Co-occurrence of discourse markers in English: from juxtaposition to composition

Abstract (max 200 words)

In this paper, we report on a qualitative analysis of co-occurring discourse markers, that is, sequences of adjacent discourse markers that belong to the same unit but may express different function(s). We examine several formal and functional features of these co-occurring strings on the basis of authentic corpus examples extracted from conversational data in English. In particular, we focus on scope, meaning-in-context (or functions), syntactic category and position. Our analysis reveals several degrees of integration: differences in scope allow us to differentiate juxtaposition and combination of markers. In the case of combination, difference in meaning integration allows us to distinguish addition from composition of markers. We pay particular attention to in-between and ambiguous cases such as *and so* or *and then*, which instantiate different degrees in our cline of co-occurrence depending on the meaning that can be interpreted from the cluster. We finally discuss the implications of such fine-grained distinctions for the perspective of systematic corpus annotation.

Key-words : discourse markers ; DM co-occurrence ; pragmatic functions ; scope ; corpus ; English

1. Introduction

Among the vast literature on discourse markers (henceforth DMs) and discourse-relational devices in general, one aspect of their behaviour has been somewhat overlooked until recently, namely their co-occurrence. It is frequently the case that two or more DMs co-occur, as in the case of *and if, but when* or *so for instance if*, where DMs are juxtaposed, or in the case of *but actually*, *and so, and then, and therefore, and in fact* or *but anyway*, where they combine. Discourse analysis and corpus annotation show that co-occurrence is a relatively frequent phenomenon.

In Cuenca & Marín's (2009) corpus of spoken Catalan and Spanish, co-occurring markers account for 16.48% (149 out of 904 instances of DMs) and 17.25% (156 out of 905 instances of DMs) of all the identified markers, respectively. In a similar vein, in Crible's (2018) spoken corpus study, 15.77% (670 cases / 4249 total) of all occurrences in English and 23.85% (1072 cases / 4494 total) in French are coded as part of a co-occurring string. Although the criteria for including or excluding certain strings may vary among studies and languages, the proportions point to a phenomenon that cannot be ignored.

DM co-occurrence is a multi-faceted phenomenon, since not all cases display the same degree of integration. The authors that have accounted for this phenomenon distinguish different types of co-occurrence depending on a number of syntactic and functional criteria (see, e.g., Luscher 1993; Hansen 1998; Pons 2008, in print; Cuenca & Marín 2009). However, the concept, types and criteria are still to be explored and many key-questions that arise when annotating a corpus, especially with spoken discourse, remain unclear. Specifically, DM co-occurrence poses a challenge for corpus annotation since (i) it is not always clear whether two co-occurring DMs remain independent from each other or whether they should be considered as one token, and (ii) senses can be influenced by co-occurring DMs during disambiguation. This study sets out to provide criteria for different degrees of co-occurrence on the basis of corpus examples.

The purpose of this study is to revisit Cuenca & Marín's (2009) three-fold classification (namely, juxtaposition, addition and composition of DMs), and refine the criteria to distinguish each degree of co-occurrence, in order to be able to apply them systematically to corpus data. Specifically, we propose to differentiate juxtaposition and combination of discourse markers based on scope: Juxtaposed DMs take scope on different units, whereas combined DMs take scope over the same discourse unit. In the case of combined DMs, the degree of integration can be different. Added DMs combine but keep their individual meaning. Compound DMs include DMs that can occur independently but, when combined, they jointly act as a single marker and their individual meaning cannot be disentangled. In the case of a compound DM, as opposed to added DMs, no prosodic boundary can

exist between the DMs, and substituting DM2 by a synonym or near-synonym is not possible or implies substantial changes in the meaning and function of the whole.

The proposal relies as much as possible on objective features such as syntactic scope, grammatical category and prosody. In doing so, we will distinguish criterial features that always apply to a specific degree of co-occurrence (e.g. juxtaposed DMs have different scopes) from typical features which do not always apply (e.g. juxtaposed DMs typically correspond to a conjunction and express propositional meanings). We will also show that some clusters (e.g. *and then*) resist systematic classification into one of the three degrees of co-occurrence, hence advocating for a flexible, context-bound approach to the most ambiguous markers.

The analysis is applied to English, in contrast with many previous corpus analyses specifically devoted to the topic, which mostly refer to Romance languages (French, Spanish and Catalan).[1] To this end, the materials and method are described in the next section.

2. Corpus and method

For this study, we used a sample of English conversational data from the *DisFrEn* dataset where DMs were already identified (Crible 2017): 109 DM clusters were thus extracted, from a total of 34,534 words (about 3 hours of recordings).[2] Half of the transcripts correspond to private conversations between friends or relatives (sampled from the British component of the International Corpus of English, Nelson et al. 2002). The other half consists of face-to-face interviews sampled from the

---

[1] For previous analyses based on English corpora, see Oates (2000, 2001), Koops & Lohmann (2015), Lohmann & Koops (2016) and Tagliamonte (2016).

[2] The annotations in the *DisFrEn* dataset include more types of discourse markers than what is included in the present study. The 109 extracted clusters were selected based on criteria developed in Section 4.

Backbone corpus (Kohn 2012): this setting is less interactive as the interviewee primarily holds the floor while they talk about their profession or place of living.

We considered multi-word markers as one DM (e.g. *so that, even if, I mean*) and restricted the analysis to connective DMs (that is, three-slot constructions, i.e. two content units and a marker). This criterion includes conjunctions (*and, but, because, when*, etc.), other connectives (*actually, however*, etc.) and other expressions such as *well* or *I mean*, provided that they connect two linguistically expressed units. It excludes, however, markers which only affect one content unit without linking it to another one, often with a modal or epistemic meaning, such as *you know* or *I think* (see Cuenca 2013).

For each cluster, we manually encoded four features, namely number of elements in the cluster, syntactic category of each DM, scope (same or different), position (turn- initial, utterance-initial or utterance-medial). We then discussed whether the elements of the cluster expressed the same meaning (or function) or not, and then decided on the degree of integration of the adjacent DMs.

As for the syntactic category of the DM, assuming Cuenca's (2006, 2013) proposal, we differentiate three classes, namely, conjunctions, parenthetical connectives and pragmatic connectives.

  (i)   Conjunctions (Cj) are linking words that indicate grammatical relationships (subordination/coordination) and can express various discourse relations such as addition, disjunction, contrast, condition or purpose. Conjunctions typically introduce clauses in compound sentences but some of them can also connect at text level. The conjunctions that were involved in a co-occurrence in our corpus are: *although, and, as, because, but, even if, if, or, so[3], when, whereas*.

  (ii)   Parenthetical connectives (ParentC) are appositional syntactically detached items indicating basic logico-argumentative meanings, namely,

---

[3] As a discourse marker, *so* can function either as a parenthetical connective or as a conjunction. As a parenthetical connective, it can combine with *and,* whereas this combination is not possible when *so* is a conjunction. This dual syntactic category clearly relates to the polyfunctionality of *so* (see also Koops & Lohmann 2016: 442).

addition, disjunction, contrast and consequence. They can combine with conjunctions preceding them and connect both at sentence and at text level. The parenthetical connectives that were involved in a co-occurrence in our corpus are: *actually, anyway, for example, for instance, in fact, nevertheless, otherwise*, *secondly, so, then, therefore.*

(iii) Pragmatic connectives (PragC) are appositional syntactically detached items that combine frame and modal meanings. They are mainly used in oral texts, bracketing units of talk such as interventions, turns or units within turns while also indicating an interpersonal meaning. The pragmatic connectives that were involved in a co-occurrence in our corpus are: *well, I mean* and *now*.[4]

Directly related to pragmatic connectives, the corpus also includes *okay* (used in our corpus once preceding *so*), which typically acts as an interjection but can also introduce units in discourse, especially in dialogue.

Following Cuenca's proposal (2013), based on Halliday's (1970) functions of language (i.e., ideational, textual and interpersonal), we differentiate three types of general discourse marking meanings, corresponding to three domains where DMs can occur, namely, propositional, structural and modal.

• Propositional meanings arise when connection takes place at the content-level, that is, when the marker relates ideas. Propositional markers link discourse units and indicate relationships such as addition, disjunction, contrast, concession, condition, cause or consequence (or variants of these general meanings) between their contents. These meanings are typically expressed by conjunctions and parenthetical connectives.

• Structural meanings are the result of a marker bracketing a unit of talk such as the text, a sequence, a turn or a unit within a turn. A structural marker indicates meanings such as start, closing, pre-closing, continuity, topic

---

[4] The corpus also includes a case of *listen* (*but listen when*), but its use is not clearly connective and so it has been excluded.

change or reformulation. Some conjunctions, parenthetical connectives and pragmatic connectives can fulfil these functions.

- Modal meanings are inherently interpersonal in that they put forward the attitude, knowledge or stance of the speaker with respect to what is being said or to the hearer. Pragmatic connectives are typical vehicles of modal meanings such as agreement, disagreement or emphasis combined with connective functions related to discourse organization (e.g. turn-taking, topic change, topic retrieval and so on).

The analysis of language-in-use examples will highlight which aspects (scope, function/meaning, syntactic categories, prosody) are criterial in the definition of the degree of integration and which ones only point to tendencies.

3. Previous analyses of DM co-occurrence

Analyses of the co-occurrences of connectives and DMs in general are relatively recent but there are some outstanding previous contributions that should be taken into account. The co-occurrence of DMs has been identified at least since Dik's (1968: Chapter 6, Section 4.1) monograph on coordination. Dik (1968) proposes a test to differentiate coordination and subordination consisting in the possibility of co-occurrence of two subordinators (e.g. *although when*) versus the impossibility of combining two coordinators (which accounts for *and, nor, or, but* and *for*). Dik (1968) also observes that conjunctive adverbs such as *yet, still* or *nevertheless* can follow *and*.

Franchini (1986) identifies the latter elements as *matizadores* (roughly, 'shaders' or 'nuancers'), that is, connectives that combine with conjunctions (especially *and* and *but*) and specify their meaning. Franchini (1986: 196) reformulates Dik's test and proposes two structures:

   a) M_1 *y* x M_2
   b) M_1 x M_2 *y* x M_3

where
x = the element to be identified in M_1 x M_2

y = neutral coordinator *and*

M = discourse units

Considering the previous structures, Franchini formulates the following principles:

   (1) Apply structure (a). If x accepts it, x is a *matizador*

   (2) Apply structure (b). If x accepts it, x is a subordinator

   Otherwise,

   (3) *x* is a coordinator

This test allows to differentiate subordinators and coordinators, but also other markers that can combine with the general coordinator *and*.

On the lines of Franchini, Cuenca (1990, 2001, 2002, 2006: chap. 3, 2013) identifies the group of connectives that can combine with conjunctions, defines the category as related but different from both conjunctions and adverbials, and characterizes it for Catalan (1990) and Spanish (Cuenca 2001). Cuenca (1990) used the term *matizador/matisador* but in later papers she adopts the term *parenthetical connective*, following Rouchota (1998), who defines the class for English.[5]

Papers specifically dealing with co-occurrences propose several criteria to distinguish different degrees of integration, namely difference in meaning, in compositionality and in scope and position. Luscher (1993) uses syntactic and semantic scope to distinguish between "additive" and "compositional" sequences. He defines the latter as applying to two adjacent DMs which are semantically similar (e.g. French *mais pourtant* 'but however'), one of them being more restricted or specific in its meaning than the other. A similar observation is made by Oates (2001, 2002), who argues that combined DMs tend to follow the pattern: 'weak marker + strong marker', where weak stands for more polyfunctional/ambiguous

---

[5] The term *matizador* is borrowed by Franchini from Barrenechea & Manacorda (1971). It metaphorically refers to the idea that connective items such as *therefore, thus, in contrast, in addition* etc. add a *shade* to the basic meaning ('colour') of the conjunction they follow.

and strong means that it "can only cue a single relation" (Oates, 2001: 43). Hansen's (1998) distinction between summative and combinatory sequences adopts a different perspective and depends on whether the elements in the sequence retain their individual meaning (French *ah bon* 'oh really') or form a new complex one (*eh bien* 'well'). Hansen argues that DM sequences are summative when it is possible to reconstruct the meaning of each element. Pons (2008) analyses the co-occurrences of the Spanish modal marker *bueno* with other discourse markers and concludes that oral discourse segmentation allows to differentiate two different configurations: the cases in which two markers are simply adjacent from the cases in which they combine, according to whether they apply to different or to a unique structural unit. In a more recent paper, Pons (in print) describes the co-occurrence of DMs, including connective, modal and interpersonal markers, in conversational Spanish. He differentiates three types of relationships, namely, adjacency (unrelated DMs, even belonging to different non-syntactically related units), combination (related DMs, e.g. *y luego/entonces, bueno pero/pues*) and lexicalization (conversational idioms, e.g. *pues bueno*). His analysis focuses on combinations from the point of view of discourse units and the positions that DMs can adopt within a discourse unit.

Fraser (2013) examines the sequencing of two contrastive markers in English, as in the case of *but, on the other hand* or *but instead*. Fraser identifies 15 markers that can combine (namely, *alternatively, but, contrary to expectations, conversely, despite, however, in comparison, in contrast, nevertheless, on the contrary, on the other hand, rather, regardless of that, instead, yet*) and investigates the conditions under which a co-occurrence of two contrastive DMs is acceptable (i.e., ability to stand alone and compatibility of the meanings of the two DMs). In a later study, Fraser (2015) extends this line of reasoning to inferential or "implicative" DMs (such as *so*) and examines combinations of DMs across semantic classes.
Dostie (2013) and Crible (2015) consider other types of cues in DM use that provide evidence for stronger degrees of combination, such as phonological reduction (*eh bien* to *eh ben*), new spellings (*ou sinon* 'or else' to *aussi non*) and new contexts of use (initial to final position for *ou sinon*).

Koops and Lohmann (Koops & Lohmann 2015, Lohmann & Koops 2016) provide an interesting analysis of sequences of discourse markers. Their research on a corpus of American English telephone conversations addresses several key-questions: (i) Which DMs tend to combine? (ii) Are there any constraints in combinations? (iii) What determines the order of co-occurring DMs? (iv) Which are the patterns, if any, that determine the linearization of DMs? Koops and Lohmann identify all clausal-initial sequences of DMs and conclude that "the order in collocating DM sequences is highly systematic, except for collocates involving the DM *you know*" (Lohmann & Koops 2016: 440). They identify the most frequent sequences including one of the eleven DMs analysed in Schiffrin (1987), namely, *oh, well, and, but, or, so, because, now, then, you know* and *I mean.* Their frequency analysis shows that the top of the rank corresponds to *and then* and *but then,* and that many combinations include *you know*, either as DM1 or as DM2; in addition, some markers tend to specialize in the first position (most conjunctions, *well* and *oh*), whereas some others tend to have second position (*then, now, I mean*). Lohmann & Koops (2015, 2016) conclude that the ordering tendencies can be accounted for by considering (i) two separate grammaticalisation processes for DMs that have a sentence-level origin, by which some elements tend to precede others, whereas others do not have to, and (ii) the persistence of ordering constraints from their sources. Lohmann & Koops (2016) also take into account functional factors that explain the order of co-occurring discourse markers and point out that, although "DM's functional orientation in discourse seems promising" (2016: 437), no clear conclusions can be drawn.

Let us finally turn to Cuenca & Marín (2009), which is the starting point of our proposal here. Cuenca & Marín (2009) discuss and illustrate a three-fold distinction in a corpus of Spanish and Catalan oral narratives obtained through a semi-structured interview protocol. By considering several features in a prototype approach (namely, grammatical pattern, meaning integration, position, domain), they differentiate three types of co-occurrences: juxtaposition, addition and composition. The three types can be related to the categories of the DMs and the type of function or domain that they typically express:

- *juxtaposition*, when the DMs do not combine syntactically nor semantically;
- *addition*, when the DMs combine locally but their functions remain distinct;
- *composition*, when the DMs function as a single complex unit and jointly contribute to indicating a discourse function at a global level. [6]

Cuenca & Marín's (2009) fine-grained analysis allows to identify recurrent formal and functional tendencies for each of these types of co-occurrences, which are summarised in Table 1 (Cuenca & Marín 2009: 911, Table 5; adapted).

Table 1. Types of co-occurrences according to Cuenca & Marín (2009)

	Dominant grammatical pattern(s)	Dominant position(s)	Dominant domain(s)
Juxtaposition e.g. Sp. *y mientras* 'and meanwhile'	Cj Cj	Act internal	Propositional
Addition e.g. Cat. *perquè a més* ('because, in addition') Cat. *i bueno* ('and well')	Cj ParentC Cj PragC	Turn/act internal Minor transition places	Propositional Structural
Composition e.g. Sp. *pues vale* ('then OK') Cat. *clar, a veure* ('of course, well')	ParentC PragC PragC PragC	Beginning End Major transition places	Structural-Modal

According to Cuenca & Marín's (2009) analysis, juxtaposition typically involves two conjunctions acting at the ideational or propositional domain; *addition* usually involves a conjunction followed by a parenthetical connective that connect at a

---

[6] The term *composition* refers to the lexical process by which two items integrate so that they act as a single unit of the same kind. This process should not be mistaken for *compositionality*. In fact, from a semantic point of view, the process generally implies that the sequence is not interpreted compositionally anymore, that is, the meaning of the whole is not the sum of the meaning of its parts.

local level, take scope over the same unit and express propositional and structural functions; finally, *composition* is the result of a combination of parenthetical and/or pragmatic connectives that express a single structural-modal function.

Crible (2018) attempted to apply Cuenca & Marín's (2009) classification through systematic annotation and was confronted with problematic, borderline cases (e.g. *and so* or French *et alors* 'and then') which raised concerns about some features, such as the difference in function and the grammatical category, pointing especially to the fuzzy border between addition and composition. Crible also discusses the role of frequency in the definition of these levels, and suggests an additional degree to deal with cases of "reinforcement" (e.g. *but in fact*). Her study draws the attention to the advantages of an adequate treatment of DM co-occurrence for corpus annotation (token identification and sense disambiguation). Similarly, in the guidelines of the Penn Discourse TreeBank 2.0, Prasad et al. (2007) mention that multiple (i.e. co-occurring) connectives should ideally be annotated as such and differentiated according to the (in)dependence of their elements in order to improve predictive features and classifiers.[7]

4. Clusters of discourse markers in the corpus

Two or more DMs co-occur when they are contiguous and have partial or total scope over a discourse unit (Q) that they connectively relate to a previous segment or discourse unit (P): "P DM1 DM2 (DMx) Q".

(1)   you switch the wheels off (0.370) and you turn the propeller on (0.230) and then it behaves as if it was exactly like a (0.420) boat a little bit slow **but nevertheless** it's still a boat working on the water (INT 20)

This definition excludes:

---

[7] The PDTB 2.0 distinguishes between "multiple" and "conjoined" connectives, the latter referring to a very restricted number of uses such as *if and when*, which are annotated as one item.

(i) two consecutive DMs, one linking backwards (final position) and the other linking forward (initial position) (Q DM — DM P'), as in example (2),

(2)   it becomes (0.100) more intensive (0.310) particularly in the summer **actually when** it's easier to make trips out (INT 26)

(ii) markers separated by longer pauses (3),

(3)   they call me the boss so anything that goes wrong is my fault (0.250) but we do have an activities organiser who deals with all of that (0.020) in fact (0.060) we have (0.310) two people: one working at the practical level of making the things happen (0.400) and one who produces the information that you 've seen on the website // **okay (0.380) and (0.100) so** what is your (0.020) um role here as manager what sort of things do (0.020) you take care of here is it the day-to-day running (INT 28)[8]

(iii) cases of re-starts or repetition of markers due to performance effects, as in the case of *so* in (4).

(4)   man you like involved interested sense of humour warm open intelligent (0.250) central quality involved prepared to take an interest in (2.640) **so (0.633) so anyway so** all I was saying originally (2.480) is that it's very interesting (CONV 22)

In (2), *actually* is a connective in final position (the first segment could be expressed as *actually it becomes more intensive particularly in the summer*) and *when* introduces a time clause modifying the main clause. In (3), the pause following *okay* suggests that it does not co-occur with *and*, which does not co-occur with the following marker, *so,* either. Finally, in (4) *so* is repeated three times but only *so anyway* is considered as a co-occurrence.[9]

We have also excluded cases in which one element is a pragmatic marker with no connective function. This includes a number of co-occurrences with the phatic

---

[8] Double backslashes (//) in the transcription of examples correspond to speaker changes (i.e., a new turn).

[9] In a similar vein, Lohmann & Koops (2016: Section 2.2) exclude repeated DMs as a reflect of a false start or self-repair, markers that are not strictly adjacent and also DMs that are "temporally adjacent but which the speaker most likely did not intend to utter together" (2016: 424).

marker *you know*, which is frequently used as a filler or a transition marker inside an utterance and tends to cluster with other DMs. There are 9 sequences in the corpus including *you know*, which could be considered as co-occurrences: *but you know* (3 examples), *but you know if* (1)*, and you know* (1) *and so you know* (1), *I mean you know if* (1), *well you know* (1), *well I mean you know so* (1).[10] Let us consider example (5):

> (5) at the end of the day she got a very indifferent degree which I can comfort myself with (1.090) <laughing/> I mean // she's happy // *yes* (0.200) **well I mean you know so** (0.200) that's what I'm trying to say that you know all these things that Linda sets such great store by at the end of the day (0.600) don't add up to a row of beans (CONV 68)

The speaker in (5) is trying to elaborate an idea and starts with an affirmative marker (*yes*) and then, after a pause, tries to reassume the argument by using four more markers: two pragmatic connectives (*well I mean*) that reorient the message, the phatic marker *you know* and the continuative *so*. The example in (5) could be considered a co-occurrence of four (or even five) DMs. However, if we restrict the analysis to connective markers and exclude *you know,* only *well I mean* counts as a co-occurrence*.*

The corpus analysed includes 109 DMs clusters, most of them corresponding to 2 DMs (13 different DM1s combined with 18 different DM2s, 103 examples), and some including 3 DMs (6 examples), as shown in Table 2.

Table 2. DM co-occurrences in the corpus

Configuration	DM1	DM2	N	Degree of integration

---

[10] On *you know* frequently occurring in DM combinations, see the results in Lohmann & Koops (2016). They include this marker in their study but exclude other units such as agreement and disagreement markers (*yes, no, okay*) and interjections (*oh*) considering that they do not exhibit a linking function and can constitute a turn on their own.

Cj – ParenC **(49 cases)**	and	so (16), then (13), actually (2), in fact (2), therefore (1), secondly (1), I mean (1)	**36**	**addition** (composition)
	but	I mean (3), anyway (2), nevertheless (2), actually (1), in fact (1)	**9**	
	so	for example (1), for instance (1), in fact (1), anyway (1)	**4**	
Cj – Cj **(32 cases)**	and	if (8), as (2), because (2), even if (1), although (1), when (1)	**15**	**juxtaposition**
	but	if (3), as (1), because (1), when (1)	**6**	
	so (that)	when (3), if (2)	**5**	
	because	if (3), when (1)	**4**	
	or	if	**1**	
	whereas	when	**1**	
PragC/ParenC – Cj/ParenC **(22 cases)**	well	if (5), I mean (5), actually (3)	**13**	**juxtaposition** (addition, composition)
	I mean	because (1), when (1), for example (1)	**3**	
	now	then (2), if (1)	**3**	
	for instance	If	**1**	
	then	if	**1**	
	therefore	if	**1**	
Interj– ParenC	okay	so	**1**	**addition**
Cj – ParenC – Cj **(5 cases)**	and	I mean when	**1**	**juxtaposition**
	because	for example when (1), I mean if (1)	**2**	
	but	anyway I mean	**1**	
	so	if for example	**1**	
PragC – Cj – ParenC	well	because otherwise	**1**	**juxtaposition**
Total			**109**	

As Table 1 shows, DM1s are: *and, because, but, for instance, I mean, now, okay, or, so/so that, then, therefore, well, whereas.* DM2s are: *actually, although, anyway, as, because, for example, for instance, I mean, if/even if, in fact, nevertheless, otherwise, secondly, so, then, therefore, when*. Only *because, for instance, I mean, so, then* and *therefore* occurred in both positions in our corpus.

Co-occurrences of three DMs are not very frequent (6 cases out of 109 sequences) and consist of a conjunction followed by a parenthetical connective (often indicating exemplification or re-orientation) and by another conjunction: *and I mean when, because for example when, because I mean if*. However, other configurations are possible: *but anyway I mean, so if for example, well because otherwise*.

The six strings containing three DMs only occurred once, and most of them in conversation. By contrast, some of the two-DM co-occurrences are relatively frequent: *and so* (15 cases), *and then* (13 cases), *and if* (8 cases), *well if* (5 cases), *well I mean* (5 cases), *because if, but if, but I mean, well actually* (3 cases each). It can be concluded that, in the corpus analysed, co-occurrence of DMs typically involves one general conjunction (mainly *and* or *but*) or the very frequent oral marker *well*. The most frequent configuration in our spoken English corpus is 'conjunction + parenthetical connective' (49 examples), followed by the two juxtaposed conjunctions (31 examples).[11]

5. A revised cline of co-occurrence

In this Section, the different degrees of co-occurrence will be defined and illustrated, namely, juxtaposition, addition and composition. The defining features will be established and the examples classified and analysed in order to refine Cuenca & Marín's (2009) original proposal.

---

[11] On the linear ordering of co-occurring DMs, see the discussion in Lohmann & Koops (2016).

5.1 Juxtaposition of discourse markers

Two or more DMs are juxtaposed when they take scope over different units. Juxtaposition usually involves two consecutive conjunctions (6).

> (6)　he said he seemed quite quite happy to meet you (0.320) I'm I'll attempt not to turn this off // well I mean it's no problem [*because* [*if* he doesn't turn up if he doesn't turn up] I'll just uhm (0.020) you know go and get some sandwiches or something] (CONV 53)

In (6), *if* takes scope over the protasis of a conditional sentence ('he doesn't turn up'), whereas *because* takes scope over the conditional sentence as a whole ('if he doesn't turn up I'll just go and get some sandwiches···'). The meanings are clearly different (cause and condition).

In the cases of juxtaposition of conjunctions in the corpus, DM1 is either *and, but, because, so* and less frequently *or* and *whereas*, and DM2 corresponds to the conjunctions *if* (and *even if*), *because, as, although* and *when*. The clusters of two conjunctions identified in the corpus (32 examples) are shown in Table 3:

Table 3. Co-occurrences of two conjunctions (juxtaposition)

and	if (8), as (2), because (2), even if (1), although (1), when (1)	**15**
but	if (3), as(1), because(1), when (1)	**6**
so (that)	when (3), if (2)	**5**
because	If (3), when (1)	**4**
or	if	**1**
whereas	when	**1**

There are some other configurations of two DMs taking scope over different units (10 examples):

Table 4. Other juxtaposed DMs

well	if (5), because otherwise (1)	**6**
now	if	**1**
for instance	if	**1**

then	if	1
therefore	if	1

The latter cases of juxtaposition include a conjunction as DM2, corresponding to the conditional *if* in most cases, whereas DM1 is a parenthetical connective (*therefore if, for instance if, then if*) or a pragmatic connective (*well if*, *now if, well because otherwise*) that acts as a discourse organiser and typically occurs after a pause.

(7)   the reason being that uhm direct exports between (0.830) EEC countries uhm (1.660) does not attract any subsidy (0.330) but because the EEC is collectively concerned to increase its revenue with other trading blocs (1.050) a subsidy's paid for exports sent to countries outside the block (0.960) **therefore if** you (0.273) if you (0.540) whiz your Ed your boring Edam presumably down through France nip it into Andorra you've gone outside the EEC I think because Andorra is not officially part of the EEC (0.487) and you then send it to Germany (0.847) and somewhere along the lines someone collects an enormous amount of money which we are contributing to (CONV 39)

(8)   it really does have to be uhm two twenty // mm // well (0.030) two thirty maybe you know // mm (0.050) **well if** he's gone against the agent's advice already (0.020) and slapped another fifty (CONV 5)

(9)   [playing Scrabble] why have I got such a terrible collection of letters here // mm (0.920) so have I // and the board's not much better (0.060) Uhm (0.220) **now if** I had an S (3.450) I could do a really clever word (0.410) ah there I go // trip // twelve // oh yes (0.220) well that would be very nice for anybody with an E (0.270) haven't got one (CONV 64)

Finally, it is worth noticing that clusters of three markers involve juxtaposition at least of one of the markers (10).

(10)   the conversation next to you someone's saying he's really boring blah blah blah he's really materialistic (0.070) this is what we do all the time we sit and describe other people **and I mean when** people got stuck I'd just say look just listen you imagine you're in a pub and I've said to you what's he like (CONV 18)

In (10) the DMS *and* and *I mean* combine, whereas the conjunction *when* is juxtaposed to the previous sequence, since its scope is more reduced.

5.2 Addition of discourse markers

There is addition of two co-occurring DMs when the markers take scope over the same discourse unit and they exhibit distinct but compatible meanings. Specifically, the second marker either narrows down or reinforces the meaning of the first one, which is more general and underspecified.

(11)   I like it like that // oh God you just don't (0.050) first of all you don't score so much (0.030) **and secondly** you only get rid of two letters (CONV 66)

(12)   when she answers the phone she says Mrs French's residence and I always thought that she's she's taking the piss **but in fact** she's absolutely dead serious // (CONV 1)

In (11) *and* introduces an enumeration and *secondly* specifies that the following segment is the second of the list; in (12) *in fact* reinforces the contrastive meaning of *but*. In these cases, as opposed to the cases of juxtaposition including a parenthetical or a pragmatic connective as DM1 discussed in the previous section (5.1), the meanings are not completely different but similar or compatible. The degree of integration, however, is lesser than in the case of a compound DM (see section 5.3) and the substitution of DM2 by a synonym or near-synonym is possible with no substantial meaning change (*secondly > also/in addition; in fact > actually*).

The cases of added DMs in the corpus correspond to a generic conjunction followed by a parenthetical connective, as in the previous examples. In our corpus, the prototypical clusters of added DMs include *and* or *but*, and also *so*. The most frequent case is that of *and so*.

Table 5. Added DMs (conjunction + parenthetical connective)

and	so (15), actually (2), in fact (2), secondly (1), then (1), therefore (1)	**22**
but	nevertheless (2), actually (1), in fact (1)	**4**
so	in fact (1), for example (1), for instance (1)	**3**

The general conjunctions *and* and *but* combine with other markers that specify or reinforce their meaning. It is also the case of *so*, although its polyfunctionality must be taken into account: it can act as either as a conjunction (DM1) or as a parenthetical connective (DM2), and it can express consequence but also conclusion, continuity or other structural functions alike.

As for their meanings, the addition of a conjunction and a parenthetical marker in the corpus illustrate different possibilities:

(i) *and* + ParenC:
addition + sequencing: *and secondly*
addition + contrast: *and actually, and in fact*
addition + consequence: *and therefore, and so, and then* (equivalent to *and as a consequence*)
(ii) *but* + ParenC:
contrast + specification (or reinforced contrast): *but in fact, but actually*
contrast + concession (or concessive contrast): *but nevertheless*
(iii) *so* + ParenC:
consequence/continuity + specification: *so in fact*
consequence/continuity + exemplification: *so for example, so for instance*

The clusters *and so* and *and then* are instances of addition when they are equivalent to 'and as a consequence' (13), but the string can also instantiate composition when it indicates discourse re-orientation or change (14), as will be developed in Section 6.2.

> (13)   at one stage the m- (0.120) the partner that was in charge of marketing went on a sabbatical (0.190) to South Africa (0.670) and she said to me would I like to take over the marketing function because I'd been doing networking events (0.500) actively going out and trying to gain business (0.320) um **and so** I said yes // so it was more circumstances then that led you to have a bit of (0.150) a career change (INT 48)
>
> (14)   they know if the baby isn't quite as well as it was (0.340) you know because the nurses (0.270) big beaming smile and saying hi (0.500) you know lovely to see you (0.020) you know it's kind of (0.650) you know we need to just talk about

things (0.430) **and so** they're very sensitive to body language from us (0.220) so you do have to be (0.050) take care that you don't frighten them (INT 10)

Other cases of addition of DMs generally imply one marker with a structural function. This includes the cases of *and I mean* (2 cases), *well actually* (3 cases) and *so anyway* (1 case), which will be illustrated and discussed later (Section 6).

5.3 Composition of discourse markers

Two co-occurring DMs form a compound DM when they have the same scope and jointly express a single meaning, that is, the contribution of the individual markers can no longer be disentangled. As González (2004: 297) suggests, their "combinatory functions result in a change of attentional state of the speaker, or shift in cognitive frame, and/or a remarkable emphasis on the illocutionary point of the segment."

(15)   imagine you're in a pub and I've said to you what's he like (1.900) and then just describe them (0.200) he's this and this // I think people aren't used to describing peoples's personalities // no no no it wa- **but anyway** all I'm saying is it's been a very interesting way of meeting people

(16)   it has to comply with all the safety features it has to have emergency exits the engine has to have a cut off (0.580) uh point and all sorts of things of that sort (0.510) **and then** on the water (0.300) it has to be (0.350) registered with the maritime and coastguard agency to carry passengers and again there are a range of safety (0.360) implications (INT 24)

In (15) *but* and *anyway* act together as a repair of an incomplete utterance prefacing a concluding remark. In (16) *and then* announces a different subtopic ("on the ground" vs. "on the water") and introduces a new item in a non-numbered list of requirements. This kind of combination exhibits a structural or a mixed structural-modal meaning equivalent to a single DM used at the structural domain, with a major discourse-structuring function. The cluster acts at a more global level than markers that simply add. A preceding pause is very frequent, whereas there

cannot be any intermediate pause between the two markers. The clusters identified are the following:

Table 6. Compound DMs

and	so (3), then (9)	**12**
but	anyway (2), I mean (3)	5
well	I mean	5
now	then	2

The cases of composition in English typically include a general conjunction, mainly *and* or *but*, and a parenthetical connective. The compound marker indicates a single function at the structural domain. This is the case of *but anyway* and *and then*, both indicating topic or subtopic change or re-orientation. As already pointed out (Section 5.2), the clusters *and then* and *and so* share the fact of being ambiguous, since they both can be equivalent to 'and as a consequence' (added DMs) or can act as structural markers indicating (sub)topic change or discourse reorientation (compound DM). When they are to be interpreted as a compound DM, they tend to occur after a significant pause or a prosodic boundary, and substitution by 'and as a consequence' in an example such as (16) would be odd.

Other co-occurrences that may be considered as cases of compound DMs are: *now then, but I mean, well I mean* (5 cases), all indicating re-orientation in discourse. The strings including *I mean* will be discussed in the next Section.

(17)   the reason I think is that uhm modern medicine (1.120) now enables people to cope // you're getting cheese on your (0.900) jumper // it'll improve the flavour it'll improve the flavour (0.020) Uhm it enables people to come through Uhm yah // Mhm improve appearance more Uhm // **now then** Uhm // by the way Liz is ok for going to the uhm (0.030) Verdi in Oxford (CON 23)

In summary, the two components of a compound DM act as a single marker and the substitution or paraphrase of DM2 by an equivalent element, such as 'as a consequence'(in the case of *then* or *so*), 'in any case'(in the case of *anyway*)or 'in other words'(in the case of *I mean*), would be odd or would change the meaning in

a significant way. Even though the examples do not always behave the same, the substitution test helps differentiate when two markers are added or have integrated into a compound marker.

6. Discussion: fuzzy boundaries and ambiguity

6.1 Juxtaposition or addition?

Some examples in our English corpus pose a challenge for classification. For instance, when the first marker of a string (DM1) is not a conjunction, it is not always obvious whether the scope of the markers is different (juxtaposition) or the same (combination).

> (18)   I never think of anybody as scheming I mean maybe I'm naive (1.560) scheming manipulative a liar // I'd I'd describe manipulative people as manipulative // **I mean because** what you what you what you get when you when you do this is you think (1.210) here are all these people describing other people as (2.090) liars manipulative (CONV 24)

In (18) *I mean* and *because* take scope over the same unit but the functions of the two markers are completely different and do not combine in any sense. In these cases (e.g., *I mean because, I mean for example*), the consideration of the cluster as an example of juxtaposition is certainly as controversial as classifying it as an addition, since the scope of the two markers could be different if the utterance was longer or more complex. This is also due to the sometimes fragmented syntax of spoken language, and to the blurry, variable scope of "fillers" such as *I mean*.

Similarly, in the co-occurrences *and I mean* (19) and *well actually* (20) the scopes seem to converge; however, the functions of the two markers remain sufficiently distinct so that the combination of discourse meanings that addition entails is missing.

> (19)   the conversation next to you someone's saying he's really boring blah blah blah he's really materialistic (0.070) this is what we do all the time we sit and describe other people **and I mean** when people got stuck I'd just say look just listen you imagine you're in a pub and I've said to you what's he like (CONV 18)

(20)   that sounds to me like a sort of you know nineteenth century thing (0.060) a glass (0.020) meaning a mirror // **well actually** a glass means two things it means a barometer or a mirror (CONV 3)

In (19) *and* introduces a continuation whereas *I mean* indicates re-orientation. In (20) *well* prefaces a reacting turn and *actually* indicates contrast.

It is also worth noticing the case of *okay so*, although it can be excluded as a DM co-occurrence and considered just a collocation on syntactic grounds (see Lohmann & Koops, 2016: Section 2.1). DM1 is a modal marker indicating agreement, similar to *well*, and DM2 is a structural marker initiating a new move in discourse. This double move makes it difficult to determine whether the scope of the two markers is the same (addition) or different (juxtaposition). DM1 and DM2 introduce the same segment, but, as an agreement marker, DM1 points backwards, too.

(21)   he's not very old is he (0.120) he's forty odd I would have thought // Uhm not from not from not from where I stand but uhm // **okay so** you're an old man I think we all know that (CONV 32)

Although strings combining a backward-forward movement are common in conversation (see, e.g. Lohmann & Koops (2016: 420) and Pons (in print)), only one case has been found in the English corpus so that no further conclusions can be drawn at the moment.[12]

6.2 Addition or composition?

---

[12] Pons (in print) identifies a case of Sp. *ya però* ('okay/yes but') that he describes as follows: "While *ya* is pointing backwards, showing agreement with the previous intervention, *pero* points forward, anticipating an objection (this is the co-oriented/anti-oriented schema proposed for adversative conjunctions by Anscombre and Ducrot 1983)".

Some of the combinations including the reformulation marker *I mean* (*but I mean*, *well I mean*) can be considered as cases of addition or of composition (as opposed to the cluster *and I mean*, an instance of added DMs where the two meanings are distinct because DM1 is additive and DM2 is reformulative-contrastive; see Section 6.1).

(22) it's it's more cost effective to actually book both nights at the same time (2.440) **well I mean** you can do one one night and one a week later (CONV 11)

(23) it doesn't have to be uhm Le Manoir we could always go to Shinfield and see uhm // Burton race // well if he's handy and his delightful wife yes (0.070) uh which is on on the way almost (0.420) **but I mean** I I do so enjoy (0.910) the atmosphere at Le Manoir (CON 36)

In (22) and (23) the markers share a contrastive meaning or nuance. The cluster instantiates composition if we consider that the two markers jointly perform a single function of non-paraphrastic reformulation or re-orientation. The fact that *I mean* can be used as a filler also points to a unified or merged interpretation of the whole string since its contribution is certainly vague. The difficulty of identifying a stable meaning for pragmatic connectives that are semantically bleached, explains the fuzzy status of some clusters including *I mean* between addition and composition.

Similarly, the cluster *so anyway* incorporates the structural meanings of continuity and change of topic, but it can also be seen as a compound marker of (sub)topic change.

(24) man you like involved interested sense of humour warm open intelligent (0.250) central quality involved prepared to take an interest in (2.640) so (0.633) **so anyway** so all I was saying originally (2.480) is that it's very interesting (CONV 22)

6.3 Ambiguity: The case of and then

The highly frequent cluster *and then* instantiates different configurations and degrees of integration, as already pointed out. The first (and most frequent) use of

and then is when *then* is interpreted as a temporal adverb (equivalent to 'at that time' or 'afterwards') following the conjunction *and* (25).

> (25) they buy the books say for a couple of pounds (1.420) **and then** return it and get half (CONV 16)

In the second use (26), the meaning of consequence (a meaning which can be derived –but differs– from the temporal meaning of *then*) combines with the general meaning of addition, so that the two DMs are added.

> (26) I've got people coming I'll get some salmon from the stall and when you get down there you find he hasn't actually got any **and then** it throws you into a complete quandary (CONV 60)

Lastly, *and then* can express one global function of continuity or sequencing at discourse level (i.e. not temporality or consequence between facts) with contrastive nuances. In this case, the meaning of the cluster is not (strictly) the sum of its parts and *and then* can be considered a compound discourse marker (27-28).

> (27) people do tend to describe themselves [⋯] a lot of people describe people as jealous [⋯] **and then** there are the really bland ones (CONV 30)
>
> (28) so the two uh universities are very different they're complementary (0.320) in their approach (0.610) where you have the more traditional (0.670) uh Bristol university with its many endowments and its traditional approach and its academic curricula and large numbers of postgraduates (0.710) uh many doctoral students (0.480) **and then** you have UWE as it's known university of the West of England Bristol (0.900) uh based in Frenchay (INT 40)

In (28) we can identify the typical structure of sequencers (Hempel & Degand, 2008), namely: an introductory phrase including a quantifier and a classifier (*two universities*) followed by the first element of the sequence (*Bristol University*) and the second element (*UWE*) prefaced by the compound discourse marker (*and then*).

It can be concluded that a single cluster (*and then*) can instantiate different categorical configurations (namely, 'conjunction + adverb' or 'conjunction +

parenthetical connective'), and can also vary along the cline of co-occurrence (either added DMs or compound DM). The previous examples highlight the importance of a flexible, context-bound approach to the issue in future annotation endeavours. These distinctions are subtle and require to take a lot of contextual information into consideration. Yet, they can and should be systematically accounted for, especially since *and then* is also quite frequent in writing (cf. *but then* or *so for instance*, mentioned in the PDTB guidelines). Additional features (e.g. prosody, length and type of host unit) can be investigated to further support this portrait of *and then*.[13]

7. Conclusions

Co-occurrences of DMs are relatively frequent in English speech. In our corpus, most co-occurrences include a conjunction as the first item (*and* (16 cases), *but* (13), *so* (8), *because* (4)) and to a lesser extent the pervasive pragmatic marker *well* (6 cases). According to our corpus, the most frequent configurations in English include two DMs, the first one being a conjunction and the second one either another conjunction or a parenthetical connective (Cj+Cj or Cj+ParenC), among which *and so* (15 cases) and *and then* (13 cases) exhibit the highest scores.

If we compare our results with those reported in Lohmman & Koops (2016: 429), some similarities can be uncovered, despite the fact that the DMs are partially different, theirs being the 11 markers identified by Schiffrin (1987). DM1 corresponds to a conjunction (*and, but, because, so*, *or*) in 15 out of 30 most frequent sequences identified. The markers *well* and *you know* are also relatively frequent as DM1 (5 cases each). DM2 correspond to *then, I mean, now, so* or *you*

---

[13] The role of prosody is certainly important but difficult to apply consistently. See Lohmann & Koops (2016: Section 2.2) for a detailed discussion of prosody as a criterion of DMs integration.

know in 25 sequences out of 30. Finally, the most frequent clusters in their corpus are *and then* and *but then*.[14]

The qualitative analysis presented here shows that not all features are equally important to decide the degree of structural and functional integration of two or more co-occurring DMs. Considerations of scope and of function are criterial in the definition of the degree of integration, whereas prosody (i.e. contiguous pause) and syntactic categories only point to tendencies. Considering the different relevant features, the revised cline of integration of co-occurring DMs proposed here is the following:

(i) *juxtaposition*, when the DMs take scope over different units (mostly when two or more conjunctions co-occur);

(ii) *combination*, when the DMs have the same scope and their functions mix. Combination can lead to addition or to composition of markers:

    a. *addition*, when the DMs have the same scope but distinct compatible meanings that add so that the second DMs narrows down or reinforces the meaning of the first DM,

    b. *composition*, when the DMs have the same scope and jointly express one single meaning.

A further distinction could be made in the case of addition between those cases where the functions of the markers are clearly different and those cases where the functions are so compatible and proximal that they can even be considered one function with a difference in the degree of specificity. The general proposal is presented in Table 7.

Table 7. Degrees of integration of co-occurring DMs (cline)

Type	Subtype	Distinctive features
Juxtaposition		Different scope, different functions
Combination	**Addition**	--Same scope, different functions

---

[14] It must be noticed that some instances of *then* may correspond to the adverbial, which would reduce the final figures if our analysis was applied to that corpus.

		--Same scope, compatible but distinct functions
	Composition	Same scope, one complex function

As for categories, the syntactic behaviour of the DM is not determinant but some tendencies can be drawn, as Table 8 shows, considering only the most frequent configurations.

Table 8. Main configurations and degree of integration

Main configuration	N	Typical degree of integration
Cj – ParenC (e.g. *and in fact, but anyway*)	54	**addition** (composition)
Cj – Cj (e.g. *but when*)	32	**juxtaposition**
PragC/ParenC – Cj (e.g. *well if, I mean because*) PragC/ParenC – ParenC (e.g. *well actually, now then*)	23	**juxtaposition** (addition, composition)

Co-occurrences of DMs usually imply a conjunction followed by either a parenthetical connective or another conjunction. Juxtaposition is the most frequent relationship between two adjacent DMs in our English corpus. Addition and composition are related to the presence of a parenthetical connective as DM2. Addition and composition are not always clearly distinct and sometimes the same cluster can be classified as either added or compound depending on the interpretation and function of the markers involved. For this reason, it may be useful to group addition and composition into the more general category of "combination", especially in the perspective of systematic corpus annotation, in order to avoid such ambiguities.

The study presented here has identified several contexts that can or should be excluded from the concept of co-occurrence (i.e. two consecutive markers relating

to completely different units, markers separated by longer pauses, cases of re-starts or repetition of markers due to performance effects). It has also refined the criteria for each degree of co-occurrence. Although the proposal builds upon the classification developed in Cuenca & Marín (2009), it defines juxtaposition in a more precise way in terms of scope, all the other features (i.e. dominant grammatical pattern, position and domain) being a consequence of difference in scope. The proposal here also points out the fuzzy boundaries especially between added and compound DMs, which are grouped into one type under the term *combination*. Moreover, criteria to locate any specific case in a certain area of the cline are presented and discussed. Final decisions on classification ultimately depend on the definition of the types of co-occurrences and the features considered in the study, which must be as explicit as possible.

The proposed distinction can shed some light on the discussion about the motivation for DMs combination, as discussed in Lohmann and Koops (2016: Section 3.1). Some authors point to functional similarity as the cause of combining, whereas other authors point to complementarity of the combined DMs. Authors pointing to functional similarity, such as Aijmer (2002) or Flores-Ferrán (2014), deal with cases of composition or near composition and focus on oral data. Authors who highlight cases in which the first marker is typically a coordinator and the second one is a more specific and 'strong' marker (e.g. *but nevertheless*), such as Oates (2000, 2001), deal with cases of addition and focus on written data (BNC). As a consequence, the two 'motivations' for DMs combining can be re-interpreted by considering different degrees of integration.

In line with Crible & Cuenca (2017), we suggest that DM annotation endeavours should consider including information about co-occurrence, minimally by identifying clusters, ideally by distinguishing between degrees of integration following the criteria that we have developed in this study. This is particularly crucial for sequences such as *and then* (and its cross-linguistic equivalents, e.g. French *et puis*, Sp. *y entonces*, Cat. *i llavors*), which do not display a unique functional profile depending on co-occurrence degree and categorical adscription.

Our criteria and analysis pave the way for fruitful comparisons across languages, and also across spoken and written registers.

The analysis presented in this paper is applied to speech, as most of the previous analyses of DM co-occurrences. An extension to written text will surely show similarities but also interesting differences.

Finally, our previous experience in analysing Romance language data (namely, French, Spanish and Catalan) points to a challenging research avenue, that of cross-linguistic analysis. At the moment, we can only hypothesize that Romance languages tend to include more complex (and maybe more varied) co-occurrences than English, and that this tendency is connected with the fact that Romance clusters more frequently include connective-modal markers, especially in the case of compound DMs (see, e.g., author 2009, author 2018, Pons in print).

Acknowledgements

References

Aijmer, Karin, 2002. English Discourse Particles: Evidence from a Corpus. Amsterdam: John Benjamins.

Barrenechea, Ana María, Mabel Manacorda de Rosetti, 1971. Estudios de Gramática Estructural. Buenos Aires, Paidós.

Crible, Ludivine, 2015. Grammaticalisation du marqueur discursif complexe ou sinon dans le corpus de SMS belge : spécificités sémantiques, graphiques et diatopiques. Le Discours et la Langue 7 (1), 181-200.

Crible, Ludivine, 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, Ludivine, 2018. *Discourse Markers and (Dis)fluency. Forms and Functions across Languages and Registers.* [Pragmatics & Beyond New Series] Amsterdam: John Benjamins.

Crible, Ludivine, Cuenca, Maria Josep, 2017. Discourse markers in speech: Characteristics and challenges for corpus annotation. Dialogue and Discourse 8 (2), 149-166.

Cuenca, Maria Josep, 1990. Els matisadors: connectors oracionals i textuals. Caplletra 8, 149-167.

Cuenca, Maria Josep, 2001. Los conectores parentéticos como categoría gramatical. Lingüística Española Actual, XXIII, vol. 2, pp. 211-235.

Cuenca, Maria Josep, 2002. Els connectors i les interjeccions. In J. Solà (coord), Gramàtica del Català Contemporani. Barcelona, Empúries, vol. Sintaxi, pp. 3173-3237.

Cuenca, Maria Josep, 2006. La vonnexió i els vonnectors. Perspectiva oracional i textual. Vic, Eumo.

Cuenca, Maria Josep, 2013. The fuzzy boundaries between modal and discourse marking. In L. Degand, B. Cornillie & P. Pietrandrea (eds), Discourse Markers and Modal Particles: Description and Categorization. Amsterdam, John Benjamins, pp. 191-216.

Cuenca, Maria Josep, Marín, Maria Josep, 2009. Co-occurrence of discourse markers in Catalan and Spanish oral narrative. Journal of Pragmatics 41, 899-914.

Dik, Simon Cornelis, 1968. Coordination. Its Implications for the Theory of General Linguistics. Amsterdam, North-Holland.

Dostie, Gaëtane, 2013. Les associations de marqueurs discursifs - De la cooccurrence libre à la collocation. Linguistik Online 62 (5).

Flores-Ferrán, Nydia, 2014. So *pues entonces*: An examination of bilingual discourse markers in Spanish oral narratives of personal experience of New York City-born Puerto Ricans. Sociolinguistic Studies 8 (1), 57–83.

Franchini, Enzo, 1986. Las Condiciones Gramaticales de la Coordinación Copulativa en español. Bern, Francke Verlag.

Fraser, Bruce, 2013. Combinations of contrastive discourse markers in English. International Review of Pragmatics 5, 318-340.

Fraser, Bruce, 2015. The combining of discourse markers – A beginning. Journal of Pragmatics 86, 48-53.

González Condom, Montserrat, 2001. Els marcadors pragmàtics compostos en el relat oral anglès i català. Caplletra 30, 73-93.

Halliday, M. A. K., 1970. Functional diversity in language as seen from a consideration of modality and mood in English. Foundations of Language: International Journal of Language and Philosophy 6, 322-361.

Hansen, Maj-Britt Mosegaard, 1998. The Function of Discourse Particles. A Study with Special Reference to Spoken Standard French. Amsterdam, John Benjamins.

Hempel, Susanne, Degand, Liesbeth, 2008. Sequencers in different text genres: Academic writing, journalese and fiction. Journal of Pragmatics 40, 676–693.

Kohn, Kurt, 2012. Pedagogic corpora for content and language integrated learning. Insights from the BACKBONE project. The Eurocall Review 20 (2).

Koops, Christian, Lohmann, Arne, 2015. A quantitative approach to the grammaticalization of discourse markers. International Journal of Corpus Linguistics 20 (2), 232-259.

Lohmann, Arne, Koops, Christian, 2016. Aspects of discourse marker sequencing – empirical challenges and theoretical implications. In Evelien Keizer, Gunther Kaltenböck & Arne Lohmann (eds.) (2016) Outside the Clause: Form and Function of Extra-Clausal Constituents. Amsterdam/Philadelphia. John Benjamins, 417-446.

Luscher, Jean-Marc, 1993. La marque de connexion complexe. Cahiers de Linguistique Française 14, 173-188.

Nelson, Gerald, Wallis, Sean & Aarts, Bas, 2002. Exploring Natural Language: Working with the British Component of the International Corpus of English. Amsterdam: John Benjamins.

Oates, Sarah Louise, 2000. Multiple Discourse Marker Occurrence: Creating Hierarchies for Natural Language. In Proceedings of the 3rd CLUK Colloquium, Brighton, 41-45.

Oates, Sarah Louise, 2002. Multiple Discourse Marker Occurrence: Creating Hierarchies for Natural Language Generation. MA dissertation, University of Brighton.

Pons, Salvador, 2008. La combinación de marcadores del discurso en la conversación coloquial: interacciones entre posición y función. Estudos Linguísticos/Linguistic Studies, 2. Lisboa, Edições Colibri/CLUNL, pp. 141-159.

Pons, Salvador, in print. The combination of discourse markers in spontaneous conversations: keys to undo a gordian knot. Revue Romane.

Prasad, Rashmi, Miltsakaki, Eleni, Dinesh, Nil, Lee, Alan & Joshi, Aravind, 2007. The Penn Discourse Treebank 2.0 Annotation Manual. IRCS Technical Reports Series, University of Pennsylvania, Institute for Research in Cognitive Science.

Rouchota, Villy, 1998. Procedural meaning and parenthetical discourse markers. In A. H. Jucker & Y. Ziv (eds.), Discourse Markers. Descriptions and Theory, Amsterdam/Philadelphia, John Benjamins, pp. 97-126.

Schiffrin, Deborah,1987. *Discourse Markers*. Cambridge: CUP.

Tagliamonte, Sali A. 2016. Teen Talk: The Language of Adolescents. Cambridge: Cambridge University Press