



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Weak and strong discourse markers in speech, chat and writing

Citation for published version:

Crible, L 2020, 'Weak and strong discourse markers in speech, chat and writing: Do signals compensate for ambiguity in explicit relations?', *Discourse Processes*, vol. 57, no. 9, pp. 793-807.
<https://doi.org/10.1080/0163853X.2020.1786778>

Digital Object Identifier (DOI):

[10.1080/0163853X.2020.1786778](https://doi.org/10.1080/0163853X.2020.1786778)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Discourse Processes

Publisher Rights Statement:

This is an Accepted Manuscript of an article published by Taylor & Francis in Discourse Processes on 14.8.2020, available online:
<https://www.tandfonline.com/doi/abs/10.1080/0163853X.2020.1786778?journalCode=hdsp20>

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Weak and strong discourse markers in speech, chat and writing: Do signals compensate for ambiguity in explicit relations?

Ludivine Crible

Department of Psychology, University of Edinburgh

Author Note

Ludivine Crible, Department of Psychology, University of Edinburgh.

This research was supported by the Marie Skłodowska-Curie grant number 794575.

Correspondence concerning this article should be addressed to Ludivine Crible, Department of Psychology, University of Edinburgh, 7 George Square, EH8 9JZ, Edinburgh, United Kingdom. E-mail: ludivine.crible@uclouvain.be

To cite this paper: Crible, L. (In press). Weak and strong discourse markers in speech, chat and writing: Do signals compensate for ambiguity in explicit relations? *Discourse Processes*. doi:10.1080/0163853X.2020.1786778

Abstract

Ambiguity in discourse is pervasive, yet mechanisms of production and processing suggest that it tends to be compensated in context. The present study sets out to analyze the combination of discourse markers (such as *but* or *moreover*) with other discourse signals (such as semantic relations or punctuation marks) across three genres (discussion, chat, and essay). The presence of discourse signals is expected to vary with the ambiguity of the discourse marker and with the genre. This analysis complements Das and Taboada's (2018) approach to discourse signalling by zooming in on the different types of discourse markers with which other signals combine. The corpus annotation study uncovered three categories of marker strength – weak, intermediate and strong – thus refining the concept of 'explicitness'. Statistical modeling reveals that weak discourse markers are more often compensated than intermediate and strong markers, and that this compensation is not affected by genre variation.

Keywords: discourse markers, ambiguity, discourse signals, explicitness, information density

Weak and strong discourse markers in speech, chat and writing: Do signals compensate for ambiguity in explicit relations?

Discourse, either written or spoken, is built upon coherence relations, which specify the nature of the link that connects clauses and other discourse segments. These coherence relations can be additive, contrastive, causal, conditional, etc. Interpreting coherence relations is crucial for comprehenders to achieve understanding of a discourse (e.g. Sanders & Noordman, 2000). They have repeatedly been studied in connection with the category of discourse markers, which act as “cue phrases” or signals for particular relations (e.g., Knott & Dale, 1994).¹ Following Das & Taboada’s (2018) recent line of research, this study sets out to examine the polyfunctionality of discourse markers (henceforth DMs) in relation to other signals that may compensate for their ambiguity. Corpus analyses test whether the ambiguity of DMs and their co-occurrence with other signals vary across genres and across coherence relations.

DMs are the prototypical signals for coherence relations, although they can also perform other functions, such as topic organization or interpersonal management (e.g., Schiffrin, 1987). Despite this ambiguity or polyfunctionality of DMs, no one would question that *whereas* in (1) expresses a contrastive relation.

(1) You cannot overdose with marijuana, whereas you can overdose with alcohol.

The subordinating conjunction signals a contrast between the two clauses, making the relation *explicit* in this context. In many approaches to coherence relations, the same example without the DM would be classified as an *implicit* relation, as in (2).

(2) You cannot overdose with marijuana, you can overdose with alcohol.

The contrastive interpretation can arguably be maintained from the content of the

segments, despite the absence of *whereas*. However, the apparent dichotomy between implicit and explicit relations has been refined in recent works (Das & Taboada, 2018; Hoek, Zufferey, Evers-Vermeul, & Sanders, 2019; Péry-Woodley, Ho-Dac, Rebeyrolle, Tanguy, & Fabre, 2018), which also consider additional signals of coherence relations beyond DMs. In examples (1-2) above, the different polarity between the two segments (negative-positive) and the syntactic parallelism between them act as signals for the contrastive relation, since they contribute to the interpretation of contrast and comparison. The role of signals such as polarity or parallel constructions is thus paramount in these examples. They can be considered as reinforcing the connective, as in (1), or compensating for its absence, as in (2).

Despite the growing interest for these other discourse signals, they have not yet been systematically investigated in relation with the type of DM with which they co-occur, especially in terms of DM ambiguity. Not all DMs are equal in how strongly they signal a coherence relation (Asr & Demberg, 2012), depending on the number and frequency of other relations which they can express: a monosemous marker like *whereas* is stronger in expressing contrast than the additive conjunction *and*. Yet, in (3), a relation of contrast can still be easily interpreted.

(3) You cannot overdose with marijuana, and you can overdose with alcohol.

Here, the relation relies more on the negation and the parallelism than on the DM itself; still, such less standard co-occurrences are not uncommon and deserve further investigation. One could hypothesize that the “weaker” the DM for a given relation, the more it will be compensated by other signals, and that this compensation will vary across genres, under the influence of planning pressure and/or recipient design (Spooren, 1997). This study thus examines the interaction between discourse signals on

the one hand, and ambiguity of the DMs on the other, including genre as a potential factor of variation. In doing so, it refines the divide between explicit and implicit relations by introducing weak, intermediate and strong DMs in the continuum.

The next section takes stock of previous discourse studies on the role of signals in coherence relations. Two main accounts of their distribution are confronted, one in terms of genre, the other in terms of compensation. The materials and annotation procedure are detailed in the following section. DMs expressing a subset of five relations (addition, specification, consequence, concession and contrast) are categorized into degrees of strength and coded for presence of reinforcing signals. Multivariate models test whether signals can be predicted by DM ambiguity, type of relation and genre. Finally, we discuss the significance of these results for discourse analysis and processing.

Coherence relations and their signals

Discourse markers in coherence relations

In most frameworks, such as the Penn Discourse Treebank 2.0 (henceforth PDTB; Prasad et al., 2008), coherence relations are classified into labels that correspond to distinct (sub)types of relation. Originally designed for written corpora, such taxonomies cover both explicit and implicit relations (i.e. with and without a DM). Several corpus studies have shown that some relations are more often left implicit than others, and that factors such as expectedness, cognitive complexity or genre seem to at least partially motivate the explicit vs. implicit marking of a relation (Asr & Demberg, 2012; Hoek, Zufferey, Evers-Vermeul, & Sanders, 2017; Taboada, 2006). From a cognitive perspective, implicit relations are less informative and thus more demanding for the reader or hearer than their explicit versions. For instance, Sanders and Noordman

(2000) found that coherence relations (problem-solution and list relations, in their study) are processed faster when they are connected by a DM than when they are “implicit”. Similarly, Millis and Just (1994) observed faster and better answers to comprehension questions for causal relations when “because” (and “although”) was used.

With much of the attention being paid to explicit and implicit relations, one DM has been particularly neglected so far in studies on the marking of coherence relations: the additive conjunction *and*. Although often excluded from studies (experimental or corpus-based) because of its ambiguity, *and* is one of the most frequent DM in spoken and written English (Crible, 2017; Prasad et al., 2008) and can express relations as varied as addition, consequence, specification, contrast or concession. Little is known about the production and comprehension of *and*, which is often discarded as a low-information additive DM. One exception is Spooren (1997), who adopts a Gricean approach to the underspecified use of *and* to express a causal or temporal relation. Spooren showed that such ambiguity is less frequent in proficient speakers than in children or second-language learners, suggesting that this use of *and* relates to a principle of speaker economy rather than of hearer economy. Cain and Nash (2011) further showed that sentences linked by *and* take longer to be read, which is indicative of the time spent computing the appropriate relation between two clauses, which is implied but not stated (2011, p. 436). It is therefore not trivial to locate relations marked by *and* on the implicit vs. explicit continuum: a DM may be present but not fully encode the relation at hand.

This line of reasoning can be extended to other DMs as well: is a contrastive relation expressed by *when* truly explicit, considering that this DM is temporal in most of its uses? Many DMs express more than one relation (e.g. *while* for concessive and

temporal relations, *actually* for concession and specification), although not always with the same frequency, as can be seen from large corpus annotation studies (Webber, Prasad, & Lee, 2019b). Asr and Demberg (2012) have developed a measure of “cue strength” that gives the probability for a given DM to express a given relation by taking into account its other meanings and the other DMs that express that relation. They found that only a few DMs are very strong cues for their most frequently expressed relation, while *and* and *but* are weaker because distributed over several coherence relations. If we repeat the example from the Introduction and use *but* instead of *whereas*, as in (4), we can see that there is an explicit DM which does encode contrast, but its marking strength is lower than that of *whereas* because of its other uses (for instance in concessive relations), thus suggesting yet another point on the scale from implicit to strongly marked (explicit) relations.

(4) You cannot overdose with marijuana, but you can overdose with alcohol.

From a psycholinguistic perspective, Zufferey and her collaborators found that polyfunctional DMs were a source of difficulty, both for native speakers (Zufferey, Mak, Degand, & Sanders, 2015) and learners (Zufferey & Gygax, 2017), in reading and coherence judgment tasks. However, more recently, Zufferey and Gygax (2020) found that monosemous DMs that are specific to formal writing, such as French *en outre* ‘moreover’, were also a source of difficulty for adult speakers with low exposure to print.

Other signals in coherence relations

Whether a DM is weak or strong, present or absent, other discourse signals may be used to convey or reinforce the interpretation of the coherence relation. Some

experimental studies have addressed the impact of specific contextual elements on DM processing (e.g., Mak, Tribushinina, & Andreiushina, 2013; Scholman, Rohde, & Demberg, 2017). It is Das and Taboada (2018) who have provided the most extensive study of these other signals so far. They started from the observation that many relations do not contain a DM or contain an ambiguous DM, and built a taxonomy of signals for manual identification in the RST Signalling Corpus (Das, Taboada, & McFetridge, 2015). There are nine types of signals: DMs (e.g. *whereas*), reference (e.g. demonstrative), lexical (e.g. alternative expression), semantic (e.g. synonymy), morphological (tense), syntactic (e.g. imperative), graphical (e.g. parentheses), genre (e.g. inverted pyramid scheme) and numerical (same count). These cues are identified if they are relevant to the annotated relation, i.e. when they “could best function as the indicator(s) for that relation instance” (Das & Taboada, 2018, p. 754). They report on proportions of relations signalled by DMs and/or other signals or no signal, and show that this proportion and the type of signal vary with the relation: 76% of result relations are signalled, against 95% for elaboration relations. One particularly interesting finding is that the elaboration relation (labelled “specification” in the present framework) is signalled by a variety of devices, against the common assumption that this relation is often implicit.

Their study is a milestone in our understanding of discourse marking, and the annotation is impressive and fine-grained. Some features and classifications are surprising: for instance, some semantic classifications are questionable (word pairs with sometimes very little in common); parallelism and auxiliary inversion are considered as syntactic units (on a par with relative clauses) rather than constructions; genre as a signal for a coherence relation is quite far-fetched given that all the relations in the text belong to the same genre. There are other ways in which discourse signals can be

analyzed which would be perhaps less subjective (see the Discussion), yet Das and Taboada's (2018) approach, where syntax and semantics are building up discourse, is enlightening and central to the present endeavor, which closely complements it.

Finally, one should note that such comprehensive coverage of discourse signals extends the line of work already present in the PDTB 2.0 (Prasad et al., 2008) with the annotation of *alternative lexicalizations*, or AltLex: they are expressions with an anaphoric reference that signal the meaning of the coherence relation, such as “the main reason for this situation is” (cause) or “this effort resulted in” (consequence); when an AltLex is used, an explicit DM would be superfluous. In the PDTB 3.0 (Prasad, Webber, & Lee, 2018), the category of AltLex-C (for *construction*) has been added to the annotation and corresponds to lexico-syntactic constructions which are specific to a given relation, such as auxiliary inversion for conditional relations or “so [Adj] that” for consequence relations.

Factors behind discourse signalling

Previous studies have been focusing on discourse signalling across different relations in one genre (newspaper articles). By contrast, two complementary factors of variation are analyzed here: genre and DM strength. They will each be developed below.

Firstly, different genres have different conventions regarding the amount of information that should be given to the hearer or reader. Genre is here operationalized as the medium of communication and the degree of formality and planning that goes with it. Written texts such as argumentative essays are not subject to planning pressure and require a relatively high degree of formality: the reader is not physically present and can only use the information in the text, which thus has to be maximally informative;

the writer can provide this information by taking the time to plan and edit their sentences (Chafe, 1984). Spoken conversation such as group discussions are more spontaneous and typically less formal: the speaker is under some pressure to speak in due time (Clark, 2002) and has limited cognitive resources for optimal lexical retrieval; the hearer can make use of other contextual cues such as prosody or multimodality to help interpret utterances. Computer-mediated communication such as chat conversations are somewhat intermediate: chat participants can use the affordances of writing to plan and edit their discourse production, but the rhythm and informality of most chat conversations resemble those of speech (Degand & van Bergen, 2018). The effect of genre in coherence relations and DMs has been repeatedly evidenced in previous studies (Crible, 2017; Kunz & Lapshinova-Koltunski, 2015; Louwerse & Mitchell, 2003; Taboada, 2006). Psycholinguistic studies have also shown that genre interacts with how DMs are processed, facilitating some interpretations over others (e.g., Canestrelli, Mak, & Sanders, 2016). However, nothing is known about its effect on the compensation of DMs by other signals.

Secondly, the presence of compensating signals in addition to the DM might be influenced by the strength (or ambiguity) of that DM: the stronger the DM (i.e. more monosemous, less ambiguous) is in expressing a given relation, the fewer additional signals it will co-occur with. This potential factor relates to the Uniform Information Density hypothesis (Levy & Jaeger, 2007), which suggests that speakers tend to avoid troughs and peaks in information density in their production, so that if the interpretation is already achieved by one part of the sentence, there is no need to further signal it by extra markers. For instance, if a contrastive relation is connected by the strong DM *whereas*, there is no apparent need to reinforce or “compensate” the DM with other signals such as antonyms or syntactic parallelism; if, however, the ambiguous *and* is

used, we can expect that its low informative value will be compensated by clearer signals of contrast.

Hypotheses

The present analysis aims to compare the frequency of DMs with different marking strengths across genres, to analyze in what proportion these different DMs are compensated by other signals, and to examine whether this compensation varies across genres. Three main hypotheses will be tested. Ambiguous (polyfunctional) DMs are expected to be more frequent in speech than in written genres, which favor more precise DM types (Hypothesis 1, henceforth H1). Moreover, signals are also expected to be more frequent in formal and written genres than in informal speech, as an effect of formality and planning pressure (H2). Finally, the Uniform Information Density hypothesis suggests that weak DMs such as *and* will more frequently co-occur with reinforcing signals to compensate for their ambiguity, compared to stronger, less ambiguous DMs like *whereas* that do not require such compensation (H3). This third hypothesis takes us further than distributional considerations of genre and strives to address cognitive strategies of discourse production. The three variables under scrutiny (DM strength, genre and compensation) are operationally defined in the next section.

Methodology

Corpus data

The analysis of DM ambiguity and compensation has been carried out on the Loyola CMC Corpus (Goldstein-Stewart, Goodwin, Sabin, & Winder, 2008). This corpus contains samples of text and talk from six genres, namely email, essay, interview, blog, chat and discussion. All 21 participants produced each of the six

genres; they were all U.S. university students, remunerated for their participation. The topics of conversation were also controlled: the texts are all argumentative and ask the students' opinion on six socially-relevant topics such as the legalization of marijuana or the war in Iraq. This controlled corpus design vouches for high comparability between texts.

Samples from three genres out of the six were analyzed for the present study: discussion, chat and essay. In the discussions and chat conversations, there is a moderator (a researcher) and four or five participants. Discussions were face-to-face; chat conversations were set on a private online environment; essays were typed by individual participants on a computer and submitted online. These genres were selected because they represent different constraints and demands in terms of planning pressure and formality, with a highly spontaneous spoken task, a spontaneous written task and a formal written task. The whole data from chats and essays were included, amounting to 72,466 words and 64,864 words, respectively, in addition to a stratified sample of 89,515 words from discussions, which includes a similar number of texts per topic and per participant, for a total corpus size of 226,845 words.

Annotation

Identifying discourse markers.

All tokens of DMs were manually identified in the data, following Crible's (2017) operational criteria:

Strict syntactic optionality (thus excluding prepositions such as *because of*);
discourse-level scope (i.e. presence of a finite predicate); high degree of
grammaticalization (excluding idiosyncrasies such as *and all that kind of jazz*);
procedural meaning expressing either a discourse relation (e.g. cause, contrast),

meta-comments, a structuring function (turn or topic structure) or managing the speaker-hearer relationship. (p. 252)

In other words, any procedural expression connecting clauses with propositional content was identified. This identification is straight-forward: most DMs are utterance-initial, can be removed from the sentence without altering its grammaticality, and express some coherence relation. The present identification phase was conducted by the author. A sample of two texts in each genre (1,034 DMs) was analyzed by a second expert annotator in order to measure inter-rater reliability on the DM identification phase. We report on a 95.83% agreement rate, calculated over the total number of tokens identified by either annotator. More specifically, 18 disagreements were found in chats, 27 in the discussions, and none in the essays. Disagreements mostly concerned the following DM types: *I mean, okay, like, now*. These are very frequent DMs typical of spoken-like language, and disagreements are likely due to simple omissions over their very high frequency.²

Following this definition, 109 English DM types (12,710 tokens) were annotated, which includes coordinating (*and*) and subordinating conjunctions (*because*), adverbials (*however*), prepositional phrases (*on the other hand*) and verb phrases (*I mean*). Some of them are typical of speech, some of writing, and many are shared in the three genres. Complex DMs, made of the combination of two otherwise independent DMs, are annotated as one token if they jointly express the same function and cannot be separated without change in meaning, such as some instances of *and then* (see Cuenca & Crible, 2019; Webber, Prasad, & Lee, 2019a). Not all identified DMs are strictly connecting or relational, but they are considered traditional DMs in the spoken discourse analysis literature, such as *alright, kind of, like* or *well* (Schiffrin, 1987). All tokens were only annotated if they met the criteria listed above.

Identifying coherence relations.

For the annotation of the coherence relations and other functions that DMs may perform, Crible and Degand's (2019) system was manually applied to all DM tokens. As mentioned above, this taxonomy distinguishes between 15 functions and most of them correspond to typical coherence relations: addition, alternative, cause, concession, condition, consequence, contrast, hedging, monitoring, specification, temporal, agreeing, disagreeing, topic, quoting (see Crible & Degand, 2019 for operational definitions). The function is assigned in context, making use of all available information for the interpretation: *and* will not always be annotated as "additive" nor *when* always as "temporal", even though a conservative bias towards the "dictionary" meaning of the DM was used in case of hesitation. Only one function is assigned for each DM, as multiple simultaneous functions are rare, difficult to process quantitatively and mostly correspond to cases of hesitation between two values, which are solved through the bias mentioned above. This system is somewhat intermediate in terms of granularity as it makes fewer distinctions of subtypes than other models and does not specify the order of the segments (as opposed to CONCESSION.ARG1-AS-DENIER vs. CONCESSION.ARG2-AS-DENIER in the PDTB 3.0, for instance). It does, however, specify a domain of use for all functions (ideational, rhetorical, sequential or interpersonal), but this layer of analysis is not relevant to the present study and will not be discussed any further.

The annotation was carried out by the author, with regular consistency checks to ensure the stability of the analysis. A random sample of 50 DMs in each genre was independently double-coded by a second expert annotator for inter-rater reliability measures. We report on a 90.7% of agreement on DM functions and a Fleiss' kappa of $\kappa = .892$, which is an agreement level way above the satisfying threshold of $\kappa = .7$ for such complex discourse disambiguation (Spooren & Degand, 2010). This score breaks

down to 86% in discussion, 88% in chat and up to 98% in essay texts. The improvement of reliability in the essay corpus can be explained by the narrower functional spectrum of DMs in writing, where highly polyfunctional expressions such as *I mean* or *well* almost never occur. An error analysis indeed shows that most DMs are disambiguated consistently across annotators except for *I mean* (seven out of 11 disagreements) and *well* (four out of eight disagreements).

Despite these high reliability scores, discourse annotation remains very challenging (Spooren & Degand, 2010), especially with numerous values in the coding scheme, and the analysis is always in part subjective since it relies on the researcher's interpretation of the context. Although this situation is shared among all discourse annotation frameworks, it remains a limitation to the present study, and the partly subjective nature of this analysis should be borne in mind.

Identifying signals.

Given the objective of the present analysis to study the combination of DMs with other signals, the identification of these other discourse signals was only performed when an explicit DM was present (thus excluding “implicit” relations) and only on a sample of 1,968 annotated DMs. This sample includes occurrences of a subset of five relations (addition, specification, consequence, concession, contrast; the rationale is given in the Results). Given the extremely high frequency of *and* expressing addition (2,057 tokens), we only annotated signals for a random sample of 301 tokens of this use of *and*.³

The taxonomy of signals is partly inspired by Das and Taboada (2018) and was designed through multiple testings on corpus data, adding a new signal type whenever it was found. The final list includes 26 signal types, which can be classified in four groups

(Table 1). These signals cover syntagmatic properties of the DM itself (the co-occurring signals) as well as morphosyntactic features of the segments (sentence signals), specific syntactic constructions (cf. the AltLex-C in the PDTB 3.0) and semantic features such as the presence of a semantic relation across the two connected segments or of an alternative lexicalization. Much like the approach in Das and Taboada (2018), any given DM occurrence is associated with one or several signals that are relevant to or “compatible with” (Das & Taboada, 2018, p. 765) the interpretation of the particular relation, in other words, signals that encode congruent information that reinforces the meaning of the relation and of the DM. The procedure requires that, for each DM, the annotated relation is matched with signals that share its conceptual features, such as reference continuity or contrast, causality, hierarchical status of the segments, etc.

Consider the following examples:

- (5) While it may have been a larger issue in the past half-century, it does not pose much of a problem to our society today. [essay]
- (6) they weren't attracted to their wives and they didn't want to have kids with their wives [chat]
- (7) even if he started to believe that we shouldn't be there, what should we do? [discussion]

In (5), the relation of contrast expressed by *while* is further signalled by the difference in polarity and difference in verb tense between the two segments, the comparative adjective *larger* and the deictic expressions *in the past half-century* vs. *today*. In (6), *and* expresses an additive relation, which is reinforced by the lexical repetition of *wives*, the same verb tense and same subject referent in the two segments (*they*). Even though the subject of the segments is also the same in (5), this feature is not

relevant to the relation of contrast, and is therefore not counted as a signal in this case. In (7), the concessive relation marked by *even if* is not reinforced by any other signal. Such cases where no relevant signal can be identified (besides the DM) are classified as “without signals”. It is important to note that, in the framework of the present study, the main goal of the signals annotation is to distinguish between signalled and unsignalled cases (i.e. cases where the DM is reinforced or not). As such, exhaustivity in the signals identification may not always be reached: in (6) above, it could be argued that the shared verb tense between the two segments is not a strong signal for the relation of addition compared to the lexical repetition and same subject referent, so that this signal could be omitted. The focus of the present analysis is not on the number of signals that reinforce a particular DM, but rather on the presence or absence of at least one signal. As an indication, Das and Taboada (2018) identify only one or two signals (including DMs) in 88.95% of all signalled relations, which suggests that multiple signals are very rare.

The signals identification was carried out by the author, with regular consistency checks to ensure the stability of the analysis. A random sample of 50 DMs in each genre was independently double-coded by a second expert annotator for inter-rater reliability measures. The two annotators only coded for presence or absence of at least one signal reinforcing the interpretation of the DM, not for the nature of the signal. We report on a 85.3% of agreement on signals identification and a smaller Fleiss' kappa of $\kappa = .696$, penalized by the binary decision. Agreement is lowest in the chat genre (78%), where difficulties mainly concern cases of *and*. Such discordant cases were discussed amongst annotators, and the signalling identification for all *and* tokens in the corpus was once more checked afterwards. Das and Taboada (2018) report on similar scores (between $\kappa = .67$ and $.71$) on a similar small sample of 130 relation tokens, although their coding

scheme includes ten values, instead of two in the present study. We share with the authors the view that “what is more required than arriving at an acceptable measure of agreement is an acceptance of the intrinsic difficulty of annotation, together with a reasonable explanation of how the annotation was performed” (Das and Taboada, 2018, p. 756).

Differences with Das and Taboada’s (2018) method include: some signals are present in both taxonomies but not grouped under the same categories (for instance deictics and demonstratives, or repetition counted as a semantic relation); some features are more precise or more inclusive in one model or the other (for instance “numerical” is restricted in their approach to “same count”, whereas any numeral can be considered in the present approach). Overall, while the general approach is similar in the two projects (i.e. identifying signals that are relevant – but not specific to – the relation at hand), the specific values vary slightly, and the overarching goals are distinct and complementary.

The signalling analysis requires careful examination of the immediate linguistic context of the DM. As subjective as it may be, it is very much in line with the approach in the RST Signalling Corpus (Das et al., 2015) and strives to be as exhaustive and systematic as possible. Limitations of this subjective analysis should nevertheless be borne in mind in the remainder of this paper. In the following sections, results on the functions and compensation of DMs will be discussed and compared across genres and across degrees of DM strength.

Results

As a reminder, this study sets out to test three hypotheses. Firstly, DMs are expected to be more polyfunctional and ambiguous in informal genres (H1). Secondly,

signals reinforcing the interpretation of DMs are expected to be more frequent in formal writing (H2). Thirdly, these signals are also expected to vary with the degree of ambiguity or “strength” of the DM, with more signals compensating for weak DMs (H3). This section first compares the polyfunctionality of DMs across genres (H1). The analysis then focuses on a subset of five relations in order to study the variation of DM strength in genre. The hypotheses about signalling (H2, H3) are then tested in the final subsection.

Are discourse markers more polyfunctional in informal speech?

A total of 12,710 DM tokens have been annotated. Unsurprisingly, *and* (2,495 tokens) is the most frequent and most polyfunctional DM in the sample, that is, it expresses the most different functions. Highly frequent DMs (over 100 tokens) tend to be polyfunctional: among the 15 DMs with over 100 tokens, eight express more than one relation (*and, but, like, I mean, well, then, however, since*), whereas only four out of the remaining 94 types are polyfunctional (*actually, while, as, on the other hand*).

Table 2 shows the proportion of polyfunctional DMs (against monosemous ones) across genres.⁴ It appears that they take up more than half of all DMs in all genres. These proportions are significantly different and go in the expected direction (cf. Hypothesis 1): the proportion of ambiguous DMs decreases with formality and the written modality; in other words, writers tend to use less ambiguous, more specific DMs than speakers, especially in formal texts. This finding brings some quantitative support to a general trend of language, first noted by Zipf (1949), according to which ambiguity and frequency go hand in hand.

Does DM strength vary across genres?

Apart from its basic additive meaning, *and* shows substantial (albeit much less frequent) occurrences as a marker of specification (the second segment introduces more detailed information about the first one), concession (one segment denies the expectation brought about by the other one), consequence (the second segment is the result of the first one) and contrast (the relation points at a difference between at least two aspects in the segments). These uses occur in all three genres with a similar proportion (around 10% for specification, around 3% or less for the other relations).

In the remainder of this paper, we will focus on the five relations expressed by *and* (addition, specification, consequence, concession and contrast) and will compare the different DMs that express them across genres, in order to locate *and* on a continuum from weaker to stronger DMs. This sample amounts to 1,968 annotated DMs, comprising the following 20 types: *and*, *but*, *so*, *however*, *actually*, *while*, *also*, *then*, *though*, *even if*, *although*, *whereas*, *therefore*, *for example*, *plus*, *on the other hand*, *yet*, *thus*, *furthermore* and *moreover*. The five relations all show DM types that do not express any other relation (e.g. *whereas* for contrast): these DMs are considered “strong”. Some are also signalled by DMs that are strongly related to that relation but also express other relations (usually two in total), hence they are more ambiguous (e.g. *but* for contrast): these are “intermediate” DMs. These two categories of DM strength are based on a simple frequency criterion measured by the annotations in the corpus (one relation expressed vs. more than one relation expressed). In addition, *and* will be treated separately as a third category, given its extreme polyfunctionality over five (very different) relations. Therefore, *and* constitutes the lowest point on the scale of DM strength. The resulting scale, broken down for each relation, is the following:

- Addition: *and* < *plus*, *also*, *furthermore*, *moreover*;
- Specification: *and* < *actually* < *for example*;

- Consequence: *and* < *then* < *so*, *therefore*, *thus*;
- Concession: *and* < *actually*, *while*, *however*, *on the other hand*, *but*, *though* < *although*, *even if*, *yet*;
- Contrast: *and* < *though*, *but*, *on the other hand*, *while*, *however* < *whereas*.

The distribution of *and*, intermediate and strong DMs across genres is given in Table 3. Although the effect size is weak (Cramer's $V = .1213$), the differences are statistically significant ($\chi^2 = 57.908$, $df = 4$, $p < .001$) and confirm our hypothesis: strong and intermediate DMs are more frequent in essays than in informal contexts. Pearson's residuals indicate that the distribution of DM strength is intermediate in chat (not very different from the other genres). By contrast, the gap is much larger between discussion and essay, with a decrease of *and* from 45.49% in discussions to 26.26% in essays.

In sum, DM strength is contingent on genre (operationalized as formality and modality) such that stronger DMs are more frequent in essays, while *and* is more frequently used in speech.

When are DMs compensated by signals?

The signalling analysis was carried out on the same sample of 1,968 DMs classified into three categories of DM strength. A total of 1,349 DMs were annotated as co-occurring with other signals in their immediate environment, leaving only 32.55% of the DMs non compensated (Table 4). This result stands in sharp contrast with the data in Das and Taboada (2018, p. 757), who report on a much smaller difference between relations exclusively signalled by DMs (10.65%) and those signalled by both DMs and other signals (7.55%). This could be due to methodological differences in the identification of DMs and/or of signals, as well as to the selection of five relations under scrutiny here.

To test the effect of genre (Hypothesis 2), DM strength and relation type on the presence or absence of signals, we ran a logistic regression model. Forward model selection was carried out with the *step()* function in R: genre did not significantly improve the model, which indicates that the distribution of signals is not affected by genre variation, against our hypothesis. The final model ($R^2 = .261$, $C = .758$) with relation and strength as predictors was run using the *lrm* function from the {rms} package (Harrell, 2015). It returns significant effects for intermediate and strong DMs, which both significantly decrease the likelihood of compensation (intermediate: $\beta = -0.5156$, $SE = 0.17$, $p < .01$; strong: $\beta = -1.1519$, $SE = 0.15$, $p < .001$), which confirms Hypothesis 3. The percentage of null compensation steadily increases from 18.81% with *and* to 31.1% and 56.6% with intermediate and strong DMs.

Relation type is also highly significant: compared to addition as the reference level, DMs expressing consequence are less often compensated ($\beta = -0.9728$, $SE = 0.18$, $p < .001$), whereas DMs expressing contrast ($\beta = 1.5455$, $SE = 0.25$, $p < .001$) and specification ($\beta = 1.3059$, $SE = 0.22$, $p < .001$) are more often compensated; concession does not significantly differ from addition. Das and Taboada (2018) do not report on findings for signalling by DMs alone vs. DMs plus other signals for each coherence relation, so that our results cannot be compared.

The association between ambiguity and compensation confirms our hypothesis (H3) and shows that the more informative the DM, the less it needs to be reinforced by other discourse signals which would be superfluous, in line with Levy and Jaeger's (2007) Uniform Information Density hypothesis.

Types of signals

A total of 142 signal types or signal combinations have been annotated. Besides “no signal” (i.e. DMs alone, 32.5%), the most frequent signal is negative polarity (16.3%), which accompanies many contrastive and concessive relations. The next type is pronouns in co-referential chains (5.8%), which are clearly associated to *and* and additive relations, as are “same subject” (2.8%) and “subject elision” (2.2%), in their capacity as markers of continuity. Other DMs in the close vicinity of the annotated item, either adjacent or further away in the segment, take up 5.7% of all signals and are more specific to intermediate and strong DMs. Syntactic constructions (4.2%) are almost exclusive to *and* and mostly signal specification relations. Alternative lexicalizations (e.g. “by that logic” for consequence or “one of them is” for specification, 3.6%) are, on the contrary, more frequent with intermediate and strong DMs and, therefore, more specific to writing. Signal types and combinations are too numerous to be detailed here. Closer investigation of the association between signal types, relation types and DM strength is beyond the scope of this paper.

Discussion

The present analysis of ambiguity and compensation of DMs across genres shows results that are consistent with the Uniform Information Density hypothesis (Levy & Jaeger, 2007). The association between the degree of DM strength on the one hand and their frequency of compensation or reinforcement by other signals on the other is particularly significant, and prevails over hypotheses of genre variation. In this respect, the study goes beyond previous approaches to discourse signalling by linking it not only to types of coherence relations but to the specific DMs that express them. The relatively large sample size (12,710 DMs overall; 1,968 for the signalling analysis) and high comparability of the corpus components vouches for robust statistical analyses.

Discourse annotation on several written and spoken genres remains rare in the field, and the attention paid to this factor is another innovative aspect of the study.

Nevertheless, it is not without its limitations. From a methodological point of view, the partial subjectivity of the analyses cannot be ignored. Both the annotation of the functions of the DMs and that of the other relevant signals in the context depend on the analyst's interpretation. There are other ways to achieve the same goal which would overcome the issue of subjectivity: Levshina and Degand (2017) and Péry-Woodley et al. (2017) both resort to semi-automatic methods, combining statistical modeling with fine-grained annotation. The latter suggest that "a more satisfactory approach to cuesets would involve attributing weights to individual [...] cues in order to account for the fact that several weak cues may do the same work as one strong cue" (Péry-Woodley et al., 2017, p. 93), which is a very promising avenue indeed.

Weighted approaches to discourse signals would not only address the issue of subjectivity, but they would also be more useful in terms of computational applications: if a "signal" is shared among many coherence relations, it loses its predictive power for that relation. For instance, many signals identified as relevant for addition or specification in this analysis as well as in Das and Taboada's (2018) correspond to basic configurations of sentences, such as tense, subject reference, co-reference, polarity or even text genre, and are therefore often present in relations which have little or nothing to do with the concept of continuity, such as concession or contrast. Similarly, negative polarity, which has been associated with contrastive relations, can also occur in many other relation types, including addition. Automatic discourse parsing (e.g., Marcu, 2000) would indeed benefit from signals other than DMs, as Das and Taboada (2018) have already pointed out, but more efficiently so if these signals are specific, even exclusive to particular relations, excluding *de facto* any other possible interpretation.

Hoek et al. (2019) make a more theoretical contribution in this direction, by proposing to distinguish between three ways in which DMs and other signals interact: “division of labor” (when the DM and the signal are independently self-sufficient and make the presence of the other redundant), “agreement” (when the DM and the signal share some conceptual features) and “general collocation” (when the DM and the signal are *a priori* unrelated but often appear jointly). As the authors put it, “not all non-connective signals for coherence relations are created equal” (2019, p. 27), and this fine-grained view of signal types mirrors and complements the present fine-grained approach to DM strength. Adding Hoek et al.’s (2019) categories to the analysis of discourse signalling would undoubtedly further clarify the picture of a very complex phenomenon.

A final point of discussion relates to implicit relations (i.e. where no DM is used), which were not addressed in this paper. The study has shown that signals tend to compensate for ambiguous DMs. It could logically be expected that signals other than DMs will therefore be more frequent in implicit than explicit relations, since the absence of a DM may generate ambiguity, even compared to the weakly informative *and*. This is suggested by the results in Das and Taboada (2018), who study both “explicit” (with a DM) and “implicit” relations (without one) and show that only 7.55% of all relations are jointly signalled by a DM and other signals, while 74.54% of relations are signalled only by signals other than DMs. However, this result could be due to methodological decisions in the identification of DMs and of signals. Replicating the present analysis to implicit relations would therefore add a fourth category of ambiguity next to *and*, intermediate and strong DMs, namely “implicit relations”, and refine our understanding of discourse signalling. Experimentally testing the effect of

these categories on processing would then establish whether and how combinations of signals impact discourse comprehension.

Conclusions

The present study set out to analyze the combination of discourse markers (such as *but* or *moreover*) with other discourse signals (such as semantic relations or punctuation marks) across three genres that differ in their modality (spoken vs. written) and degree of formality (informal vs. formal). On the basis of a corpus study where the functions of 12,710 DMs have been annotated, three categories of DM strength have been established, based on the polyfunctionality of the DMs: weak (*and*), intermediate and strong. Statistical modeling of the association between DM ambiguity, presence of compensating signals, and genre, on a sample of 1,968 DMs where other signals have been identified, reveals that *and* is indeed more often compensated than intermediate and strong DMs, and that this compensation is not affected by genre variation. Formality does explain the distribution of polyfunctional DMs in the corpus, but their combination (or not) with other discourse signals seems more dictated by the strength of the DM and by the relation it expresses, with specification and contrast particularly prone to reinforced signalling, as opposed to consequence relations.

This study has a number of implications for discourse analysis and psycholinguistics. First of all, it confirms that the low informative value of some discourse markers does tend to be balanced by other elements in the context, supporting a reconciling, Gricean view of production and comprehension processes (e.g. Rational Speech Act theory, Frank & Goodman, 2012; Yung, Duh, Komura, & Matsumoto, 2017). Experimental studies would be useful to confirm the role of DMs and signals as triggers of comprehension processes (cf. Kleijn, Pander Maat, & Sanders,

2019, for a recent comprehension study). We also showed that genre variation, while influential in the polyfunctionality and strength of DMs, has no significant effect on signals, which restricts the impact of planning pressure and formality conventions to DMs themselves, and not their linguistic environment. Further research might reveal more genre differences between specific discourse signals, as our qualitative analysis of signal types tentatively suggested.

The most important implication, however, concerns the continuum from implicit to explicit discourse marking. Das and Taboada (2018) already questioned the traditional understanding of “implicit” relations by showing that, while DMs are only present in a small percentage of coherence relations, these are very often signalled by other textual devices. The present analysis has tried to show that not only the implicit-explicit dichotomy should be refined, but even within the “explicit” category, distinctions should be drawn between different categories of DMs on the basis of their ambiguity and informativeness (see also, Fabricius-Hansen, 2005). In other words, if implicitness should be redefined by taking into account other discourse signals besides DMs, explicitness, in turn, can be refined by teasing out high-information (strong, non compensated) and low-information (weak, compensated) DMs. Seminal processing studies such as Murray (1997) or Millis and Just (1994) could be extended by incorporating such refined distinctions of DM strength, taking the present results as hypotheses for the role of DMs and signals on comprehension.

Several areas remain open for further investigation, such as the role of text type (e.g. argumentative, narrative, expository), crosslinguistic variation, individual preferences, a more fine-grained and statistical approach to specific signal types, and experimental testing of all these factors. Despite the significant predictors identified in this study, they cannot entirely explain why speakers and writers sometimes combine

strong DMs with other discourse signals, making one or the other redundant. As

Fabricius-Hansen (2005) puts it:

Although the connective may not be strictly necessary in order to arrive at the interpretation triggered by its presence, it may be preferable because it eliminates ambiguity, reduces undeterminacy, or prevents incoherence – or garden path effects – by blocking an interpretation that would otherwise be natural at the local (sentence) level but lead to incoherence at the (global) level of discourse interpretation. (p. 43)

Accounting for all these possibilities seems like a daunting challenge, and it is one definite future direction of the present study to design controlled experiments that would test for the effect of these various contexts of use, one small piece of the puzzle at a time.

References

- Asr, F., & Demberg, V. (2012). Measuring the strength of linguistic cues for discourse relations. In E. Hajičová, L. Poláková, & J. Mírovský (Eds.), *Proceedings of the COLING Workshop on Advances in Discourse Analysis and its Computational Aspects (ADACA)* (pp. 33–42). Mumbai, India: The COLING 2012 Organizing Committee. Retrieved from <https://www.aclweb.org/anthology/W12-4703>
- Cain, K., & Nash, H. (2011). The influence of connectives on young readers' processing and comprehension of text. *Journal of Educational Psychology, 103*, 429–441. doi:10.1037/a0022824
- Canestrelli, A., Mak, P., & Sanders, T. (2016). The influence of genre on the processing of objective and subjective causal relations: Evidence from eye-tracking. In N. Stukker, W. Spooren, & G. Steen (Eds.), *Genre in language, discourse and cognition* (pp. 51–74). Berlin: De Gruyter. doi:10.1515/9783110469639-004
- Chafe, W. (1994). *Discourse, consciousness, and time*. Chicago: University of Chicago Press. doi:10.1017/S0142716400010523
- Clark, H. (2002). Speaking in time. *Speech Communication, 36*, 5–13. doi:10.1016/S0167-6393(01)00022-X
- Crible, L. (2017). Discourse markers and (dis)fluency in English and French. Variation and combination in the DisFrEn corpus. *International Journal of Corpus Linguistics, 22*, 242–269. doi:10.1075/ijcl.22.2.04cri
- Crible, L., & Degand, L. (2019). Domains and functions: a two-dimensional account of discourse markers. *Discours, 24*. doi:10.4000/discours.9997
- Cuenca, M. J., & Crible, L. (2019). Co-occurrence of discourse markers in English : From juxtaposition to composition. *Journal of Pragmatics, 140*, 171–184. doi:10.1016/j.pragma.2018.12.001

- Das, D., & Taboada, M. (2018). Signalling of coherence relations in discourse, beyond discourse markers. *Discourse Processes*, *55*, 743–770.
doi:10.1080/0163853X.2017.1379327
- Das, D., Taboada, M., & McFetridge, P. (2015). RST signalling corpus, LDC2015T10.
Retrieved from <https://catalog.ldc.upenn.edu/LDC2015T10>
- Degand, L., & van Bergen, G. (2018). Discourse markers as turn-transition devices: Evidence from speech and instant messaging. *Discourse Processes*, *55*, 47–71.
doi:10.1080/0163853X.2016.1198136
- Fabricius-Hansen, C. (2005). Elusive connectives. A case study on the explicitness dimension of discourse coherence. *Linguistics*, *43*, 17–48.
doi:10.1515/ling.2005.43.1.17
- Frank, A., & Goodman, N. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998. doi:10.1126/science.1218633
- Goldstein-Stewart, J., Goodwin, K. A., Sabin, R. E., & Winder, R. K. (2008). Creating and using a correlated corpora to glean communicative commonalities. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008*. Marrakech, Morocco: European Language Resources Association. Retrieved from http://www.lrec-conf.org/proceedings/lrec2008/pdf/771_paper.pdf
- Harrell, F. (2015). rms: Regression Modeling Strategies. R package version 4.3-1.
Retrieved from <https://hbiostat.org/R/rms>
- Hoek, J., Zufferey, S., Evers-Vermeul, J., & Sanders, T. J. M. (2017). Cognitive complexity and the linguistic marking of coherence relations: A parallel corpus study. *Journal of Pragmatics*, *121*, 113–131. doi:10.1016/j.pragma.2017.10.010

- Hoek, J., Zufferey, S., Evers-Vermeul, J., & Sanders, T. J. M. (2019). The linguistic marking of coherence relations: Interactions between connectives and segment-internal elements. *Pragmatics & Cognition*, *25*, 275–309.
doi:10.1075/pc.18016.hoe
- Kleijn, S., Pander Maat, H.L.W., & Sanders, T.J.M. (2019). Comprehension effects of connectives across texts, readers, and coherence relations. *Discourse Processes*, *56*, 447–464. doi:10.1080/0163853X.2019.1605257
- Knott, A., & Dale, R. (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, *18*, 35–62. doi:10.1080/01638539409544883
- Kunz, K., & Laphinova-Koltunski, E. (2015). Cross-linguistic analysis of discourse variation across registers. *Nordic Journal of English Studies*, *14*, 258–288.
doi:10.35360/njes.347
- Levshina, N., & Degand, L. (2017). Just *because*: In search of objective criteria of subjectivity expressed by causal connectives. *Dialogue & Discourse*, *8*(1), 132–150. doi:10.5087/dad.2017.105
- Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems (NIPS)* (pp. 849–856). Cambridge, MA: MIT Press. doi:10.5555/2976456.2976563
- Louwerse, M. M., & Mitchell, H. H. (2003). Toward a taxonomy of a set of discourse markers in dialogue: a theoretical and computational linguistic account. *Discourse Processes*, *35*, 243–281. doi:10.1207/S15326950DP3503_1
- 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, 557–576.

doi:10.1080/0163853X.2013.841075

Marcu, D. (2000). The rhetorical parsing of unrestricted texts: A surface based approach. *Computational Linguistics*, 26, 395–448. Retrieved from <https://www.aclweb.org/anthology/J00-3005>

Millis, K. K., & Just, M. A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*, 33, 128–147.
doi:10.1006/jmla.1994.1007

Murray, J. (1997). Connectives and narrative text: The role of continuity. *Memory & Cognition*, 25, 227–236. doi:10.3758/BF03201114

Péry-Woodley, M.-P., Ho-Dac, L.-M., Rebeyrolle, J., Tanguy, L., & Fabre, C. (2017). A corpus-driven approach to discourse organisation: from cues to complex markers. *Dialogue & Discourse*, 8(1), 66–105. doi:10.5087/dad.2017.103

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). The Penn Discourse Treebank 2.0. Paper presented at the 6th International Conference on Language Resources and Evaluation (LREC 2008), Marrakech, Morocco. Retrieved from <http://www.lrec-conf.org/proceedings/lrec2008/>

Prasad, R., Webber, B., & Lee, A. (2018). Discourse annotation in the PDTB: The next generation. In H. Bunt (Ed.), *Proceedings of the 14th Joint ACL-ISO Workshop on Interoperable Semantic Annotation* (pp. 87–97). Santa Fe, NM: Association for Computational Linguistics. Retrieved from <https://www.aclweb.org/anthology/W18-4710>

- Sanders, T., & Noordman, L. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes, 29*, 37–60.
doi:10.1207/S15326950dp2901_3
- Schiffrin, D. (1987). *Discourse markers*. Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511611841
- Scholman, M., Rohde, H., & Demberg, V. (2017). “On the one hand” as a cue to anticipate upcoming discourse structure. *Journal of Memory and Language, 97*, 47–60. doi:10.1016/j.jml.2017.07.010
- Spooren, W. (1997). The processing of underspecified coherence relations. *Discourse Processes, 24*, 149–168. doi:10.1080/01638539709545010
- Spooren, W., & Degand, L. (2010). Coding coherence relations: Reliability and validity. *Corpus Linguistics and Linguistic Theory, 6*, 241–266.
doi:10.1515/cllt.2010.009
- Taboada, M. (2006). Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics, 38*, 567–592. doi:10.1016/j.pragma.2005.09.010
- Webber, B., Prasad, R., & Lee, A. (2019a). Ambiguity in explicit discourse connectives. In S. Dobnik, S. Chatzikiyriakidis, & V. Demberg (Eds.), *Proceedings of the 13th International Conference on Computational Semantics (IWCS 2019) – Long papers* (pp. 134–141). Gothenburg, Sweden: Association for Computational Linguistics. doi:10.18653/v1/W19-0411
- Webber, B., Prasad, R., & Lee, A. (2019b). The Penn Discourse Treebank 3.0 Annotation Manual. Retrieved from <https://catalog.ldc.upenn.edu/docs/LDC2019T05/>

Yung, F., Duh, K., Komura, T., & Matsumoto, Y. (2017). A psycholinguistic model for the marking of discourse relations. *Dialogue & Discourse*, 8(1), 106–131. doi: 10.5087/dad.2017.104

Zipf, G. K. (1949). *Human behaviour and the principle of least effort: An introduction to human ecology*. Cambridge, Massachusetts: Addison-Wesley.

Zufferey, S., & Gygax, P. (2017). Processing connectives with a complex form-function mapping in L2: The case of French “en effet”. *Frontiers in Psychology*, 8. doi: 10.3389/fpsyg.2017.01198

Zufferey, S., & Gygax, P. (2020). “Roger broke his tooth. *However*, he went to the dentist”: Why some readers struggle to evaluate wrong (and right) uses of connectives. *Discourse Processes*, 57, 184–200.
doi:10.1080/0163853X.2019.1607446

Zufferey, S., Mak, P., Degand, L., & Sanders, T. (2015). Advanced learners’ comprehension of discourse connectives: The role of L1 transfer across on-line and off-line tasks. *Second Language Research*, 31, 389–411.
doi:10.1177/0267658315573349

Footnotes

¹The term “discourse marker” will be used throughout this paper in order to be consistent with previous publications and with the bulk of the literature, even though it mostly refers to what other authors call “connectives.” Operational definitions are provided in the methodology.

²I would like to thank Prof. Liesbeth Degand and Fang Yang for their precious help with the double coding of the data.

³The annotations on the 1,968 DM sample, as well as the R script used to conduct the analysis, are available as supplementary materials and on <https://osf.io/6abyn/>.

⁴This is excluding a small number of very rare DMs (20 tokens or fewer) that showed one or two tokens in more than one function. The full table of DM types and their assigned functions can be found at <https://osf.io/6abyn/>.

Table 1

Taxonomy of Discourse Signals in the Context of DMs

Signal group	Signal type	Values
Co-occurring	Adjacent DM	<i>the DM token</i>
	Other DM in the	<i>the DM token</i>
	Adjacent pause	silent or filled pause
	Punctuation	dashes, parentheses, commas
	Disfluency	interruption, repetition, repair
	Response particle	yes, no
Sentence	Mood	declarative, interrogative, imperative, exclamative
	Polarity	positive, negative + combinations
	Polarity difference	same, different
	Verb tense	present, past, future, conditional + combinations
	Tense difference	same, different
	Subject referent	same, different
	Unit type	full, relative, completive, non-finite, elision,
Syntax	Construction	parallelism, SV inversion, cleft, presentational, imperative + <i>and</i> , dislocation
Semantics	Semantic relation	synonymy, antonymy, metonymy, hyperonymy, hyponymy, equivalence, comparison
	AltLex	expression encoding the meaning of the relation
	Evaluative	expression of stance
	Epistemic language	expression of reasoning
	Speech-act	expression of speaking
	Deictics	time and place references
	Proper nouns	names referring to places, groups or persons
	Numerals	the unit contains numbers or cardinals
	Demonstratives	including possessives
	Pronouns	referential chain between the two segments
Repetition	exact lexical repetition in the two segments	

Table 2

Proportion of Polyfunctional DMs Across Genres

	Discussion	Chat	Essay
Tokens	4028	2682	1221
% of all DMs	65.51	60.58	57.22
<i>z</i>	5.187	2.602	-6.849
<i>p</i>	<.001	<.01	<.001

Note. The *z* score provides a measure of statistical significance for the difference

between two proportions. The *z* scores in each column correspond to the difference with the column to the right (discussion-chat, chat-essay, essay-discussion).

Table 3

Frequency of DM Strength Categories Across Genres

	Discussion		Chat		Essay	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
<i>and</i>	313	45.49	254	38.66	172	26.26
intermediate	232	33.72	253	38.51	306	46.72
strong	143	20.78	150	22.83	177	27.02
Total	688	100	657	100	655	100

Table 4

Presence of Discourse Signals Across Degrees of DM Strength

	With signals		Without signals		Total	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
<i>and</i>	600	81.19	139	18.81	739	100
intermediate	526	69.12	235	30.88	761	100
strong	202	43.16	266	56.84	468	100
Total	1328	67.48	640	32.52	1968	100