results of the participants to the Cognitive Reflection Test (CRT), a simple measure of cognitive ability associated with analytic thinking [10]. The latter parameter adds robustness to the model as it is highly correlated with the intercept ($r = 0.821$).Table 1 shows the results for fixed effects (with default contrast treatment). The intercepts represent the original *filler repetition in the compound shape condition* ($M = (e^{660.82/1000}) = 1.94$ sec.). Increased knowledge about participants (age, gender, musical knowledge or bilingual aptitude) were not included in the model as they did not increase the explained variance.

Firstly, the negative estimates on the CRT reveals that the lower the participants score on the Cognitive Reflection Test (i.e. make assumption or anticipate on the upcoming content), the fastest participants react to the stimuli. Overall, RTs were indeed 1.65 sec. shorter for simple NPs than for complex ones. Moreover, the fluent condition elicited longer RTs than all other conditions, for both compound and simple shapes ($+1.01$ sec., $p < 0.05$).

Table 1: Fixed effects

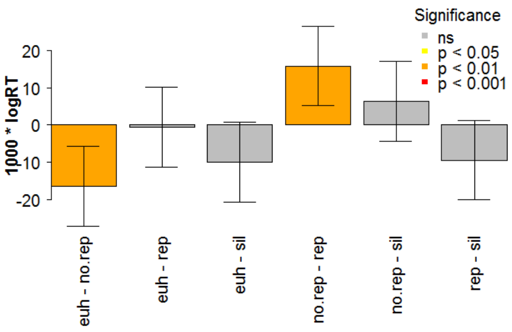
	Estimate (log)	Estimate (sec)	Df	t value	Sig. (p)
(Intercept)	660.82	1.94	66.13	20.62	<0.001***
Filler rep.- without arrows	-50.25	-1.65	90.52	-10.78	<0.001***
No rep. (fluent)	+13.95	+1.014	90.20	2.12	0.03*
Contiguous rep.	+0.60	1.000	90.23	0.09	0.93
Silent rep.	+10.79	+1.011	90.80	1.64	0.11
CRT scores	-1.79	-1.001	60.00	-4.01	<0.001***

Little variance is attested between other fluency levels. Post-hoc tests did show a significant difference among fluency conditions in the simple condition, ($F(3,58) = 2.79, p < 0.05$), but no significant difference in the compound shape condition ($F(3,58) = 1.9, p = 0.131$). These results contrast with [19] who reported and inverted significance. As a result, the second hypothesis according to which RTs to compound shapes would be shorter when there is a repetition than when there is none, and that RTs to simple shapes might be longer with a repetition than without can not be fully supported. In fact, due to the lack of interaction between fluency and complexity factors, we can solely conclude that, no matter the complexity condition, overall sentences with repetitions did induce shorter RTs than fluent ones. No matter the complexity condition, sentences with repetitions did induce shorter RTs than fluent ones.

The third hypothesis is partially confirmed. Overall, repetitions including a FP as an editing term show shorter RTs than their homologue silent version ($t(60) = 4.018, p < 0.001$), thus leaning in favour of the interpretation of the filler and repetition additive weight. Nonetheless, there is no strong difference in latency between the filler condition and contiguous repetition condition, encouraging to consider that the decrease in latency may be due more to the presence of a repetition than to the presence of FP. Moreover, within the simple condition, replacing the FP by a silence doesn't significantly increase the latency, same is said when erasing the FP altogether ($+1.011$ and $+1.00$ sec. respectively).

To summarize, Figure 3 illustrates the differences in least square means across fluency conditions and their confidence interval (based on Kenwar-Roger approximation). The least mean square differences were also found shorter for sentences including a

Figure 3: Differences in Least Square means for fluency factor with 95% Confidence Intervals



repetition with silence than for fluent ones, but not significantly. In decreasing order, fluent sentences, repetitions with silent pauses and contiguous repetitions show longer RTs than repetitions with filler. Furthermore, latency to contiguous repetitions are also significantly shorter than their fluent version ($p < 0.001$ for no.rep-sil).

4. DISCUSSION AND CONCLUSION

The study tested the hypothesis according to which hearers expect rather long or complex phrases when preceded by a disfluency. Our results do not entirely converge with previous studies: no interaction was found between fluency and complexity factors. Repetition and its editing terms do not seem to facilitate processing or bias hearers' expectations when complexity rises which does go in line with [13] and [8]. Significant difference in RTs was observed within the fluency condition for simple shapes but not for compound ones. Results show that native speakers of French are quicker to respond to compound and simple shapes when there is a contiguous repetition or a repetition with a FP. In decreasing order, fluent sentences, repetitions with silent pauses and contiguous repetitions show longer RTs than repetitions with filler. The findings seem to support that the combination of repetition with FP does not have an additive effect. According to the results, repetitions and FPs would cause hearers to react faster; they can thus be interpreted as a cue to catch their attention rather than as a complexity cue.

5. REFERENCES

[1] Alter, A. L., Oppenheimer, D. M. 2009. Uniting the tribes of fluency to form a metacognitive nation. *Pers. Soc. Psychol. Rev.* 13(3), 219–35.
[2] Arnold, J. E., Fagnano, M., Tanenhaus, M. K. 2003. Disfluencies signal theee, um, new information. *Journal of psycholinguistic research* 32(1), 25–36.
[3] Boersma, P., Weenink, D. 2015. Praat: doing phonetics by computer.
[4] Bortfeld, H., Leon, S., Bloom, J., Schober, M., Brennan, S. 2001. Disfluency Rates in Conversation : Effects of Age, Relationship, Topic, Role, and Gender. *Lang. Speech* 44(2), 123–147.
[5] Bosker, H. R., Quené, H., Sanders, T., de Jong, N. H. 2014. The Perception of Fluency in Native and Nonnative Speech: Native and Nonnative Fluency Perception. *Lang. Learn.* 64(3), 579–614.
[6] Brennan, S. E., Williams, M. 1995. The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *J. Mem. Lang.* 34, 383–398.
[7] Clark, H. H., Tree, J. E. F. 2002. Using uh and um in spontaneous speaking. *Cognition* 84(1), 73–111.
[8] Corley, M., Hartsuiker, R. J. 2011. Why Um Helps Auditory Word Recognition: The Temporal Delay Hypothesis. *PLoS ONE* 6(5), e19792.
[9] Fox Tree, J. E. 1995. The effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech. *J. Mem. Lang.* 34(6), 709–738.
[10] Frederick, S. 2005. Cognitive Reflection and Decision Making. *J. Econ. Perspect.* 19(4), 25–42.
[11] Goldman, J.-P. 2011. EasyAlign: an automatic phonetic alignment tool under Praat. *Proc. of Interspeech 2011.* Firenze, Italy. ISCA Archive. 3233–3236.
[12] Grosman, I. 2014. Dynamics of disfluencies within Basic Discourse Units. *NCLF* 31, 45–50.
[13] MacGregor, L. J., Corley, M., Donaldson, D. I. 2009. Not all disfluencies are are equal: The effects of disfluent repetitions on language comprehension. *Brain Lang.* 111(1), 36–45.
[14] Mertens, P. 2004. Un outil pour la transcription de la prosodie dans les corpus oraux. *Traitement Automatique des langues* 45(2), 109–130.
[15] Moniz, H., Batista, F., Mata, A. I., Trancoso, I. 2014. Speaking style effects in the production of disfluencies. *Speech Commu.* 65, 20–35.
[16] Moniz, H., Trancoso, I., Mata, A. I. 2009. Classification of disfluent phenomena as fluent communicative devices in specific prosodic contexts. Schuller, B., Steidl, S., Batliner, A., (eds), *Proc. of Interspeech 2009.* Brighton, U.K. ISCA Archive. 1719–1722.
[17] Segalowitz, N. 2010. *Cognitive Bases of Second Language Fluency*. New York: Routledge.
[18] Shriberg, E. 1995. Spontaneous Speech : How People Really Talk and Why Engineers Should Care. *Proc. of Eurospeech.* ISCA Archive. 1781–1784.
[19] Watanabe, M., Hirose, K., Den, Y., Minematsu, N. 2008. Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Commun.* 50(2), 81–94.