Fine phonetic details for discourse marker disambiguation: a corpus-based investigation

Yaru Wu, Mathilde Hutin, Ioana Vasilescu, Lori Lamel, Martine Adda-Decker, Liesbeth Degand

Fine phonetic details for DM disambiguation: a corpus-based investigation

Yaru Wu[1,2], *Mathilde Hutin*[1], *Ioana Vasilescu*[1], *Lori Lamel*[1], *Martine Adda-Decker*[1,2], *Liesbeth Degand*[3]

[1]*Université Paris-Saclay, CNRS, LISN, Bât. 507, rue du Belvédère, 91405 Orsay, France*

[2]*Laboratoire de Phonétique et Phonologie (UMR7018, CNRS-Sorbonne Nouvelle), France*

[3]*Université catholique de Louvain, Institut Langage et Communication*

Abstract

In this study we examine phonetic variation of discourse markers in French, using for this purpose the 4-hour richly annotated LOCAS-F corpus. Both linguistic factors and stylistic variables are considered: speech style, part-of-speech category, mean phone duration and vowel formant distributions with respect to the word status. The results show that the use of discourse markers increases with the degree of spontaneity of the speech. Coordinating conjunctions are the part-of-speech which is most frequently used for discourse markers. Moreover, the mean phone duration tends to be shorter and the vowel space more centralized when words are employed as discourse markers, suggesting that discourse markers undergo hypoarticulation and, more generally, reduction.

Introduction

Languages are known to be ambiguous, and the primary sources of ambiguity in the lexicon are polysemy and homophony, both resulting in similar phonological forms referring to different objects.

Polysemy refers to similar phonological forms with different, yet related meanings. For example, in French, /kafe/ refers to the coffee plant, to the drink made of this plant and to the place where this beverage is served. So-called "word sense disambiguation", i.e. "the ability to computationally determine which sense of a word is activated by its use in a particular context" (Navigli, 2009) is a well-known challenge for word recognition systems (Schütze, 1998; Jurafsky and Martin, 2008). It is solved as best as possible with various methodologies that nevertheless all still need to mature (Singh and Saraswat, 2019).

Homophony refers to two identical phonological forms that have different meanings, and sometimes even belong to different grammatical categories. For instance, in French, the sequence /sɑ̃/ can stand either for the numeral *cent*, "hundred", for the noun *sang*, "blood", or the conjugated verb *sens*/*sent*, "(I, you, he/she/it) smell(s)". Homophones or near homophones, sometimes called "phonological neighbors", represent approximately 4% of the words in the world's languages (according to Dautriche (2015)'s study on 67 languages). Moreover, studies have shown that the acquisition of a novel word is harder when the new entry has a homophone in the existing lexicon of the learner (Swingley and Aslin, 2007), especially if the new entry and the existing homophone belong to the same grammatical category (Dautriche et al., 2015), and that the identification of a new homophone is more difficult when the semantic context it appears in is close to the lexical field of the existing homophone (Cole et al., 1978). In sum: phonological, grammatical and semantic proximity greatly complexify homophone disambiguation.

This paper proposes to add to the literature on disambiguation by investigating a question that seems to have been less studied: that of achieving disambiguation by modeling fine-grained phonetic details. We know that phonetic variation depends on word frequency (Pierrehumbert, 2008; Phillips, 1984). For instance, it has been shown that word-frequency influences word-duration, allowing homophonous nouns and verbs in English to be distinguished (Lohmann and Conwell, 2020), and that more frequent words appear to have more centralized vowels due to shortening (Dinkin, 2008). Phonetic variation also depends on grammatical function in that function words are shorter, acoustically poorer and more prone to reduction (Adda-Decker and Snoeren, 2011; Ernestus, 2011). Finally, we know that phonetic variation depends on pragmatic usage, since fine prosodic cues can help disambiguate the function of words such as discourse markers (see Didirková et al. (2018, 2019) in French, Lee et al. (2020) in French and English). For example, the French discourse marker *alors*, "then, so" will display major differences in the presence/absence of preceding pauses, word duration and pitch reset as a function of whether it conveys consequence (the basic meaning) or topic-shift and specification (less predictable relations) (Didirková et al., 2018).

Our working hypothesis is that, since fine phonetic characteristics are word-specific, and thus will not apply equally to similar phonological forms with different meanings and/or functions (and therefore use), they can be exploited to disambiguate problematic phonological sequences, whether they be homophones

or polysemic words. To test this hypothesis, we take into account a well-known case of ambiguity: discourse markers (henceforth DMs). DMs are words or expressions such as *well*, *you know*, *I mean*, that are highly frequent in language use and have been shown to play an essential role as fluency devices (Crible, 2018) and in discourse planning and communication management (Levelt, 1989; Hasselgren, 2021). They are especially interesting to study because they usually emerge from other parts of speech with which they then co-exist in the language (Degand and Fagard, 2011). For example, Fr. *alors*, "then, so", has been grammaticalized from Lat. *illa hora*, "at this hour", to become a temporal adverb as in *J'avais alors 16 ans*, "I was 16 *at the time*" and is now in parallel used as a DM, for instance to mark a boundary-opening as in *Alors, comment ça va?*, "*So*, how are you?".

Following the cited literature, we aim at observing how DMs are realized in terms of their mean phone duration per word (suggesting local speech rate) and their vowels' pronunciations. We also specify how DM use correlates with speech style, which has been less investigated. To that extent, we use a 4-hour long French corpus that is richly annotated in part-of-speech and other metadata to show whether different grammatical functions of similar phonological forms bring about different phonetic behaviors. In the following, we will mainly compare the behavior of DMs (that we call "DM uses", e.g., opening boundary *alors*) with (i) the behavior of all the other words in the corpus (henceforth "Others", e.g., *table*, "table" or *manger*, "eat"), and (ii) the behavior of the same word-types (e.g., *alors*) not used as DMs (that we call "Non-DM uses"; e.g., temporal adverb *alors*) (Bolly et al., 2014).

In the remainder of this paper, we first describe our corpus (Section Corpus) and methodology (Section Methodology). Results are presented in Section Results regarding the relation between speech style and part-of-speech and DM usage, and the effect of DM usage on mean phone duration and vowel realization by comparing DMs ("DM uses") as opposed to all other words ("Others") and similar word-types not used as DMs ("Non-DM uses").

Corpus

For the present study, we chose to study the 4h long LOCAS-F corpus because of its already fine-grained manual annotations for parts-of-speech and DMs. The LOCAS-F corpus was put together by the linguistic research team at the University of Louvain (Degand et al., 2014b). It is composed of 42 sound tracks of 3 to 5 minutes audio files, including in total 48 speakers. Multiple social practices were included in this corpus of primarily Belgian and metropolitan French, such that several speech styles are represented. This corpus will thus allow us to study DMs with regards to parts-of-speech as a function of phonetic cues as well as stylistic variables.

Methodology

Thirteen large part-of-speech categories were identified manually by specialists, all of which are considered in our analyses. In the section dedicated to part-of-speech, only the nine parts-of-speech that are relevant to DMs are presented (see table 1 for the complete names of the parts of speech). The recordings were categorized into three speech styles according to the degree of preparation: (formal) prepared speech, (less formal) semi-prepared speech and (informal) unprepared speech. More details on the speech styles can be found in Degand et al. (2014a). Concerning the section on vowel space, measurements of the first and second formants were extracted using PRAAT (Boersma, 2006). Given that /o/ and /ɔ/ were manually annotated with similar standards in the LOCAS-F corpus, we decided to group the two phonemes into the same group ("o-ɔ") in our vowel analyses. Vowel spaces are illustrated for discourse markers ("DM uses") *vs* all other words ("Others") and then for discourse markers ("DM uses") *vs* word-types that can fulfill the function of DM but are not employed as such ("Non-DM uses"). Due to limited space, we did not present vowel spaces as a function of speech style. Outliers were mostly observed at the bottom left corner of the vowel spaces and were excluded using a "$y <= -0.5x + 1850$" filter.

Statistical analyses were carried out on whether or not the word-token is a DM for each part-of-speech and speech style using generalized linear models (GLM) in R (R Core Team, 2013). The model was used to test the effect of part-of-speech and speech style. ADJ, PRO and DET were grouped together, given that we observe almost no "DM uses" for these parts of speech. This grouping left us 7 categories for part-of-speech. The fixed effects considered were: part-of-speech (reference: CON) and speech style (reference: prepared speech).

Table 1: Details of part-of-speech used in 1.

Abbreviation	Complete category name
ADJ	Adjective
ADV	Adverb
CON	Coordinating conjunction
DET	Determiner
ITJ	Interjection
NOM	Noun
PRO	Pronoun
PRP	Preposition
VER	Verb

Results

In this section, we investigate the DMs' distribution and phonetic properties with respect to speech style, part-of-speech categories, mean phone duration and vowel space.

Speech style

Table 2 presents the occurrences and percentages of discourse markers ("DM uses") *vs* all other word-tokens ("Others") for each speech style. These results show that discourse markers ("DM uses") are observed more in unprepared speech (7%) than in prepared speech (2%). The more prepared the speech is, the less discourse markers are observed, which is consistent with the literature as more spontaneous settings entail the use of such items for dialog management or planning purposes.

Table 2: Occurrences and rates of DM uses *vs* Others for the three investigated speech styles.

	Prepared	Semi-prepared	Unprepared
DM uses	252 (2%)	440 (5%)	1321 (7%)
Others	12009 (98%)	8270 (95%)	16472 (93%)
Total	12261	8710	17793

Table 3 shows the occurrences and percentages of word-types when they correspond to "DM uses" *vs* "Non-DM uses", for each speech style. Among word-types that can fulfill a DM function, 8% of occurrences are used as discourse markers in prepared speech and more than 20% of the occurrences (24.49%) are used as discourse markers in unprepared speech. These results suggest that words that can be used as DMs are more likely to fulfill this function in less prepared speech.

Table 3: Occurrences and rates of DM uses *vs* Non-DM uses for the three investigated speech styles.

	Prepared	Semi-prepared	Unprepared
DM uses	252 (8%)	440 (18%)	1321 (24%)
Non-DM uses	2729 (92%)	1981 (82%)	4073 (76%)
Total	2981	2421	5394

Part-of-speech and Speech style

Figure 1 illustrates the rate of discourse markers ("DM uses") for each part-of-speech and each speech style, according to the classes specified in Table 1. Similar patterns are found for the three investigated speech styles, however, the less formal the speech style is, the more DMs are observed for each part-of-speech. Categories that are most often used as DMs are Coordinating conjunctions (CON; e.g., *et*, "and"), followed by adverbs (ADV; e.g., *alors*, "so"), interjections (ITJ; e.g., *euh*, "uh") and prepositions (PRP; e.g., *pour*, "for"). Interestingly, pronouns (*moi*, "me") and determiners (DET; *le*, "the" or *mon*, "my") are barely ever used as DMs in our corpus. The GLM results show that it is less likely to observe DMs in any of the other categories than in CON ($p < 0.001$) and that it is more likely to observe DMs in less prepared speech than in more prepared speech ($p < 0.001$ for all pairwise comparison).

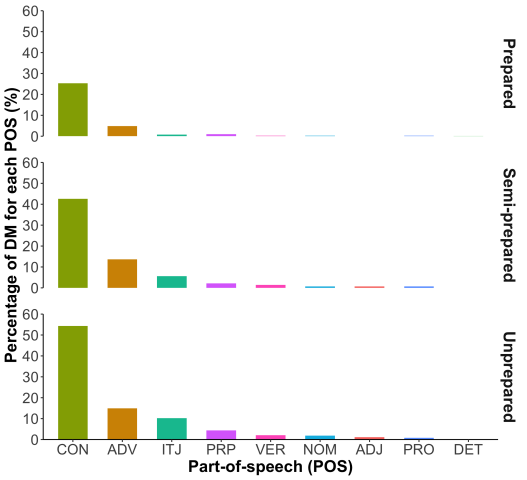

Figure 1: Percentage of discourse markers ("DM uses") for each part-of-speech in LOCAS-F corpus and for each speech style.

Mean phone duration per word

Figure 2 gives an overview of mean phone duration per word for "DM uses" *vs* "Others". The highest bar is located slightly more to the left for "DM uses" than for "Others". This suggests that the local speech rate during the production of DM is slightly higher than for other words.

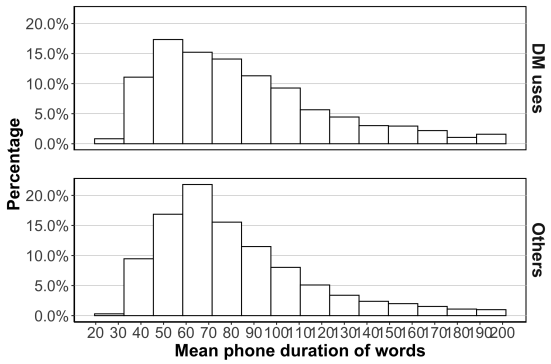

Figure 2: Mean phone duration per word for discourse markers ("DM uses") vs other words ("Others").

A similar tendency is observed for "DM uses" *vs* "Non-DM uses" (Figure 3), suggesting that the local speech rate is higher for "DM uses" than for "Non-DM uses".

While Figure 2 could be affected by the influence of

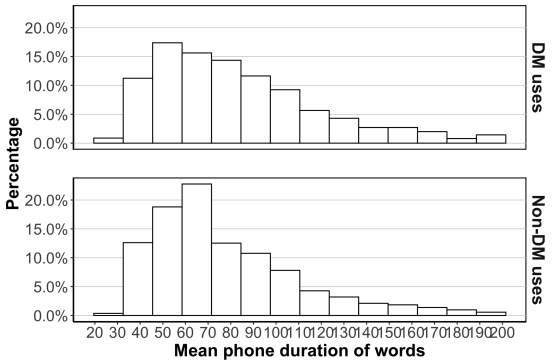

Figure 3: Mean phone duration per word for discourse markers ("DM uses") vs DM words that are not employed as DMs ("Non-DM uses").

word frequency (given that word-types that can correspond to DM realization concern words frequently used in speech and thus produced with higher local speech rate), analyses in Figure 3 which controlled for a frequency related impact, suggest that local speech rate is higher for "DM uses" than for "Non-DM uses".

Vowel space

Figure 4 illustrates the vowel space for DMs ("DM uses", left panel) *vs* other words ("Others", right panel) for female (F, top panel) and male speakers (M, bottom panel). The vowel space for other words is much larger than that for DMs, suggesting that vowels tend to be hypo-articulated for "DM uses". It is worth noting that DMs comprise only limited word-types, and thus limited vowel identities (i.e. /i/, /y/, /e/, /ɛ/, a, /u/, /o-ɔ/) were involved in the vowel space for DMs. Moreover, the larger vowel space for the other words could also be due to the large range of word frequencies for words engaged in this category.

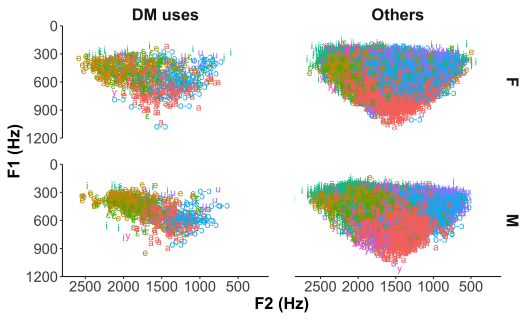

Figure 4: Vowel space for DMs ("DM uses", left panel) and other words ("Others", right panel) for female (F, top panel) and male speakers (M, bottom panel).

In order to have a better understanding of DM vowel space, we ran supplementary analyses on "DM uses" *vs* "Non-DM uses". Figure 5 illustrates the vowel space for DMs ("DM uses", left) and DM word-types that are not employed as DMs ("Non-DM uses", right) for female (F, top) and male speakers (M, bottom). The comparison thus allows us to investigate the same words used as DMs or not, and consequently comparable vocalic realizations. This comparison also allows us to control for word-frequency related variation. Similar to what is demonstrated in Figure 4, the group "DM uses" shows a smaller vowel space, suggesting that the acoustical realization of "DM uses" is more prone to reduction phenomena than in other situations ("Non-DM uses").

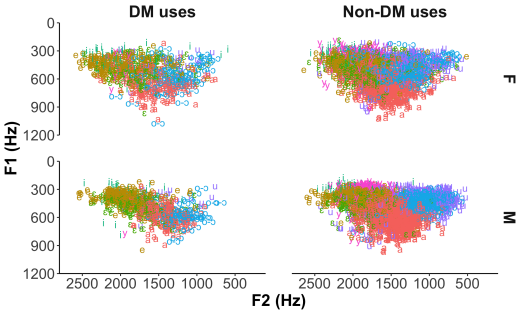

Figure 5: Vowel space for DMs ("DM uses", left panel) and DM words that are not used as DMs ("Non-DM uses", right panel) for female (F, top panel) and male speakers (M, bottom panel).

Conclusions

This study aims to characterize phonetic variation of discourse markers ("DM uses") compared to all other word-tokens ("Others") and to word-types similar to discourse markers yet not used as discourse markers ("Non-DM uses") - thus raising a question of ambiguity - using the 4-hour richly annotated LOCAS-F corpus in French. Different linguistic and stylistic parameters were investigated: speech style, part-of-speech, mean phone duration per word and vowel space. Results show that discourse markers ("DM uses") are observed more in less prepared speech. The part-of-speech that frequently fulfills the function of discourse marker is coordinating conjunction, followed by adverbs, interjections and prepositions. The distribution of mean phone duration per word, suggesting local speech rate, shows that mean phone duration tends to be shorter for "DM uses" than for "Others" and for "Non-DM uses". Vowel space is smaller for "DM uses" than for "Others" and for "Non-DM uses", suggesting that discourse markers undergo hypo-articulation, and thus reduction, compared to other usages. Overall, our results encourage for further investigations of patterns of phonetic variation as cues for disambiguation in connected speech.

Acknowledgements

This research was supported by DATAIA/MSH Paris-Saclay "Excellence" grant and by IdEX U. de Paris, ANR-18-IDEX-0001, "Emergence" grant.

References

Adda-Decker, M. and N. Snoeren (2011). Quantifying temporal speech reduction in french using forced speech alignment. *Journal of Phonetics 39*(3), 261–270.

Boersma, P. (2006). Praat: doing phonetics by computer. *http://www. praat. org/*.

Bolly, C., L. Crible, L. Degand, D. Uygur, et al. (2014). Towards a model for discourse marker annotation in spoken french: From potential to feature-based discourse markers. *Pragmatic Markers, Discourse Markers and Modal Particles: New Perspectives*, 71–98.

Cole, R. A., J. Jakimik, and W. E. Cooper (1978). Perceptibility of phonetic features in fluent speech. *The Journal of the Acoustical Society of America 64*(1), 44–56.

Crible, L. (2018). *Discourse Markers and (Dis) fluency: Forms and functions across languages and registers*, Volume 286. John Benjamins Publishing Company.

Dautriche, I. (2015, September). *Weaving an ambiguous lexicon*. Theses, Université Sorbonne Paris Cité.

Dautriche, I., D. Swingley, and A. Christophe (2015). Learning novel phonological neighbors: Syntactic category matters. *Cognition 143*, 77–86.

Degand, L. and B. Fagard (2011). Alors between discourse and grammar: the role of syntactic position. *Functions of Language 18*(1), 19–56.

Degand, L., L. Martin, and A.-C. Simon (2014a). Locas-f : un corpus oral multigenres annoté.congrès mondial de linguistique française. In *CMLF 2014 - 4 ème Congrès Mondial de Linguistique Française*, Shanghai, China, pp. 2613–2626.

Degand, L., L. J. Martin, and A.-C. Simon (2014b). Unités discursives de base et leur périphérie gauche dans locas-f, un corpus oral multigenres annoté. In *SHS Web of Conferences*, Volume 8, pp. 2613–2625. EDP Sciences.

Didirková, I., G. Christodoulides, L. Crible, and A. C. Simon (2018). Naïve annotations of french et and alors: comparison with experts and effect of implicitation. In *Cross-Linguistic Discourse Annotation: applications and perspectives. TextLink2018–Final Action Conference*, pp. 38.

Didirková, I., L. Crible, and A. C. Simon (2019). Impact of prosody on the perception and interpretation of discourse relations: Studies on "et" and "alors" in spoken french. *Discourse Processes 56*(8), 619–642.

Dinkin, Aaron, J. (2008). The real effect of word frequency on phonetic variation. In *University of Pennsylvania Working Papers in Linguistics*.

Ernestus, M. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics* (39), 253–260.

Hasselgren, A. (2021). Learner corpora and language testing.

Jurafsky, D. and J. Martin (2008, February). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Volume 2.

Lee, L., D. Jouvet, K. Bartkova, Y. Keromnes, and M. Dargnat (2020, October). Correlation between prosody and pragmatics: case study of discourse markers in French and English. In *INTERSPEECH 2020*, Shanghai, China.

Levelt, W. J. M. (1989, January). *Speaking: From Intention to Articulation*. Mit Pr.

Lohmann, A. and E. Conwell (2020, January). Phonetic effects of grammatical category: How category-specific prosodic phrasing and lexical frequency impact the duration of nouns and verbs. *Journal of Phonetics 78*.

Navigli, R. (2009, February). Word sense disambiguation: A survey. *ACM Comput. Surv. 41*(2).

Phillips, B. (1984). Word Frequency and the actuation of sound change. *Language* (60), 320–342.

Pierrehumbert, J. B. (2008). *Word-specific phonetics*, pp. 101–140. De Gruyter Mouton.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Schütze, H. (1998). Automatic word sense discrimination. *Computational Linguistics 24*(1), 97–123.

Singh, V. and K. Saraswat, Krishan (2019, June). Techniques for Disambiguation of Polysemy Words: A Review. *International Journal of Computer Applications 178*, 6–10.

Swingley, D. and R. N. Aslin (2007). Lexical competition in young children's word learning. *Cognitive Psychology 54*(2), 99–132.