4 Connectives between Syntax and Discourse

4.1 INTRODUCTION

This chapter will address a lesser-studied aspect of discourse connectives, namely their syntactic status. More precisely, an overview will be given of those studies that call on the syntactic features of discourse connectives as a means to disambiguate their function and meaning. These syntactic features cover the morphosyntactic nature of discourse connectives, their "part-of-speech" as it were, as well as their syntagmatic distribution, that is, the position they occupy in the host unit as well as their close syntagmatic dependencies.

In Chapter 1 of this book, we defined connectives as a functional, rather than a syntactic category. The homogeneity of the connectives class lies in the fact that connectives are used by speakers and writers to make explicit discourse relations holding between discourse segments. Morphosyntactically, however, the discourse connectives class is heterogeneous as it includes coordinating and subordinating conjunctions, but also adverbs, prepositions and prepositional phrases, or phrasal verbs. In other words, connectives come with a double categorial affiliation: functional and grammatical. In Pons Bordería's (2001: 228) words: "When we say that *then* is a connective what we mean is that besides being an adverb, some of its occurrences work as a connective, that is, *then* is polysemous." While most studies on discourse connectives do mention this diversity of linguistic forms belonging to the connectives class (see e.g., Fraser, 1999; Lewis, 2006), the focus has been mainly on their semantic and functional properties (see Chapter 3). Thus, little attention has been paid to the role of syntactic considerations in the definition and disambiguation of discourse connectives and their hosts. Yet, a few research strands stand out that will be worked out in this chapter.

On the basis of existing lexicons of discourse connectives, we will first give an overview of the morphosyntactic distribution of discourse connectives in several European and non-European languages (Section 4.2).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

In Section 4.3, we will address a number of studies that focus on the (semi-automatic) identification and annotation of discourse connectives in context. These include theoretical as well as applied studies, where the morphosyntactic status of discourse connectives is key in distinguishing their discourse use from other uses. This is of particular interest in the field of natural language processing (Section 4.3.1), but also in the field of contrastive linguistics, where it has been shown that syntactic categories, including those underlying the description of discourse connective uses, are not always cross-linguistically valid (Section 4.3.2). The last section of this chapter will be devoted to the relationship between the syntagmatic position of discourse connectives and their meaning, which has given rise to numerous studies at the grammar-discourse interface highlighting the fuzzy boundary between discourse connectives and discourse markers. Exemplary synchronic studies will be discussed in the present chapter, while diachronic approaches will be presented in Chapter 5.

4.2 MORPHOSYNTACTIC DISTRIBUTION OF DISCOURSE CONNECTIVES

In recent years, several lexicons of discourse connectives have been compiled in a variety of languages, among which the DiMLex project for (written) German (Stede & Umbach, 1998; Stede, 2002) and the LEXCONN project for (written) French (Roze, Danlos & Muller, 2012) stand out for their pioneering role. In both cases, the discourse connectives are listed with information regarding their orthography, syntactic category and the discourse relations they can express. The main aim of these two projects was to provide declarative knowledge about connectives for purposes of Natural Language Processing (NLP), more specifically automatic discourse analysis tasks. These two lexicons form the basis of the recently developed database Connective-Lex, a web-based multilingual lexical resource for connectives (Stede, Scheffler & Mendes, 2019). This database is a compilation of existing and purposefully built connectives lexicons that have been made compatible in their design for purposes of cross-linguistic comparison.

For the ten languages included so far in the database, the following information is provided: orthography, syntactic category, ambiguity regarding connective and non-connective usage, and coherence relations. Table 4.1 presents for each language the number of coordinating conjunctions (cco), subordinating conjunctions (csu), adverbs (adv) and

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

Table 4.1 *Morphosyntactic categories in the multilingual connective-lex*

	French	English	Dutch	German	Italian	Portuguese	Czech	Ukrainian	Arabic	Bangla
cco	11	9	13	38	77	10	29	14	92	23
csu	131	36	49	66	40	68	27	49	145	23
adv	152	95	94	158	33	99	63	41	26	68
prep	32	36	33	25	16	29	10	9	13	17
other	3	29	19	10	8	21	91	18	120	0
Total	329	205	208	297	174	227	220	131	396	131

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

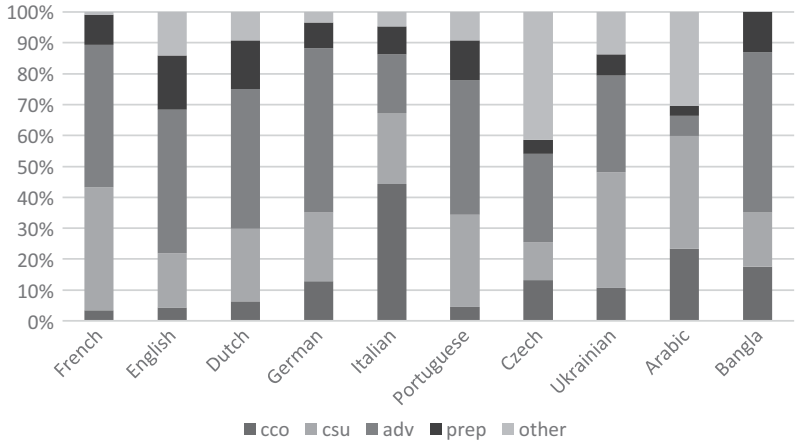


Figure 4.1 Cross-linguistic morphosyntactic distribution of discourse connectives

prepositions or prepositional phrases (prep). The variable "other" covers a variety of morphosyntactic forms, from unknown, to nouns, phrasal verbs or language-specific syntactic categories or expressions.

A closer look at the morphosyntactic distribution of connectives within each language shows that languages make a divergent use of the syntactic stock available for the expression of connectives (Figure 4.1). While Italian discourse connectives most often belong to the coordinating conjunctions category (77/174, or 44.2%), Arabic shows a strong preference for subordinating conjunctions (145/396, 36.6%) and for the category "other" (120/396, 30.3%). Strikingly, the other West-European languages (French, English, Dutch, German, Portuguese) show a similar morphosyntactic distribution, with a majority of connectives being adverbs, followed by subordinating conjunctions, prepositions and coordinating conjunctions. While this cross-linguistic overview hides the subtle idiosyncrasies within each language, the multilingual approach brings to the fore the many challenges that building connectives lexicons include: identifying a robust set of linguistic expressions that count as connectives (for a discussion, see Stede & Umbach, 1998; Pasch et al., 2003; Roze, Danlos & Muller, 2012), defining the syntactic category (see Roze, Danlos & Muller, 2012), establishing the discourse relations that the connectives can mark (in context). The latter endeavor relates to the variety of taxonomies available to describe the senses of discourse relations marked by connectives and was already discussed in Chapters 2 and 3 of this volume.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

A closer look at the criteria used to include or exclude given linguistic expressions from the list of connectives for the different lexicons shows that the syntactic criteria may play a crucial role. In other words, it is not enough for a linguistic expression to be used as an explicit marker of an underlying discourse relation to count as a connective, neither is it sufficient for a given connective candidate to belong to one of the above-mentioned syntactic categories to be automatically included in the connective category.[1] A number of additional criteria apply which are syntagmatic in nature. For purposes of illustration, we will give details for French based on Roze, Danlos and Muller (2012), and complement these with additional syntactic criteria and information from the lexicon-building efforts in the other languages.

In the French LEXCONN database, a syntactic category is associated with each connective, "which can differ a little from traditional ones" (Roze, Danlos & Muller, 2012). Strikingly, the position of the connected segments is important in determining the syntactic category of the connective (see Section 4.4). Thus, coordinating conjunctions are always at the beginning of their host clause, and the "mate segment" is always on the left (see example 1); subordinating conjunctions are also always at the beginning of their host clause, but the mate segment can be anteposed (example 2a), postposed (2b), or internal (2c); prepositions correspond to the reduced forms of subordinating conjunctions when the host clause is an infinitive verbal phrase (3), and adverb applies to connectives, which can appear in various positions in their host clause, and whose mate segment is always on the left as in example (4).

(1) Sophie devait venir aujourd'hui, **mais** elle est encore malade.
'Sophie was supposed to come today, mais/but she is still sick'

(2) a. **Alors que** la chasse est ouverte, ils ont pris le risque d'aller se promener en forêt.
'Alors que/While hunting is open, they took the risk to go for a walk in the forest.'

[1] It should be noted that individual languages may introduce more fine-grained syntactic categories such as adpositions taking an abstract nominalized argument or clitics in Arabic, pre- or postpositions with abstract nominal arguments in German or postpositions with nominal phrases in Chinese. On the other hand, the English lexicon based on the Penn Discourse Treebank (PDTB) only includes connectives belonging to the categories conjunction and adverb strictly taking two clausal arguments.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

b. Ils ont pris le risque d'aller se promener en forêt, **alors que** la chasse est ouverte.
'They took the risk to go for a walk in the forest, alors que/while hunting is open.'

c. Ils ont, **alors que** la chasse est ouverte, pris le risque d'aller se promener en forêt.
'They have, alors que/while hunting is open, taken the risk to go for a walk in the forest.'

(3) Elle a consulté le calendrier de la chasse **avant de** partir en promenade.
'She consulted the hunting calendar before going for a walk.'

(4) Il fait très beau aujourd'hui. (**Donc**) tout le monde est (**donc**) parti se promener (**donc**).
'The weather is very nice today. Donc/Hence everybody (donc/hence) went for a walk (donc/hence).'

[constructed examples]

In the German DiMLex, similar positional constraints within the sentence and information regarding linearization of the two connected arguments are mentioned in the lexicon. In addition to categorial and positional criteria, Roze, Danlos and Muller (2012) present three properties of discourse connectives: they are not integrated to propositional content (Cleft Criterion), they cannot be referential expressions (Substitutability Criterion), and their meaning is not compositional (Compositionality Criterion). Only the first one is syntactic in nature. For the authors of the LEXCONN database, connectives do not belong to the predicative structure; therefore, they cannot be focused in a cleft construction. This criterion makes it possible to distinguish a non-discourse usage (example 5) from a discourse usage (example 6) for *à ce moment-là* (at that moment, then/so). The former can be focused in a cleft construction, while the latter cannot.

(5) Il a commencé à pleuvoir. **A ce moment-là**, Paul est arrivé / C'est à ce moment-là que Paul est arrivé.
'It started raining. At that moment, Paul arrived / It is at that moment that Paul arrived.'

(6) Tu as l'air de penser qu'elle n'est pas honnête. **A ce moment-là** ne lui raconte rien / # C'est à ce moment-là que ne lui raconte rien.
'You seem to think that she's not honest. In that case/So don't tell her anything.'

[adapted from Roze, Danlos & Muller, 2012, examples 2a–3c]

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

This short overview of the syntactic information stored in connectives lexicons provides us a first glance of the importance to consider the syntactic status of these linguistic expressions, not the least for purposes of (semi-automatic) disambiguation of discourse connectives in natural language processing applications.

4.3 USING SYNTACTIC FEATURES TO DISAMBIGUATE DISCOURSE CONNECTIVES

When it comes to considering the role that syntactic features may play in the disambiguation of discourse connectives, two strands of research stand out: Natural Language Processing applications on the one hand, and (theoretical) case studies, often in a contrastive perspective, on the other.

4.3.1 Disambiguation of Connectives in NLP Applications

NLP applications are concerned with the automatization of language production (natural language generation) and language comprehension (natural language parsing) tasks, among others for purposes of data mining, text summarization and machine translation. Most, if not all of these applications, assume (correctly, we believe) that discourse connectives are the primary source of information for building the discourse structure of a text. It follows that connectives play a crucial role in (automatic) discourse production and comprehension tasks, and NLP researchers have developed algorithms for the automatic identification of connective usage.

According to Pitler and Nenkova (2009), two types of connectives ambiguity need to be resolved during (automatic) discourse processing. First, there is the ambiguity between discourse and non-discourse usage. To illustrate this ambiguity, Pitler and Nenkova mention the adverb *once*, which can be used as a temporal discourse connective, as in (7) or as a temporal adverbial meaning "formerly", as in (8). There are of course many other ambiguous pairs of this type, in many different languages (see also examples (5–6), above).

(7) **Once** they had taken the decision, the implementation of the department reorganization went smoothly.

(8) He **once** was very rich, but these times are over.

[constructed examples]

The second type of ambiguity has received more attention, both within NLP studies and in manual discourse annotation projects. It concerns

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

the ambiguity of the discourse relation marked by the connectives. For example, *since* can serve as either a temporal or causal connective (examples 9 and 10), *while* as a temporal or contrastive connective (examples 11 and 12), *but* as a contrastive or concessive connective (examples 13 and 14). Actually, the many-to-one relationship between discourse senses and discourse connectives seems to be the rule rather than the exception (see Chapter 3 for an overview in many different languages).

(9) **Since** I started studying linguistics, I am a lot happier.

(10) Tell me what to do, **since** you are in charge here.

(11) The students were very nervous, **while** they were waiting in the corridor to take their oral exam.

(12) Mary studies engineering, **while** her three brothers have left school early.

(13) Jimmy can't make it to the meeting, **but** his collaborator can.

(14) The room was not very comfortable, **but** it was very cheap.

[constructed examples]

Pitler and Nenkova (2009) demonstrate that syntactic features improve performance in the two disambiguation tasks. More precisely, taking into account the morphosyntactic category of the potential discourse connective item, that of its parent, and of its left and right siblings, the authors were able to reach a disambiguation performance close to 95 percent for both the prediction of the discourse vs. nondiscourse use, and for the prediction of the discourse senses. This confirms their assumption that discourse connectives appear in specific syntactic contexts.

Using syntactic features (mainly part-of-speech category, type and linear order of (clausal) arguments) to disambiguate and (semi-automatically) annotate discourse connectives has been applied to a variety of languages. A beacon in this domain of research is the development of the Penn Discourse Treebank (Prasad et al., 2008; see Chapter 2) for English that was manually annotated in various extensions (Webber et al., 2019) and inspired the development of different Discourse Treebanks in as many different languages (see e.g., the Chinese Discourse Treebank (Zhou & Xue, 2015), the Prague Discourse Treebank (Rysová et al., 2016) or the Turkish Discourse Bank (Zeyrek, Demirşahin & Sevdik Çallı, 2013). The principles underlying the identification and annotation of discourse connectives is

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

similar in most cases: Part of the data is manually annotated following the PDTB guidelines and is then used as a testbed for the development of automatic annotation tools. A case in point is the Leeds Arabic Discourse Treebank (Al-Saif & Markert, 2010, 2011; Alsaif, 2012). For the identification task, the authors show that lexical type features and automatically derived morphological and POS features are highly reliable to distinguish connective from nonconnective use. The method is akin to the one proposed by Pitler and Nenkova (2009) with syntactic features including the position of the potential connective (sentence-initial, medial or final), the surface strings of the three words before and after the connective and the POS-tag of the potential connective and the three preceding and following words (via the ATB/Stanford Tagger), as well as of the potential connective's parent and sibling nodes.

An alternative to annotating (manually or semi-automatically) connectives for their meaning or function is to derive their meaning through clustering techniques making use of syntactic information. Based on the assumption that semantic and functional properties of connectives arise from their distributional characteristics, Laippala, Kyröläinen, Kanerva and Finter (2021) use dependency profiles to reveal similarities and divergences in discourse connectives use. Dependency profiles are based on cooccurrence patterns of the connectives with the dependency syntax relations of the segments in which they are involved. Working with 24 discourse connectives in the Finnish Internet Parsebank of 3.7-billion tokens, the authors demonstrate that applying clustering to these dependency profiles results in linguistically meaningful groupings among the connectives, that is, functionally akin connectives end up in the same cluster. Clustering can furthermore be combined with vector analyses "to obtain generic and stable linguistic characteristics of the discourse connectives" (p. 143). While being less fine-grained than manually annotated data, the large-scale, and fully automatic, analysis may reveal distributional patterns at the interface between syntax and discourse that would remain undetectable in smaller data sets. In addition, dependency treebanks being available for many different languages (de Marneffe et al., 2021), the method can be applied cross-linguistically.

Combining syntactic and lexical information for the semi-automatic identification and characterization of discourse connectives is a promising trend in both monolingual and multilingual studies (Nivre et al., 2020; Zeyrek et al., 2020). The fact that similar surface syntactic features are successful in predicting discourse usage of connectives in several languages highlights the importance of making such syntactic criteria explicit when describing this linguistic category. This being

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

said, the theoretical case studies presented in the following section invite us to lower some of our expectations regarding the (cross-linguistic) validity of syntactic surface features when describing connectives at the discourse–grammar interface.

4.3.2 Disambiguation of Connectives in Theoretical (Case) Studies

As already mentioned, few studies focus explicitly on the syntactic properties of discourse connectives. Yet, a few case studies stand out in making explicit how syntactic features may impact on the semantics of the connectives under scrutiny. An exemplary study in this respect is Scheffler's (2005) analysis of causal *denn* ('for, because') in German in which she explicitly links *denn*'s syntactic status as a coordinating conjunction to explain "why *denn* can express causation of epistemically judged propositions and speech acts while *weil* [subordinating 'because'] cannot" (p. 219). The crux of her argumentation is that *denn* conjoins two independent assertions (syntactically expressed by main clauses) via a causal conventional implicature. "In other words, *denn* conventionally implicates that the proposition denoted by one clausal argument is caused by the proposition denoted by the other clausal argument" (p. 217). The subordinating conjunction *weil*, on the other hand, does not conventionally implicate the causal relation, rather the causal relationship is part of the embedded assertion (syntactically expressed in a subordinating clause).

A similar line of reasoning is followed by Debaisieux (2016), when she distinguishes two semantic uses of French *parce que* ('beause') on syntactic grounds rather than on pragmatic grounds. The starting point of her argumentation is the observation that in many languages, subordinating conjunctions like *parce que* are polyfunctional in that they can introduce both dependent and independent sentences. Her detailed syntactic analysis reveals that these two different uses come with different constraints, the dependent use being a case of a "micro-syntactic construction" and the independent use one of "macro-syntactic configuration" (Debaisieux 2016: 80). Among the syntactic properties at stake, she mentions: (i) the pronominal proportion criterion, that is, the (im) possibility to substitute the causal clause by the interrogative pronoun *pourquoi* (literally 'whatfor'), (ii) the syntagmatic criterion, that is, the (im)possibility to cleft the *parce que*-clause, and (iii) the contrasting option, referring to the possibility of contrasting the *parce que*-clause with a *mais*-clause, as illustrated in examples (15a–c).

(15) je lui ai proposé **parce qu**'il était au chômage
'I offered it to him because he was unemployed'

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

a. je lui ai proposé pourquoi
'I offered it to him for what'
b. c'est **parce qu**'il était au chômage que je lui ai proposé
'it's because he was unemployed that I offered it to him'
c. je ne lui ai pas proposé **parce qu**'il était au chômage mais parce que j'aime bien travailler avec lui
'I didn't offer it to him because he was unemployed but because I like working with him'

[from Debaisieux, 2016: examples 9a–c]

Given these different syntactic constraints, Debaisieux holds that the different discursive dependencies of subordinating conjunctions should be linked to their syntactic properties, rather than to distinct semantic profiles. This line of reasoning is taken up by Degand (2019) in her contrastive study of *parce que* and its Dutch counterpart *omdat*. In particular, she demonstrates that the syntactic constraints on subordinating embedding (with V-final constraints in Dutch) are lifted when the conjunction is used in a "macro-syntactic configuration" (discourse marker use, in her terms). In other words, different discursive uses come with different syntactic constraints (see also Günthner, 1996; Freywald, 2016).

At a more general level, these studies concern the question whether the distinction between subordination and coordination is parallel in syntax and discourse (Haiman & Thompson, 1989; Fabricius-Hansen & Ramm, 2008; Gast & Diessel, 2012a; Visapää, Kalliokoski & Sorva, 2014). While subordination and coordination are syntactic notions, there is a growing consensus that clause-combining has both syntactic and discursive reflections on a cline expressing different degrees of dependence between the clauses linked. Gast and Diessel (2012b: 4) list three main parameters to describe the semantic and syntactic analysis of clause linkage in a typologically relevant way: (i) the relation of dependency holding between the clause and "the attachment site", that is, the constituent of any type (noun, clause, verb, . . .) to which the dependent clause is attached, (ii) the properties of this attachment site, and (iii) the properties of the attached clause. For the purposes of this chapter, the first parameter is the most relevant one, since it directly involves discourse connectives as potential signals of the type of dependency.

Moving beyond the traditional syntactic division between coordination (or structural and functional independence) and subordination (reflecting structural and functional dependence), Gast and Diessel (2012b: 4) point out that the "term 'dependency' can be understood in at least three ways: syntactically, semantically

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

(or pragmatically) and prosodically", whereby coordination is conceived of as the absence of any of the above dependencies between the connected clauses, while subordination supposes dependency in at least one of the three properties. Yet, how this dependence is attested, how it can be measured, and how it is expressed cross-linguistically in grammar and in discourse remain questions of debate. In the parametric approach advocated by Gast and Diessel (see also Lehmann (1988) and Bickel (2010)), the distinction is regarded as a gradual one. A number of valuable proposals include fine-grained analyses of distributional features to account for this continuum. Verstraete (2007: 103), for instance, proposes that three parameters of interpersonal grammar, in other words the parameters of modality, speech function and scope "can each take two values in the context of a complex sentence construction (. . .) to set up a basic typology of four construction types". This typology of clause-combining includes coordination, on the one hand, and three types of subordination, on the other, namely bound subordination, free subordination and modal subordination. Thus, coordination is defined by the parameters [+ speech function], [+ modality] and [- scope]. In other words, coordinated clauses function independently from one another. They have equal status with a speech function and modality of their own and have scope only over their own propositional material, as illustrated in example (16):

(16) Mattéo was born 7 months ago **and**
he is the cutest baby in the world.
he is most probably the cutest baby in the world.
Isn't he the cutest baby in the world?
[constructed example]

The three types of subordination, then, are characterized by the parameter [- speech function] indicating absence of a speech functional value in the secondary clause. Bound subordination is further marked by the features [- modality] and [+ scope] indicating that the secondary clause falls entirely within the scope of the main clause and cannot have its own epistemic or deontic modal value (example 17).

(17) a. He cried a lot while you were away.
b. It is while you were away that he cried a lot.
[constructed example]

Free subordination is also [- speech function] and [- modality], but it is [- scope], indicating that the secondary clause does *not* fall within the scope of the illocutionary force of the main clause, yet it has neither a speech functional nor a modal choice of its own. In (18), the matrix

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

clause does not have scope over the *when*-clause, rather the two clauses constitute separate intonation units (they are not integrated), the second clause setting the time frame of the question in the matrix clause in a kind of afterthought.

(18) Did he cry? When I was out.

[constructed example]

Finally, modal subordination is [- speech function], [- scope] and [+ modality], which is different from free subordination in that the secondary clause can express independent deontic or epistemic modality, as in example (19).

(19) Whereas it may well be true that Mattéo is a very cute baby, his parents are not very objective.

[constructed example]

Verstraete (2007: chapter 7) convincingly shows that his four-category typology is grammatically motivated in that "the other grammatical criteria that have been used in the analysis of clause combining neatly follow the distinctions defined by the basic parameters of interpersonal grammar" (Verstraete, 2007: 186). Among these, there are, for instance, preposing, clefting and *wh*-questioning, integrated intonation and word order, but also connective selection, which is of interest to us here. Table 4.2 (based on table 32 in Verstraete, 2007: 187) displays a categorization of connectives in terms of the interpersonal typology.

The connectives *and, but, or, for* always introduce clauses associated with the value [+ speech function]. "Another set of conjunctions never allows any speech functional options besides the unmarked declarative, and is therefore invariably associated with the value [- Speech Function]. This category includes *after, before, when, whereas, while, since,*

Table 4.2 *Verstraete's (2007) clause combining typology and the categorization of connectives*

Construction type	Connectives
Coordination	*and, but, or, for, because, although*
Modal subordination	*whereas, when, while, so that, since, as, although*
Free subordination	*after, before, if, until, once, as, since, so that, when, while, because*
Bound subordination	

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

as, so that, until, once, and *if* (Verstraete, 2007: 110), which we find in free and bound subordination. A number of connectives are compatible with different clause-combining types. The connectives *because* and *although* "can have both the value [+ Speech Function] and the value [- Speech Function]. *Because*-clauses and *although*-clauses allow non-declarative clause types in final position (. . .), but they do not allow them in initial position", as shown in (20) and (21), respectively.

(20) I believe it's time 'Saint' George Galloway was officially canonised preferably using an Iraqi supergun. His intentions must be honourable because who in their right mind would go to Baghdad right after the Gulf War, in which British troops died, and praise the lunatic responsible for starting it?

(21) *Because who in his right mind would go and praise the lunatic who started the war, Galloway's intentions must be honourable.

[from Verstraete, 2007: 111]

Regarding the parameter of modality, again the values [+ modality] and [- modality] "each define their own sets of conjunctions" (Verstraete, 2007: 114). The set with a positive value for modality includes the connectives *and, or, but* and *for*, compatible with coordination, but also the connectives *although* and *whereas*, compatible with modal subordination (see example 19, above). The set with a negative value for modality, thus not allowing any subjective modal auxiliaries in their host clause, include the connectives *after, before, until* and *once* found in free and bound subordination. Finally, there is a set of connectives that is [+/- modality], namely *as, since, because, so that, when*, and *while*. According to Verstraete (2007: 117), "[i]t is not a coincidence that all of these conjunctions are ambiguous, either between temporal and contrastive (*when, while*), temporal and justifying (*as, since*), causal and justifying (*because*) and result and purpose (*so that*) meanings." In other words, the semantic distinctions correlate with the presence or absence of a modal value determining the grammatical clause-combining constructions they may occur in. This aligns with Debaisieux' (2016) findings for *parce que* ('because'), referred to above, albeit in another framework and terminology.

Summarizing the ongoing research on the role of discourse connectives in complex clauses, two research strands stand out, which we will refer to as the clause-combining vs. the discourse relational perspective. The former approach holds that the syntax and semantics of discourse connectives offer little evidence to support the assumption of a structural parallelism between syntax and discourse (for an

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

interesting synthesis, see e.g., Blühdorn, 2008). Rather, the focus of attention is on the different degrees of dependence between two clauses and whether or not "syntactic dependency" is reflected in "discursive dependency". The short answer to this question is "no", the longer question, however, requires a study of clause linkage strategies in a preferably cross-linguistic perspective (see, the studies in e.g., Bril, 2010, Dixon & Aikhenvald, 2011; Gast & Diessel, 2012a) trying to disentangle how coordination and subordination are realized both at the discourse level and the syntactic level, both in syndetic (with a signaling marker) and asyndetic (without explicit marking) combinations. In contrast, the discourse relational approach assigns a crucial role to the discourse connectives as signaling a specific discourse relation, often ignoring their syntactic nature and giving primacy to semantic aspects (Keller, 1995; Gernsbacher, 1997; Stukker & Sanders, 2012; Canestrelli, Mak & Sanders, 2013). Strikingly, in this work, the syntactic status of the linked clauses – including their degree of dependence to one another – is hardly ever accounted for. In other words, as soon as a discourse connective is identified, it is considered as a marker of a discourse relation. Implicitly, a direct mapping is established between clause combining at the grammatical level and discourse relations, focussing on their meaning and/or function but not on their syntagmatic constraints.

Yet, when addressing the question of clause linkage, the syntactic nature of the segments that are linked by a discourse connective cannot be ignored. Most studies suffice with the general statement that discourse connectives make explicit the relation between spans of texts, which by default are defined as "clause-like" (see e.g., Elementary Discourse Units segmentation (EDUs) in computational applications of Rhetorical Structure Theory (Marcu, 2000)). In the connective lexicons (Section 4.2), reference is made to the host clause containing the connective and to the mate segment to which it is related, but what these segments look like from a syntactic point of view is barely touched upon. Dupont (2021) convincingly shows that contrastive research on connectives may fare well from a more systematic syntactic account of the related segments. Focusing on "conjunctive markers of contrast" in French and English, she defined the syntactic features of the connected clauses along two main dimensions: clause types (i.e., finite, non-finite, verbless) and rank status (i.e., main, hypotactic, embedded or minor). From a morpho-syntactic point of view, her analysis demonstrates that the boundaries between coordinators, subordinators and conjunctive adjuncts is actually far from clear-cut, leading her to conceive the boundaries between the three

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

grammatical classes of markers she investigates as different levels on a cline. She also lays bare (slightly) different syntactic distributions of the contrastive connectives and their host clauses in English and French. Regarding the discourse–grammar interface, she shows interestingly that the use of contrastive connectives "in non-default (. . .) syntactic structures is often motivated by and/or creates a range of effects at the discourse level" (p. 259). For instance, she observes that the use of conjunctive markers of contrast in so-called fragmented syntactic structures (minor clauses, stand-alone hypotactic clauses, sentence-initial coordinators) can serve emphatic discourse functions enhancing the persuasive tone that is central to editorial writing (Dupont, 2021: 263), while the use of discourse connectives in non-finite and verbless clauses (referred to as "syntactic compression", p. 265) is interpreted as a space-saving strategy in newspaper prose, typical of formal writing. All in all, Dupont's corpus study convincingly demonstrates "how closely these syntactic patterns are in fact related to discourse-level strategies" (p. 270).

4.4 THE IMPACT OF THE SYNTAGMATIC POSITION OF CONNECTIVES ON THEIR MEANING

The final topic addressed in this chapter concerns the influence that the connectives' syntagmatic position may have on their function. In Section 4.2, we already mentioned that lexicons of connectives do take up the information regarding the connective's position in the host clause, with coordinating conjunctions being always in initial position of their host clause preceded by the mate segment, and subordinating conjunctions also in initial position, but with their mate segment either before, after or internal to the host clause. Adverbial connective forms, then, can appear in various positions in their host clause.

Regarding the variation in the position of the host clause *vs* the mate segment in subordinating clause combining, most studies agree that information flow plays a role in the choice between preposing or postposing of the adverbial clause hosting the connective. Renkema (1996), for instance, found that the preposing or postposing of Dutch *omdat*-clauses ('because') is related mainly with topicality and local embedding. In other words, it is not the syntactic status of main clause *vs* subordinate class that is determining in the ordering choice, rather it is the status of the propositional content, and in particular its relatedness to prior information, that is determining (see also Ramsey (1987) for a similar account of English *if*- and *when*-clauses).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

Working out this idea further, Degand (2000) confirmed the high sensitivity of such causal constructions to the surrounding discourse, not alone for the choice between preposing and postposing causal conjuncts, but also in the choice between forward and backward causal connectives (see Chapter 3). Thus, she found that writers tend to organize their discourse in function of both prior and upcoming topics of discourse. This was established in a Dutch newspaper corpus analysis based on the forward connectives *dus* ('so'), *daarom* ('therefore'), *daardoor* ('therefore' – objective use, see Chapter 3) and the backward connectives *aangezien* ('since'), *omdat* ('because'), *doordat* ('because' – objective use only). Diessel (2005), on the other hand, using corpus data from spoken and written English, found an impact of relational meaning on the choice between the two orderings. Working on the basis of clauses marked with the connectives *if, because, when, while, before, after, since, once, until, as, as soon as*, and *as long as*, Diessel found that conditional clauses tend to precede the main clause, temporal clauses are common in both initial and final position, and causal clauses usually follow the main clause. Furthermore, it is shown that final occurrence of adverbial clauses is motivated by processing, postposing being easier in terms of utterance planning (Wasow, 1997, 2002, cited in Diessel, 2005: 458), while initial occurrence results from semantic and discourse pragmatic forces (in line with the studies by Renkema (1996) and Degand (2000)) that may override the processing motivation. Verstraete (2004), however, in a theoretical constructional study, proposes that the selection of preposed adverbial clauses in English has to do with aspects of the illocutionary status of the conjoined clauses (see Verstraete (2007) presented in Section 4.3.2). More precisely, preposed adverbial clauses (whatever the connective that introduces them) do not constitute a separate speech act (being restricted to the declarative mood) and they do not fall within the illocutionary scope of the main clause. It follows that "initial clauses are interactionally 'deactivated' and backgrounded, whereas (...) their function is not restricted to purely local modification of their immediate main clause" (Verstraete, 2004: 848). This finding is compatible with the discourse pragmatic function of initial clauses found in other studies (*cf. supra*).

Regarding the syntagmatic position of the connective itself in the host clause, there is a strong tendency for initial position, even if there are typological differences with discourse connectives taking the form of sentence-final particles, (e.g., in Japanese; Onodera, 2004; Haugh, 2008; Izutzu & Izutzu, 2014) or conjunctive suffixes (e.g., in Korean; Oh, 2005). Strikingly, when the position of the connective is variable, it may play a role in disambiguating discourse usage from nondiscourse

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

usage (see Section 4.3.1). For instance, Le Draoulec and Bras (2007) showed that depending on its syntactic position in the sentence, the French marker *alors* ('then, so') performs distinct (discourse) functions. When it is used initially, *alors* functions as a conjunctive marker "expressing a dependence link" (p. 88), which may be temporal or consequential, whereas in medial position, *alors* can only function as a temporal adjunct. Similarly, Goutsos (2017) explored the position of discourse markers (functional markers in his terms) in Greek and found that markers occurring in initial position fulfil discursive functions, whereas modal markers occur in medial position. Many other studies have focused on the influence of the connective's syntagmatic position on the discursive meanings themselves, with special attention to the opposition between initial position and final position, often described as left-peripheral vs. right-peripheral uses. For instance, Degand (2014) showed that the connectives *alors* and *donc* ('so, then') express relational meanings of time, consequence and condition in initial position, but also "metadiscursive" meanings of topic management. This is more rarely the case for clause-final position. Degand (2011) finds a similar distribution for Dutch *dus* ('so').

This relation between syntactic position and discourse function was confirmed in Degand and Fagard's (2011) diachronic study of *alors*, where they showed that the marker acquired different discourse functions when moving towards the initial position (see Chapter 5). The idea has been worked out most extensively within the framework that Salameh, Estellés and Pons Bordería (2018) have labelled the *Subjectivity, Intersubjectivity and Peripheries Hypothesis*, introduced extensively in Beeching and Detges (2014), with studies on pragmatic expressions in French, English, Japanese, Italian, Chinese and Korean (see also Onodera & Traugott, 2016). Within this framework, left and right peripheral uses have been associated respectively with subjective and intersubjective meanings, a general trend that goes beyond connectives to include discourse markers and other pragmatic expressions (see also Haselow, 2017). Subjective meanings are linked to the perspective of the speaker/writer, who is drawing contextual inferences, while the intersubjective meanings refer to functions that are more dialogic in nature, concerned with speaker-hearer management (Chapter 3). While this hypothesis has been worked out and at least partially confirmed cross-linguistically in numerous case studies, it has also been challenged or nuanced. In particular, several authors have shown that this functional division of labour does not always hold, with instances of subjective uses in the right periphery, and intersubjective ones in the left

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

periphery (e.g., Haselow, 2012; Traugott, 2012; Salameh, Estellés & Pons Bordería, 2018; Heim, 2019).

More recently, Degand and Crible (2021) tackled the division of labour between the peripheries by investigating the "functional domains" expressed by discourse markers in peripheral and nonperipheral position. While Degand and Crible's study concerns not only discourse connectives, but also nonrelational discourse markers like *tu vois* ('you see/you know'), *bon* ('well'), or *hein* ('right'), their results show that initial position is overwhelmingly occupied by discourse connectives (conjunctions and conjunctive adverbials), fulfilling mainly sequential, text-structuring functions, followed closely by rhetorical and ideational uses. While the interpersonal domain is highly infrequent in initial position, it covers more than half of the cases in final position, even if rhetorical and sequential uses are not absent. In the latter case, discourse connectives are almost completely out of the picture. Besides confirming indirectly that the intersubjective meanings are indeed strongly connected to the right periphery (cf. the Subjectivity, Intersubjectivity and Peripheries Hypothesis), the study also indicates that this functional division holds for the peripheries of both clauses and turns, but not for intonation units. Discourse markers appearing overwhelmingly at the boundary of syntactic clauses (more than 90 percent), the authors suggest that "syntax is where DMs act most frequently as boundary markers" (Degand & Crible, 2021: 40) in a functionally motivated way.

4.5 SUMMARY

The starting point of this chapter was the observation that syntactic descriptions of discourse connectives have been rather scarce, even if they are not inexistent. On the one hand, there are a number of efforts aiming to describe the syntactic constraints that come with individual connectives, most notably in connectives lexicons. Such syntactic features have been applied in NLP to disambiguate connective use *vs* non-connective use, but also to disentangle different connective senses. Most of these works are corpus-based. On the other hand, there have been a number of encompassing theoretical works, in which the aim is to account for clause-combining types, considering coordination and subordination as two extremes on a scale of dependency, with either a gradient cline between these two ends or discrete intermediate categories (Section 4.3.2). Connectives are in this context seen as additional evidence for a given grammatical categorization, but they are not the primary focus of description themselves. Nonetheless, the focus on grammar and syntactic constraints should not

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004

be seen as independent from the semantic meaning expressed. In particular, it was shown that different grammatical configurations go hand in hand with different semantic meanings and categories (as presented in Chapter 3), but also that the syntagmatic distribution of discourse connectives in their host clause influences the meaning they may express. The interplay between grammar and semantics will be worked out further in Chapter 5 from a diachronic perspective.

DISCUSSION POINTS

- What is the role of syntactic features in the disambiguation of discourse connectives?
- To what extent should discourse connectives be considered as grammatical expressions?
- Are there any cross-linguistic regularities in the syntactic distribution of discourse connectives?
- What would explain the general tendency for left and right peripheral position to attract different connective meanings?
- To what extent are the functional and cognitive motivations given for preposed adverbial clauses cross-linguistically valid?

FURTHER READING

Blühdorn (2008) illustrates how connectives may provide evidence for a model of language in which the organization of subordination and coordination is considered independently in syntax, semantics and discourse. The chapter is part of the volume *"Subordination" versus "Coordination" in Sentence and Text: A Cross-linguistic Perspective* edited by Fabricius-Hansen and Ramm, of which the introduction presents a clear overview of the challenges in trying to disentangle the different categories of clause-combining and the role connectives may play therein. Gast and Diessel (2012a) pursues a similar objective in a more typological perspective. Focusing on connectives, Stede, Scheffler and Mendes (2019) is an influential contribution on (multilingual) connectives lexicon building that has inspired many authors since, both inside and outside NLP. Regarding the syntagmatic distribution of discourse connectives, in particular how position and meaning of discourse connectives interact, there is the seminal volume by Beeching and Detges (2014) and the more recent volume by Van Olmen and Šinkūnienė (2021) working out similar topics.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 10 Nov 2026 at 13:47:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108966573.004