

The Impact of Prosody on the Perception and Interpretation of Discourse Relations: Studies on “Et” and “Alors” in Spoken French

Ivana Didirková^{a*}, Ludivine Crible^a and Anne Catherine Simon^a

^aCentre VALIBEL, Institute for Language & Communication, Université catholique de Louvain, Place Blaise Pascal 1, 1348 Louvain-la-Neuve, Belgium

ivana.didirkova@gmail.com, tel.: 0033 6 34 15 75 49, fax: 00 32 10 47 25 79

Running head:

Prosody and the Perception of Discourse Relations

The Impact of Prosody on the Perception and the Interpretation of Discourse Relations: Studies on “Et” and “Alors” in Spoken French

We report on three experiments which aim at measuring the role of prosody in the acceptability and interpretation of discourse relations between utterances connected by two French discourse markers, viz. *et* ‘and’ and *alors* ‘then/well’. These two discourse markers are highly polyfunctional: *et* ‘and’ can express relations of *addition*, *specification*, *temporality* or *consequence*, while *alors* ‘then/well’ can express relations of *consequence*, *specification*, *concession* or *topic-shift*, among others. The first experiment compares the naturalness of the stimuli in their spoken and written versions. The second study uses a forced-choice task where two prosodic versions of the same stimulus are confronted. The third experiment tests whether the prosody of the first segment and of the marker projects the discourse relation continuation. Our results show that the impact of prosody varies across markers and relations, with *alors* displaying stronger prosodic profiles associated to the discourse relations it introduces.

Keywords: discourse markers, discourse relations, prosody, perception

1. Introduction

Natural language, either written or spoken, is built upon coherence links that connect discourse units through relations of causality, temporality or contrast, among many others (Mann & Thompson, 1988). These relations are often signalled by connectives or “discourse markers” (henceforth DMs), which are here defined as optional procedural operators guiding the interpretation processes. Some of the most frequent DMs in spoken French are the conjunctions *et* ‘and’ or *mais* ‘but’, the adverbials *donc* ‘so’ or *alors* ‘then, well’ as well as interjections such as *hein* ‘right?’ or *ben* ‘well’. The formal variety of DMs is only rivalled by their wide functional spectrum: DMs cover different semantic classes of discourse relations (*consequence*, *specification*, *concession*, etc.) as well as discourse-structuring and interpersonal uses such as topic change, turn-taking or monitoring (e.g. Fischer, 2006).

DMs present many interesting challenges for the analyst, as attested by the proliferating literature on the topic. Chief among them, the absence of systematic form-meaning mapping or,

in other words, the great polyfunctionality of DMs, makes their analysis – and, we argue, their online interpretation – more difficult (Degand, Cornillie, & Pietrandrea, 2013). This polyfunctionality is furthermore supported by the contribution of many other cues besides the DMs themselves to the interpretation of discourse relations. These contextual cues include syntactic constructions (e.g. negation), lexical patterns (e.g. antonyms, adverbials) and, for speech, prosodic configurations (Fabricius-Hansen, 2005; Taboada, 2006; Ouden, Noordman, & Terken, 2009; Asr & Demberg, 2015). In this study, we focus on the latter and investigate the impact of prosody on the perception of various discourse relations expressed by two highly polyfunctional DMs in spoken French: the additive conjunction *et* ‘and’ and the temporal-argumentative adverbial *alors* ‘then, well’.

We present three perception experiments which aim at measuring the effect of prosody on acceptability, preference and prediction of discourse relations marked by polyfunctional DMs. Our general hypothesis is that prosodic cues, such as pause duration before the DM, pitch reset, syllable lengthening or intonation contour, can help disambiguate between the different pragmatic uses of DMs in general and of French *et* and *alors* in particular. This project stems from the corpus-based observation that some discourse relations exhibit specific prosodic profiles (Murray, Renals, Carletta, & Moore, 2006; Kleinhans, Farrús, Gravano, Pérez, Lai, & Wanner, 2017).

The perception experiments reported in this paper are also the direct continuation of a larger research project involving annotation and production experiments with both written and spoken materials (see Didirková, Christodoulides, Crible, & Simon, 2018; Didirková, Christodoulides, & Simon, 2018). Our goals for the present study are three-fold: i) to compare acceptability scores for the same discourse stimuli in written vs. spoken form (Study 1); ii) to assess the perception of cooperating vs. conflicting prosodic profiles and their role in the construal of discourse relations (Study 2); iii) to test whether the prosody of the DM and its

preceding context projects or helps the listener predict the upcoming discourse relation (Study 3). In doing so, this paper contributes to the relatively scarce experimental literature on spoken DMs and prosody.

In the next section, we will review previous studies on DMs, discourse relations, prosody and perception (Section 2). We will then present the materials and methods (Section 3). Each study will then be developed (Sections 4, 5 and 6) before we discuss the results and conclude (Section 7).

2. Background

2.1 *The French discourse markers “et” and “alors”*

Our study focuses on two DMs which share a very high frequency and polyfunctionality in spoken French. In the spoken dataset DisFrEn (Crible, 2017), which contains 75,288 words from various dialogue and monologue settings in French, *et* ‘and’ is the most frequent DM (11.54 tokens per thousand words), while *alors* ranks fourth (3.6 per thousand words).¹ Previous research has identified a number of functions for these two DMs, which display different semantic profiles.

The conjunction *et* is primarily seen as a logical connector, with a basic semantics of *addition*: the two segments related by *et* are merely conjoined. However, most authors acknowledge that *et* can express more than *addition* in specific contexts. Luscher & Moeschler (1990) evoke *temporal*, *causal*, *narrative*, *implicative*, *counterfactual* and *oppositive* uses. Saussure & Sthioul (2002) explain this high variation by distinguishing the semantics of the conjunction on the one hand and pragmatic effects derived and enriched by the context on the

¹ In the French subcorpus of DisFrEn, 4,494 DM tokens were annotated. The five most frequent DM types are: *et* ‘and’, *mais* ‘but’, *donc* ‘so’, *alors* ‘then/well’ and *hein* ‘right’.

other. Thus, while *et* itself is always additive, it can be used as an underspecified marker of various discourse relations (cf. Spooren, 1997). Previous studies on its English equivalent *and* confirm this polyfunctionality: *and* has been described as forcing a temporal, chronological interpretation of the related segments, when a causal relation could have been inferred without *and* (Blakemore & Carston, 1999). The authors also show that *and* can encode emotions, especially in concessive contexts such as “Her husband is in hospital and she is seeing other men” (Blakemore & Carston, 1999, p. 13). In sum, *et* (and its crosslinguistic equivalents) is highly polyfunctional, although these contextual enrichments seem more frequent and more varied in speech than in writing (Crible & Cuenca, 2017).

The DM *alors* originates from a temporal anaphoric adverbial (Latin *illa hora* ‘at that time’) which has evolved into three main uses: *temporal*, *causal/argumentative*, *discourse-structuring/pragmatic* (Gerecht, 1987; Degand & Fagard, 2011). It is frequently found in conditional contexts, where it is equivalent to *then* in the *if... then* construction. From there, it developed a resultative meaning, although authors agree that *alors* always retains its original temporal meaning. The derived discourse-structuring or pragmatic uses of *alors* correspond to topic organization: introducing a new topic, opening a digression, coming back to a previous topic. Hansen (1997) adds that these discourse-structuring uses can include a nuance of *concession*. She also mentions that *alors* can be used to ask for a continuation when it is located in final position with a rising intonation. Even though such formal-functional mappings have been observed (see also Degand & Fagard, 2011, for an analysis of syntactic positions), *alors* “frequently seems to cumulate more than one function” (Hansen, 1997, p. 184), making it challenging to analyse and possibly to interpret.

In DisFrEn (Crible, 2017), some of the most frequent senses assigned to *et* are *addition* (basic conjunction of two segments; 498/869 occurrences), *specification* (the first segment is explained in more details; 78/869), *consequence* (the second segment is the logical result of the

first one; 44/869) and *temporality* (the two segments are ordered chronologically; 24/869), while *alors* was annotated as *consequence* (62/271), *specification* (42/271), *topic change* (the second segment introduces a new topic; 12/271) and *concession* (the second segment contradicts or qualifies the first one; 4/271). It can be noted that both *et* and *alors* frequently express *specification* and *consequence*. This not only points to the partial overlap between the functional spectrums of the two DMs, but it also shows the conceptual proximity of some of these meanings, as will be developed in the next section.

In sum, *et* and *alors* share part of their functional spectrum but are not synonymous, as shown by the fact that they frequently co-occur as a string *et alors*, which points to their complementarity rather than to their similarity (see Crible, 2018, pp. 120-121).

2.2 Interpreting discourse relations

Zooming out from *et* and *alors*, discourse relations in general are challenging to identify and DMs often belong to more than one functional category (either simultaneously or not; Petukhova & Bunt, 2009). In fact, the absence of consensus in the literature regarding the number and types of discourse relations and how they map with DMs in various languages is telling in itself of the complexity of the phenomenon. Let us take one of the most widespread frameworks on discourse relations, the Penn Discourse Treebank 2.0 (henceforth PDTB), to illustrate this issue. In the PDTB 2.0 (Prasad, Miltasakaki, Dines, Lee, & Joshi, 2007), relations are grouped into four semantic classes: COMPARISON, CONTINGENCY, EXPANSION and TEMPORALITY. Each class contains several relations, which are conceptually close. For instance, all relations in the CONTINGENCY class share some causality effect (subtypes of causal and conditional conditions). However, the relation of *concession* is grouped in the COMPARISON class, together with *contrast*, even though *concession* involves a causal inference as well. Another framework, the Cognitive approach to Coherence Relations

(henceforth CCR, Sanders, Spooren, & Noordman, 1992) would rather describe *contrast* as additive-negative and *concession* as causal-negative, thus pointing to what is common and what is different between the two relations.

Annotation models, such as the PDTB 2.0 and many others, tend to suggest that the categories are clear-cut and orthogonal. Yet, inter-annotator agreement scores show rather the opposite situation, and the reliability is often low for this type of study (Spooren & Degand, 2010). If we use the CCR framework to define the discourse relations expressed by *et* and *alors*, we see that some of them do not differ much: both *addition*, *temporality* and *specification* would be classified as additive-positive, *consequence* would be causal-positive while *concession* is causal-negative (see Sanders, Spooren, & Noordman, 1992 for more information on CCR).²

Besides linguistic categorization, online interpretation can also be affected by the ambiguity of the discourse relation and its marker. Cain & Nash (2011) have shown that young readers take a longer time to read sentences connected by *and* than by more specific markers like *because* or *although* (see also Murray, 1997). Zufferey & Gygax (2017) showed that German-speaking learners of French struggle with the *causal* or *confirmation* DM *en effet* ('because'/'indeed') in online and offline tasks. Nevertheless, listeners seem to be able to make the right predictions even on the basis of ambiguous or underspecified DMs, such as *a* in Russian which can either mark *contrast* ('but') or *addition* ('and'), as described in Mak, Tribushinina, & Andreiushina (2013).

How is this paradoxical situation possible? We can reconcile ambiguity with interpretation if we consider that listeners resort to other cues besides the DMs in order to interpret the discourse relation that holds between utterances. Previous research has studied many different types of such contextual cues, such as the semantics of the verb (Pyykkönen &

² Topic change is not considered as a discourse relation in the CCR framework and cannot be described in terms of these binary features.

Järvikivi, 2010), the discourse relation in the previous context (Rohde & Horton, 2014) or negation markers (Asr, & Demberg, 2015). Targeting *and* specifically, Rohde, Tyler, & Carlson (2017) mention that the conjunction can signal *additive*, *temporal*, *causal* or *contrastive* relations, in which case “the content of the conjoined sentences is crucial for inferring the speaker’s intended relation, and this inferencing depends on contextual cues and comprehenders’ real-world knowledge” (2017, p. 5). These contextual cues are the basis for automatic discourse relation recognition as carried out by natural language processing studies (see e.g. Pitler et al. 2009).

Prosody is no exception and has been repeatedly shown to intervene in the construal of discourse relations. Murray et al. (2006) suggest that prosody is a sufficient discriminating feature to distinguish between rhetorical relations, on the basis of various parameters such as pitch contour, speech rate or duration of silent frames. Kleinhans et al. (2017) qualify these results and show that only some relations can be discriminated through prosodic features such as fundamental frequency (f_0), intensity, inter- and intra-segmental features. In the same vein, Ouden et al. (2009) showed that causally related utterances are read aloud with shorter pauses between them and at a faster rate than non-causally related utterances.

The present study directly builds upon a previous production experiment in French (Didirková et al., 2018), which showed that utterances introduced by *alors* are more prosodically detached (in terms of preceding pause duration and syllable lengthening) from the previous utterance than utterances introduced by *et*. Moreover, within the relations marked by *alors*, the statistical analysis showed large discrepancies between *consequence* (the basic meaning) on the one hand and *topic-shift* and *specification* on the other (less predictable relations): in the former, no preceding pause or DM lengthening were observed and the pitch reset was present to a lesser extent.

2.3 Previous perception studies on the prosody of syntactic structures and discourse relations

A number of perception studies were carried out, aiming at identifying the influence of prosody on utterance comprehension. However, most of these studies used syntax as reference, trying to see whether intonation can guide the interlocutor in disambiguating two or more ambiguous structures. For instance, the contribution of intonation on disambiguating early from late closure has been examined in several studies (e.g. Slowiaczek, 1981; Speer, Kjelgaard, & Dobroth, 1996; Schafer, Speer, Warren, & White, 2000). However, the results seem to differ with respect to the stimuli (“natural” vs. read) and the task presented to participants. Thus, asking participants to evaluate their own comprehension of an auditory stimulus based on read speech leads to the conclusion that late closure would enhance comprehension, with or without cooperating prosody (Slowiaczek, 1981). On the other hand, when presented with “natural” stimuli produced within a game and asked to choose between two possible continuations displayed on a computer screen, the percentages of correct categorizations were higher for items with cooperating prosody as compared to the items with conflicting prosody (Schafer et al., 2000). Price, Ostendorf, Shattuck-Hufnagel, & Fong (1991) used the same experimental method (i.e. choosing between two possible continuations after listening to auditory stimuli) to verify the influence of prosodic cues on disambiguating different types of syntactic structures, such as parenthetical / non-parenthetical subordinate clauses. Their results are in line with those obtained by Schafer et al. (2000), in that naïve listeners are able to distinguish between the ambiguous structures on the basis of prosody.

Other perception studies investigated the role of intonation in interpreting syntactic relations. In French for example, Michelas & D’Imperio (2015) tried to determine whether low-level boundaries (accentual phrase and intermediate phrase) influence syntactic parsing

decisions. Their results suggest an effect of boundary strength on the decisions of participants, confirming the observations presented in Michelas (2011).

This being said, studies interested in the role of intonation in discourse interpretation remain rather scarce, even though some prosodic cues seem to enhance the identification of discourse structure (Mayer, Jasinskaja, & Kölsch, 2006). This observation led the authors to investigate the influence of pauses and pitch range on discourse structure comprehension. More specifically, Mayer et al. (2006) constructed 28 discourse sequences with two possible interpretations in terms of low / high attachment. They then asked one female speaker to read all the sentences in randomized order and adjusted pause duration and pitch range parameters post-hoc. The two parameters were manipulated either together or one at a time, in order to test their complementarity. Participants were then asked to listen to the stories and, after each of them, answer orally a question allowing the experimenter to classify the interpretation as high / low attachment or indefinite. Authors concluded that, when combined, both pause duration (longer in high attachment) and pitch range (expanded in high attachment) do help with disambiguating structurally ambiguous discourses. In another perception study on the prosody of discourse structure, Tyler (2014) also showed that participants use prosody to interpret hierarchical ambiguities in the absence of lexical cues. Namely, participants seemed to use terminal pitch, pause duration, and overall pitch and intensity contrast.

As for the prosody of discourse markers, Gravano, Benus, Chávez, Hirschberg, & Wilcox (2007) have used a perception study to investigate acoustic, prosodic and phonetic cues influencing the interpretation of *okay*. They have selected 12 spontaneous conversations from the Columbia Games Corpus in which they have identified all occurrences of the DM *okay*. These instances were then annotated for functional categories (*backchannel*, *agreement*, *cue beginning*, *pivot beginning*, other). Three of them were chosen for the study: *backchannel*, *cue beginning* and *acknowledgment/agreement* and every class was combined with three levels of

inter-annotator agreement (unanimous, majority, no-agreement), resulting in nine categories of the target word. Moreover, two versions of each token were prepared for the perception study: contextualized (with two full speaker turns for each token) and isolated. 20 participants were then asked to listen to 54 isolated tokens and 54 contextualized tokens and select one of the three functional categories (multiple-choice task). Their results show that, while participants rely more strongly on their default interpretation than on intonation, especially when the DM is isolated, the word final intonation helps the interpretation of the target word. Furthermore, the inter-annotator agreement was higher in contextualized version of the test, suggesting that the context plays an important role in such tasks. Although this study targets a different kind of marker, namely the interpersonal marker *okay*, their methodology and results are relevant for the present purposes.

The purpose of the three experiments presented below is to address the lack of perception studies trying to highlight the role of prosodic cues in interpreting discourse relations introduced by discourse markers. In other words, we intend to link the two elements, both the discourse relations and the DMs and their oral realization in order to determine whether prosody can influence the interpretation of short utterances.

3. Methods: Stimuli preparation

In this section, we present the method for preparing recorded stimuli for each of the eight investigated marker-relation pairs. Materials, procedure and participants for each study are presented in sections 4, 5 and 6.

The stimuli used in the three experiments originate from previous preliminary studies led within the framework of the same research project. Using the procedure reported in Didirková, Christodoulides, & Simon (2018), we have first pre-selected instances of the target words *alors* and *et* in the LOCAS-F corpus (Degand, Martin, & Simon, 2014). The target DMs

were embedded in utterances schematized as follows: S1+DM+S2, where S1 + DM and some of the S2s were extracted from natural spoken data. Every utterance was annotated for its discourse relation by two of the authors. We have chosen to investigate six discourse relations as presented in Insert Table 1 here.

: *concession* and *topic-shift* for *alors*, *addition* and *temporality* for *et* and *consequence* and *specification* for both DMs.

Insert Table 1 here.

The remaining S2s were constructed in order to represent the three other discourse relations for the corresponding S1+DM.

Consider the example in Table 2:

Insert Table 2 here.

This procedure gave rise to 176 utterances submitted to annotation by naïve subjects (university students with no prior training, see Didirková, Simon, & Christodoulides, 2017) and, on the basis of this annotation, discourse relations that most frequently overlapped were paired as follows: for *alors*, the *concession* relation (henceforth CCS) was paired with *specification* (SPE) and *topic-shift* (TOS) with *consequence* (CSQ), as the relations within these two pairs were most often confused with each other in the naïve annotations. For *et*, *addition* (ADD) was paired with *temporality* (TMP) and *consequence* with *specification*. Among these pairs, 64 sequences were selected, representing 32 sequences per DM and 8 per relation for each DM. Based on previous annotation experiments, we decided to keep, for every discourse relation, 2 utterances with a high inter-annotator agreement ($\geq 72.72\%$), 2 with a small agreement / without agreement (e.g. when 5 annotators chose the *consequence* relation while 6 others chose the *topic-shift*) and 4 others where the agreement was not controlled. This choice was made in order to explore the extent to which prosody can contribute to the disambiguation

of unclear discourse relations (i.e. to compare the results of our perception tests with respect to inter-annotator agreement).

These stimuli were then read by 20 naïve participants, which led to a speech production study. This production study permitted to reveal a number of differences in prosodic profiles for each discourse marker in each discourse relation (for detailed acoustic measurements, see Didirková et al., 2018). Based on acoustic parameters³, we grouped the observed prosodic profiles into three levels of prosodic integration between S1 and the DM introducing S2 (Table 3). When the DM is prosodically integrated (e.g. *et*-addition), it means that there is a short silent pause before the DM (about 125 ms) and that the pitch reset between the end of S1 and the DM is compressed (less than 1 semi-tone or ST⁴). The so-called “prosodically integrated profile” produces a smooth transition between the first utterance and the utterance introduced by the DM. On the other hand, when the DM is prosodically segmented (or “detached”), it means that there is a long silent pause before the DM (between 500 and 750 ms) and that the pitch reset between the end of S1 and the DM is expanded (3 semi-tones or more). The so-called

³ The production study resulted in 1,280 sequences (S1+DM+S2) produced by 20 naïve speakers. We opted for an acoustic analysis instead of choosing a standardised transcription system like French ToBI (Delais-Roussarie et al., 2015) or Mertens’ transcription system for French intonation and accentuation (Mertens, 2014). The reason is that we wanted to keep the option of producing synthesized stimuli for the perception studies, based on acoustic parameters. Another reason is that we learned from our own experience (Mertens & Simon 2013) that it is difficult to reach a high inter-rater agreement when prosodic transcription is concerned.

⁴ Values for pitch measurements are in semi-tones (ST) instead of Hz, since this measure is considered as being better adapted to the human auditory system (see e.g. Nolan, 2003) and to the comparison between male and female voices. These values were automatically extracted from Praat (Boersma & Weenink 2017) using the “semi-tones relatives to 1 Hz” scale, and then verified manually.

“prosodically detached (or segmented) profile” produces a perceptible interruption between the first utterance and the second utterance introduced by the DM. Finally, several DMs have a “in-between”, semi-detached prosodic profile.

Insert Table 3 here.

However, these differences were not systematically large for every utterance and every naïve speaker recorded in the production study. This led us to discard the option of using the naïve speakers’ productions for the perception studies. We also discarded the possibility of creating the stimuli by means of vocal synthesis, because of its lack of naturalness.

The decision was made to re-record all 64 utterances by an expert, female native speaker of French, in line with other perception studies (see Tyler & Warren, 1987; Tyler, 2014; Michelas, 2011; Michelas & D’Imperio, 2015). Recordings took place in a quiet room using a Thomann Shure WH20XLR cardioid headset microphone connected to an Edirol/Roland R09-HR recorder. The stimuli recorded by the expert respect the parameters produced by the naïve participants with a slight exaggeration. For instance, differences in pause duration were increased for each pair of relations so that they would be perceivable by the experimental subjects. Thus, while the difference in mean duration of pauses preceding the DM was of 75 ms for the pair CSQ – TOS (*alors*) in naïve readings, this difference has been increased to 250 ms in the expert reading. It should be noted that the pause duration parameter has been post-treated after the recordings, while pitch reset as well as syllable durations and the prosodic contour were the expert’s original ones (e.g. not manipulated after recordings). The final prosodic parameters for the four paired combinations correspond to the targeted prosodic profiles. They are presented in Table 4.

Insert Table 4 here.

Post-hoc perceptive and acoustic analyses were carried out to verify that the expected prosodic profiles had been correctly produced⁵. Firstly, we listened to each pair of stimuli in order to make sure that a clear difference between the two contrasted prosodic profiles was perceptible. Secondly, we extracted acoustic measures of duration and relative f0. Prosodic profiles for *et* are either prosodically integrated (ADD, CSQ) or semi-detached (TMP, SPE): although the difference in the DM duration is small⁶ (see Figure 1), there is a clear difference in pause duration before the DM (see Table 4) and a difference in f0 between the preceding syllable and the DM (see Figure 2): the *temporal* profile contrasts with the *consequence* profile by being melodically reset, while the *specification* profile adopts a f0 downstep that contrasts with the *addition* profile. As for *alors*, prosodic profiles are either semi-detached (CSQ, CCS) or detached (SPE, TOS). The shortest duration of the DM corresponds to *consequence*, while the longest duration characterizes the *topic change* relation (see Figure 1). As for f0 resetting, *alors* is unmarked for *consequence* and, to a lesser extent, for *concession*, while there is an intermediate melodic reset for *specification* and a strong melodic reset for *topic-shift* (see Figure 2).

Figure 1 : Mean duration of the DM (in ms) per relation. [Insert Figure 1 here.](#)

Figure 2 : Pitch reset (in ST) between the last syllable of the S1 and the (first) syllable of the DM, per relation. [Insert Figure 2 here.](#)

⁵ A few stimuli for which it was not the case were re-recorded.

⁶ The French discourse marker *et* [e] is monosyllabic and contains no consonants. It was unnatural to pronounce it with an exaggerated lengthening, so that we preferred to modify the f0 range and the duration of the pause before the discourse marker, in order to keep the stimuli as natural-sounding as possible.

Perceptive and acoustic analyses confirmed that originally recorded sound files displayed adequate prosodic profiles, and that a contrastive prosodic profile was provided for each discourse marker and each discourse relation.

4. Study 1: Acceptability

4.1 Materials, procedure and participants

In the first experiment, we used the 64 utterances recorded as explained in the previous section. The stimuli were presented in a random order (not classified by marker or relation) to the participants and could be played multiple times on LimeSurvey. The instructions specified that the utterances were connected by *alors* or by *et*, in order to draw the participants' attention to these target elements.

19 participants (10 female, 9 male; aged 22 on average), recruited via a Participants' Pool of the Faculty of Psychology of the Université catholique de Louvain (on Facebook), completed the experiment. They had to answer an acceptability question ("Indicate your degree of acceptability") through a five-point Likert scale. Subjects were remunerated for their participation.

A similar experiment had first been run with the same stimuli presented in written form (Didirková et al., 2017). The experimental conditions and the instructions were exactly the same, except that the participants read the stimuli on the screen. Different participants were recruited for each of these experiments. This pilot experiment showed good acceptability for all relations marked by *et*. On the contrary, only *consequence* achieved substantial acceptability for *alors*, while the other three relations (*topic change*, *specification* and *concession*) were never judged as acceptable (i.e. never obtained the scores of 4 or 5). Detailed results for both the written and spoken stimuli will be discussed in Section 4.3.

4.2 Hypotheses

Based on the acceptability scores of the written stimuli, we expect that the spoken modality and appropriate prosody will improve the acceptability of discourse relations marked by *alors*, which were judged highly unacceptable in their written form. In particular, we expect that the *topic change* use of *alors* is specific to spoken dialogue and will thus be better perceived in the spoken modality. Our overall hypothesis is that this modality will generally improve the acceptability of both markers and all relations, given that some stimuli judged as unacceptable were extracted from authentic spoken corpus data.

4.3 Results

Acceptability scores are higher for *et* than for *alors*, with an average score of 4.34 (out of 5) for the former (standard deviation 0.97), as opposed to 2.88 for *alors* on average with more internal variation (1.46 SD). This result was already observed for the written stimuli. All four relations marked by *et* have very similar scores, whereas only *consequence* was ranked as “acceptable” or “very acceptable” in more than 50% for *alors*. *Topic change* presents the lowest score, with less than 30% of high acceptability.

Figures 3 and 4 show the comparison between the acceptability of the stimuli in their written and spoken form. It appears that there is a clear increase of acceptability for both markers and for each relation in the spoken condition. In particular, the largest improvement concerns *alors*, although the scores remain rather low. Against our hypothesis, the *topic change* use of *alors* does not show an increase in acceptability as large as we expected. We also note that not all utterances marked by *et* are judged acceptable. This is surprising since the stimuli were either extracted from original data or constructed on the basis of original data.

For both markers, in the spoken condition, the discourse relation with the highest acceptability score corresponds to the basic meaning of the marker, viz. *addition* for *et* and

consequence for *alors*, which points to the prevalence of semantics not only for the participants' interpretation but also for their perception of spoken discourse. This slightly contrasts with the written condition, where *consequence* was ranked highest for both markers. We will come back to this preference for causal (consequential) interpretations when we discuss Study 3 (Section 6).

Figure 3: Acceptability scores by marker and relation (written) [Insert Figure 3 here.](#)

Figure 4: Acceptability scores by marker and relation (spoken) [Insert Figure 4 here.](#)

5. Study 2: Forced choice

5.1 Materials, procedure and participants

In the second study, we re-used the same 64 spoken utterances as in the previous experiment. This time, each utterance was prepared in two prosodic versions, which we refer to as:

- “cooperating” prosody, i.e. with the prosodic profile determined on the basis of the production study, without any further manipulation, cf. Table 3;
- “conflicting” prosody, i.e. with the prosodic profile of another relation, as explained in Section 3 (for *et*, *addition* contrasted with *temporality*, and *specification* with *consequence*; for *alors*, *consequence* contrasted with *topic change*, and *specification* with *concession*).

This terminology, inspired by Schafer et al. (2000), allows us to test whether or not each discourse relation and marker has an associated prosodic profile. In order to reproduce the parameters of the other discourse relation in the pair, we have constructed the “conflicting” versions by copy-pasting parts of recorded stimuli expressing different relations, which does not alter their naturalness. More specifically, we combined the second segment (S2) of the target

relation with the first segment (S1) and discourse marker of the contrasted relation. The content of the segments always remains the same, only the prosodic parameters differ. Consider the examples below:

- (1) [et ce qui est étonnant chez Proust c'est l'analyse de la vie sociale]_{S1} [**alors**]_{DM} [par moments moi il me gonfle comme tous les écrivains]_{S2-CCS}
[and what is surprising about Proust is the analysis of social life]_{S1} [**alors**]_{DM}
[sometimes he annoys me like all writers]_{S2-CCS}
- (2) [et ce qui est étonnant chez Proust c'est l'analyse de la vie sociale]_{S1} [**alors**]_{DM} [dans ses analyses il parle principalement de la haute bourgeoisie parisienne]_{S2-SPE}
[and what is surprising about Proust is the analysis of social life]_{S1} [**alors**]_{DM} [in his analyses he mainly talks about the Parisian wealthy bourgeoisie]_{S2-SPE}

We can see that the same S1 was used as the first segment in relations expressing *concession* (Example 1) and *specification* (Example 2) with different S2s. To construct the conflicting prosodic version of Example 1, we combined the S2 from Example 1 (concession relation) with the S1 and DM from Example 2 (recorded with the prosodic parameters of *specification*).

The cooperating and conflicting stimuli were presented in a random order. The participants were asked to listen to each version of the same utterance and select the one they preferred (binary choice). The stimuli could be played multiple times (Gravano et al., 2007).

Participants were then asked to answer a comprehension question. This verification was necessary given that the manipulation of prosody could re-orient the interpretation of the whole utterance. Therefore, we measured not only the participants' prosodic preference, but also their correct interpretation of the intended discourse relation. This comprehension check resorted to

simple metalinguistic paraphrases, to which the participants had to answer “true” or “false”. The (translated) paraphrases for each discourse relation are presented in Table 5.

Insert Table 5 here.

These paraphrases were specifically designed to avoid confusion with the other relation with which they compete (e.g. *addition* vs. *temporality*) as much as possible.

Two groups of 30 participants were recruited via a Participants’ Pool of the Faculty of Psychology of the Université catholique de Louvain (on Facebook), (46 female, 14 male; aged 21.5 on average). They were remunerated for their participation.

5.2 Hypotheses

Our overall hypothesis suggests that the cooperating prosodic profile (and not the conflicting version) will always be preferred by the participants, assuming that each discourse relation is associated with a given prosody in people’s representations. This hypothesis is expected to affect more particularly the *alors* stimuli, since the discourse relations marked by this DM showed clearly contrasted prosodic profiles (see Tables 3 and 4). In addition, we expect a facilitating effect of prosody on discourse interpretation such that preference for the cooperating prosodic profile should lead to correct identifications of the discourse relation.

5.3 Results

The results of the second experiment are presented in Figure 5. We distinguish between different groups according to the participants’ answers. First, participants mostly selected the cooperating prosodic profile for *alors* expressing *topic-shift*. *Concession* and *consequence* show an average of about 70% of expected answers (i.e. the cooperating prosody is selected), while *specification* (expressed by *alors*) presents a higher variation and a lower average of expected answers (about 55%).

Figure 5: Proportion of answers selecting the cooperating prosodic profile by relation and marker Insert Figure 5 here.

The picture is neater for *et*, where we can clearly see two categories: on the one hand, the cooperating prosody is mostly preferred for *addition* and *consequence*; on the other, less than half of the expected answers were provided for *specification* and *temporality*. This divide corresponds to a difference in the prosodic profiles: the unmarked (so-called *integrated*) prosody of *addition* and *consequence* is much better perceived than the stronger contrasts (expanded pitch range on the DM and longer pause before the DM) in *temporality* and *specification* relations expressed by *et*. This suggests that there might be only one acceptable prosody for *et*, namely the unmarked integrated profile, regardless of the discourse relation it expresses. In that sense, prosody reflects semantics, since *et* is a monosemous additive conjunction: *et* is associated with only one semantically encoded meaning and with one prosodic profile, to which context and the content of the related segments can add pragmatic inferences.

Turning to the two discourse relations which can be expressed by both *et* and *alors*, namely *consequence* and *specification*, we see that they behave in a similar way regardless of the marker: *consequence* shows a high proportion of cooperating prosody in both cases, whereas *specification* presents the lowest proportion (lowest average, high variation) in both cases, closely followed by *temporality* in the case of *et*. This result is confirmed by a mixed effects logistic regression model, taking as fixed effects the marker, the answer to the comprehension question, the ambiguity of the utterance⁷ and the type of discourse relation, and as random

⁷ This ambiguity is measured as the proportion of inter-annotator agreement computed for each utterance during the disambiguation task, carried out with the same stimuli but in their written form (cf. Section 3, see Didirková et al. 2017).

effects the participants and the stimuli⁸. This model, with a conditional R square of 0.145, only shows three significant main effects: overall, correct answers to the comprehension question are significantly associated with a preference for the cooperating prosodic profile ($p < 0.05$); compared to the basic meaning of each marker (*addition* for *et* and *consequence* for *alors*) as reference levels, *specification* and *temporality* decrease the chances of participants selecting the cooperating prosodic profile ($p < 0.001$). This model therefore suggests that the *temporal* and *specification* relations are not clearly associated with a given prosodic profile in people's representations, which qualifies our first hypothesis. Although we established their prosodic profiles on the basis of the production study, it could be the case that they are not stable enough to be shared in the linguistic community. Furthermore, as it has been explained in Section 3, the observed prosodic parameters were somewhat exaggerated when the expert recorded the stimuli, in order to ensure perceptible differences between stimuli. The results may suggest that *et* cannot accommodate an over-marked prosodic profile.

Once incorrect answers to the comprehension question are filtered out of this model, the same main effects remain, with the addition of a positive significant effect ($p < 0.01$) of *topic-shift* on the preference for the cooperating profile (compared to the basic meaning *consequence*), which confirms that this use of *alors* is strongly associated with its specific, cooperating prosody (cf. Figure 5). The filtered model (Model 1bis, Appendix) also suggests that there is no difference between *addition* and *consequence* for the additive marker *et*, which signals that both the basic meaning and its causal enrichment favour the selection of the cooperating prosody.

⁸ See Model 1 (Appendix) for further details.

The significant effect of correct comprehension answers on prosodic preference leads us to investigate in further details the different possible combinations between prosody and comprehension. Four cases can be identified:

- the cooperating prosody is preferred and the comprehension is correct (“helping” prosody);
- the cooperating prosody is preferred but the comprehension is incorrect (“non helping” prosody);
- the conflicting prosody is preferred but the comprehension is correct (“ignored” prosody);
- the conflicting prosody is preferred and the comprehension is incorrect (“overriding” prosody).

The proportion of each of these cases by discourse relation and marker is presented in Figure 6. We see that there is a majority of congruence in the answers to the prosody and comprehension questions for *alors* expressing *consequence* and *topic-shift* and, to a lesser extent, *et* expressing *addition* and *consequence*. In these cases, prosody can be interpreted as “helping” the identification of the discourse relation. *Concession* presents a large proportion of “non helping” cases, where participants prefer the expected prosody but misunderstand the discourse relation, which nuances the role of prosody for the interpretation of this otherwise complex relation. *Specification* (both for *et* and *alors*) and *temporality* show the largest proportions of “ignored” prosody, where participants identify the correct relation in spite of preferring the conflicting prosody, which further qualifies the relation between prosody and comprehension for these relations, as already shown by the regression model above. Lastly, *temporality* and *specification* relations marked by *et* present more than 25% of incorrect answers to both questions (conflicting prosody and incorrect comprehension). This could

indicate that, for these cases, the preferred prosodic profile influences the participant's interpretation: by listening to and selecting the profile of the conflicting relation (e.g. *addition* instead of *temporality*), they tend to interpret this other relation instead of the original one.

Figure 6: Mapping of answers for the prosodic profile and comprehension question [Insert Figure 6 here.](#)

To further examine this interrelation between prosody and discourse interpretation, we ran an additional mixed effects logistic regression model testing our second hypothesis on the facilitating effect of prosody on the identification of the relation (Model 2, Appendix). In this model (medium variance coverage, conditional R square = 0.367), selecting the cooperating prosodic profile has a significant positive effect on the correct identification of the relation ($p < 0.05$), which confirms our hypothesis. Moreover, *concession* stands out as particularly difficult to identify ($p < 0.01$), which points to the relative complexity of this relation involving counter-expectations and negated causal inferences (Sanders et al., 1992; Drenhaus, Demberg, Köhne, & Delogu, 2014).

Finally, we ran an additional regression model (Model 3, Appendix) testing the interaction between prosody (whether the participants preferred the cooperating profile) and the type of relation. In this model (conditional R square = 0.39), the following significant effects can be observed: the DM *et* has a significant negative impact ($p < 0.01$), suggesting that relations marked by *et* are harder to identify and that *concession* is difficult to identify ($p < 0.01$). The most interesting effect concerns *topic-shift*: while this relation is harder to identify without prosody ($p < 0.05$), the model shows that the interaction between *topic-shift* and cooperating prosody is highly significant ($p < 0.001$). In other words, people who recognize the typical prosody of *topic-shift* tend to correctly identify the discourse relation as well. This interaction between cooperating prosody and the *topic-shift* relation is in line with the clearly contrasted prosodic profile of this function of *alors*.

In sum, there is a clearly significant link between prosodic preference and discourse interpretation. In particular, the *temporality* and *specification* relations stand out as negatively associated with the selection of the cooperating prosodic profile, whether the relation identification is correct or not. For correct answers only, *topic-shift* increases the chances of selecting the cooperating prosody. This particular association between the *topic-shift* function and its contrasted prosodic profile was further evidenced by the regression model testing the impact of prosody on relation identification: if the cooperating prosody is recognized and preferred, the *topic-shift* relation is identified correctly. This result is particularly noteworthy given that *topic-shift* is the “rival” relation of *consequence*, which is the highly frequent basic meaning of *alors*, also easily recognized.

6. Study 3: Prosody-based prediction

6.1 Materials, procedure and participants

In the third experiment, the same 64 utterances were once more used, always in their cooperating prosodic versions (the conflicting profile was never played). The specificity of this study is that only the first segment (S1) and the discourse marker were available to the participants (not the S2) and their task was to predict the content of the following segment introduced by the marker. With this experiment, we wanted to test whether prosody alone would project a specific discourse relation as direct continuation.

Stimuli were presented in a random order to avoid any training effect and to be consistent with the previous experiments, despite the relative difficulty of the task. Participants had to select among two paraphrases the one they thought best represented the content of the segment to come. For each given S1 and DM, the participant would see the correct paraphrase and the paraphrase of the contrasted relation (e.g. *addition* and *temporality*), as well as pictures

representing symbolically the nature of the discourse relations. The paraphrases (translated from French) are presented in Table 6 and the pictures in Figure 7.

Insert Table 6 here.

Figure 7: Pictures representing (from left to right) relations of temporality, addition, consequence, specification, concession and topic-shift Insert Figure 7 here.

The paraphrases differ from those used in the forced choice study as they only focus on the S2, in order to draw the participants' attention to the projected continuation. The pictures aimed at facilitating the task and clarifying the differences between the discourse relations. By associating a relation with a picture, the participant does not have to re-read the full paraphrase during the whole experiment, thus improving their memorization and allowing them to concentrate on the projection task.

In addition, participants had to go through an online training phase before starting the actual experiment, in order to familiarize themselves with the discourse relations and with the task. The training phase was divided in two parts. In the first part, participants were presented with one written stimulus per relation and per discourse marker. Every example was followed by a reformulation of the example stressing the nature of the discourse relation, and by the paraphrase which is used during the experiment proper.

Consider the example below:

(3) J'ai deux frères **et** mon mari a une sœur.

*I have two brothers **and** my husband has one sister.*

- *Our family is composed of my brothers as well as of my husband's sister.*
- *The person adds a piece of information to the first utterance.*
- *We could add "**and in addition**" my husband has one sister.*
- *The speaker adds another piece of information.*

In the second part of the training phase, participants heard one spoken stimulus per relation and per discourse marker and had to choose one of two possible paraphrases. In this part, the full stimuli were presented (S1+DM+S2). Furthermore, while the stimuli were recorded by the same speaker as the stimuli used in the “real” studies, the ones used in this training session differed by being recorded with a “neutral” prosody (silent pauses of equal length before the DM, rather monotonous f0). Thus, no specific pauses or prosodic boundaries were inserted between the S1, the DM and the S2 of the utterance.

Two groups of 30 participants were recruited via a Participants’ Pool of the Faculty of Psychology of the Université catholique de Louvain (on Facebook) (47 female, 13 male; aged 21.65 on average). They were remunerated for their participation.

6.2 Hypotheses

For this continuation task, we expected more correct predictions for *alors*, which presents clearly contrasted prosodic profiles for each discourse relation it expresses, than for *et*, where the prosodic parameters are less differentiated across discourse relations. Study 2 also showed that correct identification of discourse relations is more strongly associated with the cooperating prosodic profile for *consequence*, *topic-shift* and *addition* (cf. Figure 6), suggesting more correct predictions for these relations.

6.3 Results

Once again, the situations clearly differ for *et* and *alors*. They can be seen in Figure 8. The additive conjunction shows a lot of variation between relations and within *addition* and *temporality*. While the average of correct predictions is highest for *addition*, the proportion varies greatly for the different stimuli. *Consequence* is correctly predicted in about 60% of the cases, which is in line with our hypothesis. Answers to *specification* relations are more consistently erroneous, with only 40% of correct predictions on average and little variation

between stimuli. Given that the *specification* relation is expressed by *et* in almost 10% of all tokens in the DisFrEn corpus (Crible, 2017), this result cannot be related to the general rarity of this relation, nor to a difference in the content of the segments, since all relations were presented with the S1 only. Instead, it confirms that the prosodic profiles of *et* are not contrasted or informative enough to substantially contribute to discourse interpretation on their own, especially for *specification* which was already identified as not clearly associated with a prosodic profile (cf. Study 2).

Figure 8: Proportions of correct predictions by relation for ET and ALORS, based on cooperating prosody Insert Figure 8 here.

For *alors*, however, the predictions based on cooperating prosody are much better, as expected: the correct prediction is always above chance level for each relation and there is less variation within each relation. *Concession*, *specification* and *topic-shift* are quite similar with an average of about 55% of correct answers. Predictions for *consequence* clearly stand out as being almost always correct. This finding has several explanations: Study 2 showed that the integrated prosody for this relation is well entrenched in the participants' perception, which can ease the recognition of the relation on the basis of S1 and prosody alone; *consequence* is a very frequent use of *alors* in spoken French (most frequent function in DisFrEn); the "rival" relation of *consequence*, viz. *topic-shift*, has a very specific prosodic profile with which it cannot be easily confused (cf. Table 4).

Another interpretation of this finding is more cognitive and relates to the causality-by-default hypothesis developed by Sanders (2005) (see also Sanders & Mak, 2017). Converging evidence from acquisition and processing studies point to a natural and universal cognitive bias towards causal interpretations (which includes *cause* and *consequence* relations) whenever possible. More specifically, in the absence of any cue indicating otherwise, readers and listeners will tend to infer a causality link between two adjacent segments, as in "She came *and* the party

was ruined” (the party was ruined *because* she came). This preference for causal and consequential interpretations is further supported by the present perception studies, which consistently showed high scores for the *consequence* relation either marked by *et* or by *alors*. Regarding Study 3 in particular, the participants’ outstanding performance for *consequence* can be at least partly related to this causality-by-default hypothesis.

Regression models for Study 3 perform well with high variance coverage (conditional R^2 of 0.43). With type of relation, markers and stimuli ambiguity as fixed effects, participants and stimuli as random effects, the logistic model returns a (marginally) significant effect for the DM *et* as negatively associated with correct predictions ($p < 0.1$) and a significant effect for stimuli ambiguity ($p < 0.05$). In other words, predictions are less likely to be correct for stimuli with *et* and more likely to be correct for non-ambiguous stimuli (that is, stimuli that had been correctly identified on a written basis). Regarding relations, there is a clear divide between *addition* and *consequence* on the one hand (relatively easy to predict) and the other four relations on the other (*specification*, *temporality*, *concession*, *topic-shift*), which are all significantly more difficult to predict (Model 4, Appendix).

In sum, the relations introduced by *et* are less obvious to predict than those by *alors* despite the help of prosody, which relates to the difference in prosodic profiles (highly contrasted or not) between the two DMs (see Table 4). *Addition* and *consequence* stand out from all the other relations as easier to predict, which we interpret as a result of their being the basic meanings of the DMs under scrutiny, in addition to the potential causality-by-default bias discussed above. The predictions are also positively influenced by the low ambiguity of the stimuli, measured with the annotation task on the written data, which points to the high variation between stimuli and the possible contribution of the content of the S1 on the identification of the discourse relation.

7. Discussion and conclusion

In this paper, we reported on three perception experiments targeting the impact of prosody on the acceptability and interpretation of discourse relations expressed by the French discourse markers *et* ‘and’ and *alors* ‘well/then’. The first study comparing written and audio stimuli showed that the spoken mode with expected prosody improves the acceptability of the utterances, especially for *alors*, where the increase of acceptability was larger, although the scores remain lower than for *et*.

The forced-choice experiment (Study 2) suggested that there seems to be a specific prosodic profile associated with each meaning-in-context of *alors*, and this profile is preferred in the majority of cases against a rival “conflicting” prosodic profile. For *et*, on the other hand, only the unmarked, integrated prosody (with the shortest pause before the DM and no expansion of the f0 span) was preferred in most cases, leaving only one prosodic profile for *et* regardless of its meaning-in-context. The helping effect of prosody is less present for *specification* (expressed by both *et* and *alors*) and for *temporal* relations. This result is in line with those observed by Kleinhans et al. (2017) who concluded on the importance of prosodic features for three out of the seven discourse relations they studied.

The continuation task (Study 3, predicting the relation with only the first segment and DM with a specific prosodic profile) indicated that the discourse relation was correctly predicted in the majority of cases, especially for *consequence* marked by *alors*, although answers vary greatly for relations marked by *et*.

Our results allow us to refine the findings in Gravano et al. (2005), who convincingly showed that the basic meaning of the marker is often preferred. In our data, the prevailing status of *consequence* for *alors* and the absence of difference between *addition* and *consequence* for *et* suggest that the perception and interpretation of discourse relations is not only influenced by the semantics of the marker, but also by a bias for consequence, which is in line with previous

research in cognitive linguistics pointing at such a natural tendency towards causal interpretations (Sanders, 2005)

These results clearly point to an association between prosody and semantics-pragmatics: the more contrasted the prosodic profiles of a DM, the more polysemous it is. The clearer contrasts in the productions of *alors* indeed correspond to the stronger differences between the various discourse relations it can express: positive (*consequence, specification, topic-shift*) and negative (*concession*) in terms of Sanders et al.'s (1992) primitives; causal (*concession, consequence*) and non-causal (*specification, topic-shift*); objective (*consequence, specification*), subjective (*concession*) and discourse-structuring (*topic-shift*).

By contrast, *et* is monosemous and only encodes *addition*, its other uses are inferred from the context (world knowledge, content of the segments), which can explain why the prosodic profiles are not as clearly contrasted. The lower proportions of correct answers for *et* in Studies 2 and 3 also relate to the conceptual proximity between the four relations expressed by *et*: *specification, temporality* and *consequence* all share a basic positive additive relation to which more specific information is given (addition + degree of specificity; addition + temporal inference; addition + causal inference, respectively). Such a mapping is not as clear for the senses of *alors*.

Finally, some discourse relations showed greater variation and different patterns of answers. In particular, *temporality* and *specification* relations marked by *et* did not clearly favor any configuration of answers to the prosody and comprehension questions in Study 2 (cf. Figure 6). Such variation can be explained by stimuli-internal factors, as suggested by the improvement of variance coverage (conditional R square) in the regression models with the stimuli as random effects. These factors can range from lexical or syntactic cues in the linguistic environment of the discourse marker to world knowledge and pragmatic inferences, which effectively orient the interpretation of the discourse relation, perhaps to a greater extent than the marker itself or

than its associated prosodic profile. Such contextual factors call for further investigation, in order to better grasp the mechanisms and various sources of coherence in spoken discourse interpretation.

Acknowledgments

This work was supported by the Belgian F.R.S./FNRS under Grant No. F 4520.16 (MIS Project “PhraDiCo: Prosodic phrasing for discourse comprehension. Perception of prosodic boundaries and their contribution to the construal of discourse coherence”, PI: Anne Catherine Simon). Technical support for cut-and-paste of paired sound files and silent pause duration adjustment was provided by George Christodoulides. We would like to thank Thomas François and Mickaël De Backer for their advice on the statistical analysis of the data, and the anonymous reviewers for their insightful comments. Any remaining errors are ours.

References

- Asr, F.T., & Demberg, V. (2015). Uniform information density at the level of discourse relations: Negation markers and discourse connective omission. In *Proceedings of the 11th International Conference on Computational semantics* (pp. 118–128). London: Association for Computational Linguistics.
- Blakemore, D., & Carston, R. (1999). The pragmatics of and-conjunctions: the non-narrative cases. *UCL Working Papers in Linguistics*, (11), 1–20.
- Boersma, Paul & Weenink, David (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.29, <http://www.praat.org/>
- Cain, K., & Nash, H. M. (2011). The influence of connectives on young readers’ processing and comprehension of text. *Journal of Educational Psychology*, 103(2), 429–441. <https://doi.org/10.1037/a0022824>
- Crible, L. (2018). *Discourse Markers and (Dis)fluency: Forms and Functions across Languages and Registers*. Amsterdam: John Benjamins.

- Crible, L. (2017). Discourse markers and (dis)fluencies in English and French: Variation and combination in the DisFrEn corpus. *International Journal of Corpus Linguistics*, 22(2), 242–269.
- Crible, L., & Cuenca, M-J. (2017). Discourse markers in speech: Distinctive features and corpus annotation. *Dialogue & Discourse*, 8(2), 149–166.
- Degand, L., Martin, L., & Simon, A. C. (2014). LOCAS-F: un corpus oral multigenre annoté. In *Proceedings of Congrès Mondial de Linguistique Française* (pp. 2613–2626).
- Degand, L., Cornillie, B., & Pietrandrea, P. (2013). Modal particles and discourse markers: Two sides of the same coin? In *Discourse Markers and Modal Particles. Categorization and description* (L. Degand, B. Cornillie, P. Pietrandrea, pp. 1–18). Amsterdam: John Benjamins Publishing Company.
- Degand, L., & Fagard, B. (2011). Alors between discourse and grammar: the role of syntactic position. *Functions of Language*, 18(1), 19–56.
- Delais-Roussarie, E., Post, B., Avanzi, M., Buthke, C., Di Cristo, A., Feldhausen, I., Sun-Ah Jun, S.-A., Martin, P., Meisenburg, T., Rialland, A., Sichel-Bazin, R. & Yoo, H.-Y. (2015). Intonational phonology of French: Developing a ToBI system for French. In S. Frota & P. Prieto (Éd.), *Intonation in Romance* (p. 63-100). Oxford: Oxford University Press.
- Didirková, I., Christodoulides, G., Crible, L., & Simon, A.-C. (2018). Naïve annotations of French et and alors : comparison with experts and effect of implicitation. In *Proceedings of Textlink Final Conference Cross-Linguistic Discourse Annotation. Application & Perspectives* (pp. 47–53). Toulouse, France.
- Didirková, I., Christodoulides, G., & Simon, A.-C. (2018). The Prosody of Discourse Markers alors and et in French. A Speech Production Study. Presented at the Speech Prosody 9, Poznan, Poland.

- Didirková, I., Simon, A. C., & Christodoulides, G. (2017). *Same discourse markers, different discourse relations: What do we learn from naïve annotation experiments?* Poster presented at the Discourse Markers in Romance Languages, Louvain-la-Neuve.
- Drenhaus, H., Demberg, V., Köhne, J., & Delogu, F. (2014). Incremental and predictive discourse processing based on causal and concessive discourse markers: ERP studies on German and English. In *Proceedings of CogSci 2014* (pp. 403–408). Quebec City, Canada.
- Fabricius-Hansen, C. (2005). Elusive connectives. A case study on the explicitness dimension of discourse coherence. *Linguistics*, 43(1), 17–48.
- Fischer, K. (2006). *Approaches to discourse particles. Studies in Pragmatics* (K. Fischer, Vol. 1). Ontario: Elsevier Sciences Ltd.
- Gerecht, M.-J. (1987). Alors : opérateur temporel, connecteur argumentatif et marqueur de discours. *Cahiers de Linguistique Française*, 8, 69–79.
- Gravano, A., Benus, S., Chávez, H., Hirschberg, J., & Wilcox, L. (2007). On the role of context and prosody in the interpretation of 'okay'. In *ACL*. Retrieved from <http://www.academia.edu/download/30624787/P07-1.pdf#page=838>
- Hansen, M.-B.M. (1997). Alors and donc in spoken French: A reanalysis. *Journal of Pragmatics*, (28), 153–187.
- Kleinhans, J., Farrús, M., Gravano, A., Pérez, J. M., Lai, C., & Wanner, L. (2017). Using Prosody to Classify Discourse Relations (pp. 3201–3205). ISCA. <https://doi.org/10.21437/Interspeech.2017-710>
- Luscher, J.M., & Moeschler, J. (1990). Approches dérivationnelles et procédurales des opérateurs et connecteurs temporels: les exemples de et et de enfin. *Cahiers de Linguistique Française*, (11), 77–104.

- Mak, P., Tribushinina, E., & Andreiushina, E. (2013). Semantics of connectives guides referential expectations in discourse: An eye-tracking study of Dutch and Russian. *Discourse Processes*, 50(8), 557–576. <https://doi.org/10.1080/0163853X.2013.841075>
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical Structure Theory: towards a functional theory of text organization. *Text*, 8(3). Retrieved from <http://semanticsarchive.net/Archive/GMyNDBjO/RST%20towards%20a%20functiona1%20theory%20of%20text%20organization.pdf>
- Mayer, J., Jasinskaja, E., & Kölsch, U. (2006). Pitch range and pause duration as markers of discourse hierarchy: perception experiments. In *INTERSPEECH*. Retrieved from https://www.researchgate.net/profile/Joerg_Mayer5/publication/221491193_Pitch_range_and_pause_duration_as_markers_of_discourse_hierarchy_Perception_experiments/links/56fa935208aede4cf4fa53bc.pdf
- Michelas, A. (2011). *Caractérisation phonétique et phonologie du syntagme intermédiaire en français : de la production à la perception* (Doctoral thesis). Université de Provence - Aix-Marseille I.
- Michelas, A., & D’Imperio, M. (2015). Prosodic boundary strength guides syntactic parsing of French utterances. *Laboratory Phonology*, 6(1), 119–146. <https://doi.org/10.1515/lp-2015-0003>
- Mertens, P. 2014. Polytonia: a system for the automatic transcription of tonal aspects in speech corpora. *Journal of Speech Sciences* 4(2), 17-57.
- Mertens, P. & Simon, A.C. (2013). Towards automatic detection of prosodic boundaries in spoken French. In Mertens, P. & Simon, A.C. (eds). *Proceedings of the Prosody-Discourse Interface Conference 2013 (IDP-2013)*. Leuven, September 11-13, 2013. pp. 81-87.

- Murray, G., Renals, S., Carletta, J., & Moore, J. (2006). Incorporating speaker and discourse features into speech summarization. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics* (pp. 367–374). New York: Association for Computational Linguistics.
- Murray, J. (1997). Connectives and narrative text: The role of continuity. *Memory & Cognition*, 25(2), 227–236.
- Nolan, F. (2003). Intonational equivalence: An experimental evaluation of pitch scales. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 771 - 774). https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/papers/p15_0771.pdf
- Ouden, H. den, Noordman, L., & Terken, J. (2009). Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports. *Speech Communication*, 51(2), 116–129. <https://doi.org/10.1016/j.specom.2008.06.003>
- Petukhova, V., & Bunt, H. (2009). Towards a Multidimensional Semantics of Discourse Markers in Spoken Dialogue. In *IWCS-8 '09 Proceedings of the Eight International Conference on Computational Semantics* (pp. 157–168). Tilburg, The Netherlands. Retrieved from </paper/Towards-a-Multidimensional-Semantics-of-Discourse-Petukhova-Bunt/9df8d66a1a9fae60193df6e67ae4704d51f1198b>
- Pitler, E., Louis, A., & Nenkova, A. (2009). Automatic sense prediction for implicit discourse relations in text. In *Proceedings of the Joint Conference of ACL-IJCNLP* (pp. 683–691). Suntec, Singapore.
- Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A., Robaldo, L., & Webber, B. (2007). The Penn Discourse Treebank 2.0 Annotation Manual. *IRCS Technical Reports Series*. Retrieved from https://repository.upenn.edu/ircs_reports/203

- Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, *90*, 2956–2970.
- Pyykkönen, P., & Järvikivi, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*, *57*(1), 5–16.
- Rohde, H., & Horton, W. S. (2014). Anticipatory looks reveal expectations about discourse relations. *Cognition*, *133*(3), 667–691. <https://doi.org/10.1016/j.cognition.2014.08.012>
- Rohde, H., Tyler, J., & Carlson, K. (2017). Form and function: Optional complementizers reduce causal inferences. *Glossa: A Journal of General Linguistics*, *2*(1). <https://doi.org/10.5334/gjgl.134>
- Sanders, T. (2005). Coherence, causality and cognitive complexity in discourse. *Proceedings/Actes SEM-05, First International Symposium on the Exploration and Modelling of Meaning*, 105–114.
- Sanders, T. J. M., Spooren, W. P. M., & Noordman, L. G. M. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, *15*(1), 1–35. <https://doi.org/10.1080/01638539209544800>
- Sanders, Ted, & Mak, P. (2017). Do we seek for causality in discourse? On the cognition of coherence relations and connectives. Presented at the 50th Annual Meeting of the Societas Linguistica Europaea, Zurich, Switzerland.
- Saussure, L., & Sthioul, B. (2002). Interprétations cumulative et distributive du connecteur et: temps, argumentation, séquencement. *Cahiers de Linguistique Française*, *(24)*, 293–314.
- Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, *29*(2), 169–182.

- Slowiaczek, M. L. (1981). *Prosodic units as language processing units*. Retrieved from http://scholarworks.umass.edu/cgi/viewcontent.cgi?article=2662&context=dissertations_1
- Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, 25(2), 249–271. <https://doi.org/10.1007/BF01708573>
- Spooren, W. P. M., & Degand, L. (2010). Coding coherence relations: Reliability and validity. *Corpus Linguistics and Linguistic Theory*, 6(2), 241–266. <https://doi.org/10.1515>
- Spooren, Wilbert. (1997). The processing of underspecified coherence relations. *Discourse Processes*, 24(1), 149–168.
- Tyler, J. (2014). Prosody and the Interpretation of Hierarchically Ambiguous Discourse. *Discourse Processes*, 51(8), 656–687. <https://doi.org/10.1080/0163853X.2013.875866>
- Tyler, L. K., & Warren, P. (1987). Local and Global Structure in Spoken Language Comprehension. *Journal of Memory and Language*, 26(87), 638–657.
- Zufferey, S., & Gygax, P. M. (2017). Processing Connectives with a Complex Form-Function Mapping in L2: The Case of French “En Effet.” *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.01198>

Appendix

Regression models

Model 1

PROSODY~DM+AMBIGUOUS+RIGHT+PROP_RIGHT+PROP_RIGHT_WRITTEN+ET.

CSQ+ET.TMP+ET.SPE+ALO.CCS+ALO.TOS+ALO.SPE+(1|RATER)+(1|FILE),

data=exp2, family="binomial"

Fixed effects: Insert Table 7 here.

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Model 1bis

PROSODY~DM+AMBIGUOUS+PROP_RIGHT+PROP_RIGHT_WRITTEN+ET.TMP+ET

.CSQ+ET.SPE+ALO.CCS+ALO.SPE+ALO.TOS+(1|RATER)+(1|FILE), data=exp2true,

family="binomial"

Fixed effects: Insert Table 8 here.

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Model 2

RIGHT~DM+PROSODY+AMBIGUOUS+PROP_RIGHT_WRITTEN+ET.CSQ+ET.TMP+

ET.SPE+ALO.CCS+ALO.TOS+ALO.SPE+(1|RATER)+(1|FILE), data=exp2,

family="binomial"

Fixed effects: Insert Table 9 here.

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Model 3

RIGHT~DM+PROSODY+AMBIGUOUS+PROP_RIGHT_WRITTEN+(PROSODY*RELATION)+(1|RATER)+(1|FILE), data=exp2, family="binomial"

Fixed effects: [Insert Table 10 here.](#)

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Model 4:

RIGHT~DM+ID_WRITTEN+ET.SPE+ET.TMP+ET.CSQ+ALO.CCS+ALO.SPE+ALO.TOS

+(1|RATER)+(1|FILE)

Fixed effects: Insert Table 11 here.

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

<i>ALORS</i>	<i>ALORS/ET</i>	<i>ET</i>
Concession		Addition
Topic shift	Consequence	Temporality
	Specification	

Table 1 : Discourse markers and the discourse relations they induced

[dès qu'on a un événement de communication on a un style de parole] _{S1} [et] _{DM} [each time we have a communicative event we have a speech style] _{S1} [and] _{DM}		
	S2	Relation
Original	[ce style de parole se caractérise par une vitesse de parole] _{S2} [this speech style is characterized by the speech rate] _{S2}	Specification
Constructed S2 no. 1	[on peut observer les gestes des locuteurs] _{S2} [we can observe speakers' gestures] _{S2}	Addition
Constructed S2 no. 2	[on va chercher à le décrire] _{S2} [we will try to describe it] _{S2}	Temporality
Constructed S2 no. 3	[on a certaines caractéristiques prévisibles] _{S2} [we have some predictable characteristics] _{S2}	Consequence

Table 2 : Example of original and constructed S2 for the sequence S1+DM

Integrated prosodic profile	<i>et-CSQ</i> <i>et-ADD</i>
-----------------------------	------------------------------------

In-between	<i>et</i> -SPE <i>et</i> -TMP <i>alors</i> -CCS <i>alors</i> -CSQ
Detached (or segmented) prosodic profile	<i>alors</i> -SPE <i>alors</i> -TOS

Table 3 : Summary of observed prosodic profiles per relation and per DM (see Didirkova et al. 2018 for detailed acoustic measurements)

		Level of prosodic integration	Mean duration of the DM	Pause duration before the DM	Mean pitch reset on the DM
1b	et-ADD	integrated	105 ms ¹	125 ms	1 ST
1a	et-TMP	semi-detached	126 ms	450 ms	5 ST
2a	et-CSQ	integrated	101 ms	125 ms	-1 ST
2b	et-SPE	semi-detached	132 ms	500 ms	1 ST
3a	alors-CCS	semi-detached	268 ms	300 ms	5 ST
3b	alors-SPE	detached	396 ms	500 ms	9 ST
4a	alors-CSQ	semi-detached	371 ms	100 ms	3 ST
4b	alors-TOS	detached	478 ms	700 ms	11 ST

¹ All (mean) durations are given in milliseconds.

Table 4 : Acoustic measures per discourse marker and discourse relation

Et	addition	The speaker provided a new piece of information.
	temporality	The speaker mentioned an event that occurred after the one she just discussed.
	consequence	The speaker gave the consequence or logical result to what she just said.
	specification	The speaker gave a more specific information, a detail or an example.
Alors	consequence	The speaker gave the consequence or logical result to what she just said.
	topic change	The speaker talked about a different theme or topic in the second sentence.
	specification	The speaker gave a more specific information, a detail or an example.
	concession	The speaker nuanced or partly contradicted what she just said.

Table 5 : Paraphrases for the whole sequence per discourse relation (Study 2)

Et	addition	The speaker is about to add another piece of information.
	temporality	The speaker is about to explain what happened next.
	consequence	The speaker is about to give a logical consequence.
	specification	The speaker is about to add a detail or an example.
Alors	consequence	The speaker is about to give a logical consequence.
	topic change	The speaker is about to introduce a new topic.
	specification	The speaker is about to add a detail or an example.
	concession	The speaker is about to provide a counter-argument.

Table 6: Paraphrases for the S2 by relation and marker (Study 3)

	Estimate	Standard error	Z value	Pr (> z)
(Intercept)	-0.1498377	0.6531368	-0.229	0.818548

Discourse marker (et)	0.1911946	0.3149412	0.607	0.543798
Ambiguity (yes)	0.1095794	0.3004116	0.365	0.715287
Correct identification (true)	0.2434076	0.1225760	1.986	0.047059 *
Proportion of correct identifications	0.0006481	0.0043722	0.148	0.882158
Proportion of correct identifications in the written version	0.0083861	0.0058631	1.430	0.152625
ET consequence	-0.0987044	0.3058387	-0.323	0.746897
ET temporality	-1.0374805	0.3131349	-3.313	0.000922 ***
ET specification	-1.3536226	0.3129095	-4.326	1.52e-05 ***
ALORS concession	0.0762063	0.3549777	0.215	0.830017
ALORS topic shift	0.4112397	0.3136158	1.311	0.189762
ALORS specification	-0.5455300	0.3101057	-1.759	0.078548 .

Table 7: Model 1

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.530329	0.690704	0.768	0.44260
Discourse marker (et)	0.038315	0.317726	0.121	0.90401
Ambiguity (yes)	0.250457	0.308964	0.811	0.41758
Proportion of correct identifications	-0.002842	0.004993	-0.569	0.56914
Proportion of correct identifications in the written version	0.007518	0.006139	1.225	0.22068
ET temporal	-1.392279	0.336473	-4.138	3.51e-05 ***

ET consequence	-0.198004	0.312722	-0.633	0.52663
ET specification	-1.370497	0.338624	-4.047	5.18e-05 ***
ALORS concession	-0.263757	0.371466	-0.710	0.47768
ALORS specification	-0.704956	0.299645	-2.353	0.01864 *
ALORS topic shift	0.889116	0.331295	2.684	0.00728 **

Table 8: Model 1 bis

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.3770	1.0198	0.370	0.71161
Discourse marker (et)	-0.3837	0.5570	-0.689	0.49082
Prosody (cooperative)	0.2434	0.1211	2.011	0.04436 *
Ambiguity (yes)	-0.6084	0.5088	-1.196	0.23179
Proportion of correct identifications in the written version	0.0196	0.0101	1.940	0.05237 .
ET consequence	-0.1837	0.5302	-0.346	0.72901
ET temporal	-1.0218	0.5281	-1.935	0.05302 .
ET specification	-1.0805	0.5212	-2.073	0.03815 *
ALORS concession	-1.7624	0.5769	-3.055	0.00225 **
ALORS specification	-0.1331	0.5700	-0.234	0.81534
ALORS topic shift	0.1128	0.5658	0.199	0.84195

Table 9: Model 2

	Estimate	Std. Error	z value	Pr(> z)

(Intercept)	1.074495	0.983842	1.092	0.274771
Discourse marker (et)	-0.972251	0.379324	-2.563	0.010374 *
Prosody (cooperative)	0.297690	0.359153	0.829	0.407180
Ambiguity (yes)	-0.685269	0.491412	-1.394	0.163170
Proportion of correct identifications in the written version	0.017669	0.009896	1.785	0.074199 .
Concession	-2.063139	0.711312	-2.900	0.003726 **
Consequence	-0.356601	0.578500	-0.616	0.537615
Specification	-0.786087	0.550480	-1.428	0.153291
Temporality	-0.726759	0.595149	-1.221	0.222034
Topic shift	-1.705771	0.749825	-2.275	0.022912 *
Cooperative prosody : concession	-0.441372	0.457841	-0.964	0.335031
Cooperative prosody : consequence	0.010868	0.445739	0.024	0.980548
Cooperative prosody : specification	-0.224882	0.424856	-0.529	0.596588
Cooperative prosody : temporality	-0.649648	0.492734	-1.318	0.187351
Cooperative prosody : topic shift	2.052968	0.570369	3.599	0.000319 ***

Table 10: Model 3

	Estimate	Standard error	Z value	Pr (> z)
(Intercept)	0.597433	0.847230	0.705	0.48071
Discourse marker (et)	-1.183997	0.717090	-1.651	0.09872 .
Proportion of correct identifications in the written version	0.018971	0.008002	2.371	0.01775 *
ET specification	-1.430823	0.659495	-2.170	0.03004 *

ET temporality	-1.806998	0.679131	-2.661	0.007780 **
ET consequence	-0.044846	0.662792	-0.068	0.94605
ALORS concession	-1.849404	0.687015	-2.692	0.00710 **
ALORS specification	-1.869019	0.680288	-2.747	0.00601 **
ALORS topic shift	-1.681186	0.710717	-2.365	0.01801 *

Table 11: Model 4