DISCOURSE COHESION AND ITS PROSODIC MARKING IN FRENCH: INTERACTIONS BETWEEN INTONATION UNIT ONSETS AND ANAPHORIC PRONOUNS IN SPEECH PERCEPTION.

Cyril Auran

Laboratoire *Savoirs, Textes, Langage*, UMR 8163 Université Lille 3 - CNRS
cyril.auran@univ-lille3.fr

ABSTRACT

This study is part of a wider project analyzing the roles of prosody and anaphora in discourse organization in English and French, and linking production and perception. More specifically, the aim of this paper is twofold: it explores the interactions of prosody and anaphora in French discourse and their consequences in terms of cognitive processing cost for the hearer; these results are based on an indirect methodology which constitutes the second aspect of this work. More specifically, this study explores the interplay hypothesis between pronominal anaphora and the phonetic realization of intonation unit onsets using cross-modal semantic priming in French.

Keywords: Prosody, discourse, cohesion, anaphora, resetting.

1. INTRODUCTION

Exploring both the prosodic and lexical marking of discourse structure in British English, [3] has shown that unstressed anaphoric pronouns seem to influence the phonetic realization of intonation unit onsets. This results in increased onset values (1.8 semitones on average), a paradoxical fact with respect to the more traditional "onset depression" associated with discourse continuity and, conversely, the resettings typically observed at initial topic boundaries (cf. [14] for English, and [12] or [7] for French).

These apparently paradoxical results receive an explanation within a model arbitrating between raising production constraints (onset raising in relation with anticipations in speech production – pre-planning or look-ahead – cf. [13]) and lowering constraints (onset depression in relation with topic continuity marking, cf. [14]). Topic continuity being guaranteed through anaphora, the lowering constraints lose relative weight, thus leading to onset raising (lookahead was statistically validated through the analysis of intonation unit durations within the Aix-MARSEC corpus; cf. [2]).

The present study constitutes a follow-up and a dual transposition of this research; here, the interactions of anaphora and onset values are analyzed from a perceptual perspective in French.

This work explores the central hypothesis according to which speaker and hearer endorse asymmetrical roles which lead to the production of speaker-optimized forms requiring extra cognitive processing on the hearer's side. Our aim is to investigate the impact of the interactions identified in speech production on speech perception. We resort to cross-modal semantic priming in order to quantify the influence of synthetically raised onset values for discourse segments containing unstressed anaphoric pronouns (the paradoxical pattern observed by [3]) on the hearer's cognitive load.

As shall be made explicit in the rest of this paper, we thus propose both an analysis of the role played by prosody in the cognitive processing of discourse structure and an original experimental methodology allowing the indirect quantification of cognitive cost.

2. HYPOTHESES

2.1. Speaker and hearer as asymmetrical roles

Discourse is seen here as an interactive cognitive and linguistic process between speaker and hearer (a position very similar to that exposed for example by [6]). However, we insist on the asymmetrical nature of the speaker-hearer relationship. Indeed, [4] have shown that intelligibility does not actually depend on the hearer's hypothesized or actual cognitive state, but rather on the speaker's. Production effort and phonetic detail thus seem to be directly related to the speaker's cognitive model of discourse, and hardly take the hearer's state into account.

These results are in line with the model proposed in [3], who postulate that onset raising solely depends on competing constraints on the speaker's side, regardless of the hearer's interpretive efforts (in relation with the interpretation of clashing cues such as higher onsets and anaphoric topic continuity cues).

2.2. Onset perception and cognitive cost

This study more specifically investigates the very existence of these extra interpretive efforts for the hearer. We thus hypothesize that onset raising is not only perceived by the hearer, but also interpreted as a mark of topic discontinuity clashing with the continuity induced by the use of unstressed anaphoric pronouns (henceforth "extra interpretive efforts hypothesis", or "EIE hypothesis").

3. METHODOLOGY

The test of the EIE hypothesis was carried out indirectly using the cross-modal semantic priming method. We will present the general principles of this method, before detailing its exploitation in the study. The experimental setup will then be evoked, with focus on stimulus generation and the experimental procedure.

3.1. Cross-modal semantic priming

3.1.1. The classical method and its applications

Semantic priming is a psycho-linguistic experimental method derived from the "probe verification" paradigm (cf. [10] for an overview). In cross-modal semantic priming, subjects, while listening to an audio stimulus, have to make lexical decisions (word/non-word) about letter sequences (*targets*) which are displayed on a computer screen at specific moments (between 250 and 700 ms after a given lexical element, or *anchor*, in the audio stimulus). Response times have been demonstrated to reflect cognitive activation of the anchor and heavily depend on the actual lexical status of the target (with non-words inducing slower responses) but also on semantic relatedness between target and anchor (semantically related targets systematically being recognized faster than unrelated targets).

Cross-modal semantic priming has been used extensively for the tracking of activation in relation with anaphoric (co)reference, where use of an anaphor as anchor has been shown to induce priming with a target semantically related to its antecedent (cf. for instance [9]).

3.1.2. An indirect application

The present study uses cross-modal semantic priming as a method for the identification of cognitive load. We used this method in relation with the manipulation of the audio stimuli in order to measure the cognitive processing cost induced by the interpretation of raised onsets.

Semantic relatedness between the target and the anchor (or the anchor's antecedent in the case of an anaphoric anchor) significantly diminishes response time in the lexical decision task; all else being equal, the differences in response time related to onset manipulation can therefore be linked with the cognitive effort involved in the interpretation of the clashing prosodic and anaphoric cues. In this perspective, the EIE hypothesis, if validated, is expected to correspond to a cancellation of semantic priming in relation with onset raising.

3.2. Experimental setup

3.2.1. Stimulus generation

Three-sentence texts were devised taking into account the classical parameters influencing semantic priming (and particularly length in syllables, lexical frequency, syntactic complexity). (1) below constitutes a typical example.

(1) La secrétaire est en train de débrancher l'ordinateur. Dans cinq minutes, elle va ranger ses dossiers une dernière fois. L'année prochaine, elle doit occuper un nouveau poste au Mans. [*The secretary is unplugging the computer. In fives minutes' time, she'll tidy up her files for the last time. Next year, she is to have a new position in Le Mans.*]

The texts were numerically recorded in a quiet room at the *Laboratoire Parole et Langage* using a Tascam US-122 sound interface and AKG C444 PP headset microphone. Each text was recorded as a 16 bit PCM WAV mono sound file, with a 44100 Hz sampling frequency.

A total of 16 original texts were thus recorded, and served as input for PSOLA resynthesis within Praat ([5]).

The resynthesis itself involved three steps:

- segmentation into intonation units and manual selection of the relevant onset (first syllable of the intonation unit in French, cf. [12] either in sentence 2 (8 texts) or sentence 3 (8 texts);
- MoMel modeling ([8]) of the F0 curve;
- automatic generation of 17 versions of each original sound file, corresponding to 16 raised-onset files (0.25-tone steps) and one "unmodified" resynthesized version (in order to neutralize bias due to resynthesis); a total of 272 stimuli were thus generated in the process.

Targets (related words, unrelated words and non-words) were then selected for each anchor within the texts, taking into account length (in terms of constituent syllables) and frequency (for words); (2) below is an example of targets for the anchor "secrétaire" (*secretary*):

(2) courrier (*mail*; related word) / cheval (*horse*; unrelated word) / luttrel (non-word).

3.2.2. Experimental procedure

12 native speakers of French from the Aix-en-Provence area were involved in the experiment. Each subject had to accomplish 2 tasks:

- Lexical decision task: the subjects had to categorize the target sequences as word/non-word as fast as possible.
- Memorization task: the subjects were lead to believe that questions would be asked at the end of the session, concerning specific passages from the texts.

The second task was actually a decoy intended to avoid a potential bias in the results due to the subjects' focusing on the visual input (targets) exclusively, without any attention drawn towards the audio stimuli.

The whole experiment was carried out within the PERCEVAL environment (cf. [1]). The sessions took place in a quiet room at the *Laboratoire Parole et Langage*, and were divided into 2 sub-sessions of 30 minutes each.

The visual targets were synchronized with the audio stimuli and were displayed on the computer's screen 500 ms after the end of the anchors. Response time was then automatically recorded for each stimulus with a precision of millisecond order.

The data was then loaded into the R Software for statistical analysis, with z-score normalized response time as the dependent variable, and onset raising, sentence position within the text and subject as the main independent variables.

4. RESULTS

As can be seen in table 1 below, our first global results confirm the significant effects of semantic priming (anova: F=30.615 ; p=8.61e-14*), with non-words, unrelated words and related words in decreasing order of normalized response time.

Table 1: Normalized response time as a function of target type.

Target type	Response time
Non-word	0.250913
Unrelated word	0.066619
Related word	- 0.184294

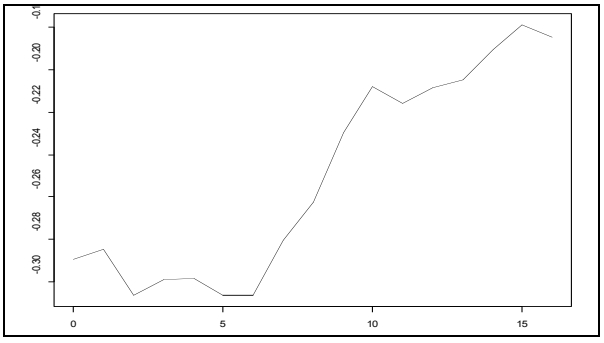
However, contrary to our predictions, a global two-factor anova indicates an apparent lack of significant influence of onset raising on response time. The interaction of target type and onset raising, however, displays a significant effect, which requires further analysis.

Table 2: 2-way anova (target type * onset raising).

Factor	F	p-value
Target type	30.615	8.61e-14 *
Onset raising	1.774	0.1830672
Target type:onset raising	7.334	0.0006736*

Formal analysis of onset raising in the "related word" condition, which particularly interests us here, reveals a significant effect (anova: F=7.9828 ; p=0.004887*).

Figure 1: Normalized response time as a function of onset raising in the "related word" condition.



The analysis of figure 1 above seems to indicate a non-linear influence of onset raising on response time in the "related word" condition, with a particularly drastic influence for an onset raising

comprised between steps 5 and 10 (1.25 to 2.5 tones) of the resynthesis. This visual inspection is confirmed by a more formal analysis, involving a segmented anova (with 3 ranges: from step 0 to step 4, from step 5 to step 10 and over step 10) and summarized in table 3 below.

Table 3: p-values for the segmented one-way anova (onset raising).

Factor	0-4	5-10	> 10
Onset raising	0.1914	0.04564 *	0.05289

5. DISCUSSION

These results obviously constitute but a first approximation of the phenomenon under scrutiny, and will be submitted to finer-grained statistical analyses in forthcoming work. Nevertheless, in their present state, they also seem to provide positive evidence in favour of our EIE hypothesis. The data indicates that onset raising above a 1.25-tone threshold actually cancels semantic priming and its benefit in terms of response time: this does suggest an increase in cognitive load hampering lexical recognition.

The distribution of the results gives rise to two questions. First, the absence of effect within the lower and upper ranges requires more explicit explanation. Our hypothesis concerning this relies on the absence of perception below step 5, and the unnatural distortions related to resynthesis above step 10 (in part due to PSOLA artifacts). This obviously remains to be investigated systematically. A second, and obviously connected question concerns the perceptual threshold which, apparently situated between 1.25 and 2.5 tones, remains to be accurately defined. [2], proposes a differential perception experiment with very similar material; the results of this experiment indicate a threshold of 1.5 tones, fully compatible with the data presented here.

More generally, these results and the questions which we have alluded to naturally lead to the gradual or categorical nature of this effect: even though our result pattern seems compatible with the sigmoid commonly observed in categorical perception, formal fitting remains to be tested. A classification of this effect as categorical would have interesting consequences on the phonological modeling of French intonation, and particularly on the status of onsets in deep phonology (cf. [11] for an overview, and particularly p. 154 for a mention of this issue).

6. CONCLUSION

This exploratory study has investigated the influence of intonation unit onset raising on speech perception of cohesive discourse in French. More particularly, through the use of cross-modal semantic priming techniques, it has provided indirect evidence and a potential quantification method concerning increased cognitive load in the processing of these conflicting cues.

The extension of this work, which is currently under way, has obvious consequences not only on our understanding of discourse, but also regarding key issues in prosodic phonology in French, particularly concerning intonational structure.

7. REFERENCES

[1] André, C. Ghio, A., Cavé, C., Teston, B. 2003. PERCEVAL: a Computer-Driven System for Experimentation on Auditory and Visual Perception. *Proc. 15th ICPhS* Barcelona, 1421-1424.

[2] Auran, C. 2004. *Prosodie et anaphore dans le discours en anglais et en français : cohésion et attribution référentielle*. PhD. dissertation, Université de Provence, France and Laboratoire Parole et Langage, UMR 6057, CNRS.

[3] Auran, C., Hirst D. 2004. Anaphora, Connectors and Resetting: Prosodic and Pragmatic Parameters Interactions in the Marking of Discourse Structure. *Proc. Speech Prosody 2004* Nara, 259-262.

[4] Bard, E., Anderson, A., Sotillo, C., Aylett, M. Doherty-Sneddon, G., Newlands, A. 2000. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language* 42, 1-22

[5] Boersma, P., Weenink, D. 2006. Praat: doing phonetics by computer (Version 4.5.14). http://www.praat.org

[6] Cornish, F. 1999. *Anaphora, Discourse and Understanding*. Oxford: Oxford University Press.

[7] Grobet, A. 2002. *L'identification des topiques dans les dialogues.* Brussels: De-Boeck-Duculot.

[8] Hirst, D., Di Cristo A., Espesser, R. 2000. Levels of representation and levels of analysis for the description of intonation systems. In: Horne, M. (ed.), *Prosody: Theory and Experiement.* Dordrecht: Kluwer, 51-87.

[9] Nicol, J. 1988. *Coreference processing during sentence comprehension*. Ph.D. Dissertation, M.I.T.

[10] Nicol, J., Swinney, D 2003. The Psycholinguistics of Anaphora. In: Barrs, A. (ed.), *Anaphora. A reference guide*. Oxford, Blackwell, 72-104.

[11] Post, B. 2000. *Tonal and phrasal structures in French intonation*. The Hague: Holland Academic Graphics.

[12] Simon, A.-C., Grobet, A. 2001. Réinitialisations (resetting) et unites prosodiques maximales : une evidence ? *Proc. Journées Prosodie 2001* Grenoble, 79-83.

[13] Swerts, M. Strangert E., Heldner M. 1996. F0 declination in spontaneous and read-aloud speech. *TMH-QPSR* 2, 23-24.

[14] Wichmann, A. 2000. Intonation in Text and Discourse: Beginnings, Middles and Ends. Pearson Education, London: Longman.