German parenthetical discourse markers between perception and cognition
An explorative approach to parallel corpus data

Regina Zieleke
Eberhard Karls Universität Tübingen
regina.zieleke@uni-tuebingen.de

When perception verbs are employed as parenthetical discourse markers, e.g. English *you see*, French *tu vois*, the concrete visual perceptual meaning of *see* is said to be expanded to a more abstract meaning of general cognition (cf. Brinton, 2008). In this paper, I will show that in German, different cognitive processes map with different parentheticals: visual parentheticals such as *(wie) du siehst* ('(as) you see') are only used in contexts of justification, whereas processes of explanation invoke the use of cognitive parentheticals such as *weißt/verstehst du* ('you know/understand') instead. For this purpose, I will explore data from the parallel corpora Europarl7 and OpenSubtitles2011. The assessment of German equivalents to the English *you see* and French *tu vois* and a paraphrase test aiming at these different cognitive processes provide a pattern linking the latter to German visual vis-à-vis cognitive parentheticals.

1 Parenthetical discourse markers and verbs of perception

Parenthetical discourse markers (also 'pragmatic markers', 'comment clauses', see e.g. Brinton, 2008) such as *you know* or *I mean* are verbal constructions that are "not syntactically connected to the rest of the clause (i.e., [are] parenthetical)" (Brinton, 2008: 1) and are "metacommunicative" in that they "comment on the truth value of a […] group of sentences, on the organization of the text or on the attitude of the speaker" (Peltola, 1982/1983, cited by Brinton, 2008: 5).

On the supposition that 'visual perception is our primary source of information in the outside world' (Bat-Zeev Shyldkrot, 1989, cited by Bolly,

2012: 3[1]), visual perception verbs can be regarded natural candidates for such constructions. However, while English *you see* as in (1) and French *tu vois* as in (2) are frequent, an equivalent construction in German is inacceptable (cf. (3a)) and has to be expressed by a construction involving the cognitive verbs *wissen* ('to know') or *verstehen* ('to understand') instead (cf. (3b)).

(1) I went to three different stores to find the perfect avocado. **You see,** I love guacamole.

(2) J'ai cherché l'avocat parfait dans trois magasins différents. **Tu vois**, j'adore du guacamole.

(3) Ich war in drei Läden, um die perfekte Avocado zu finden.
     a. **#Du siehst / #Siehst du**, ich liebe Guacamole.
     b. **?Du weißt / Weißt du / Verstehst du**, ich liebe Guacamole.

The link between (verbs of) perception and cognition is well-known (cf. 'I see your point' vs. 'I know what you mean', see e.g. Sweetser, 1990). Viberg (2015: 96), for example, states that "verbs of perception are situated in the middle of a continuum of more raw descriptions of sensations at one end and more abstract reference to thinking and knowledge at the other end". According to Brinton (2008: 159), this path can also be observed in the grammaticalization process of constructions such as English *(as) you see* or *I see*: "the concrete visual perceptual meaning of *see* is bleached or widened to a more abstract meaning of general cognitive perception".

There are two reasons to be nonetheless puzzled by this observation. First, both French and English have parenthetical markers involving an equivalent cognitive verb that are very frequent, i.e. French *tu sais* and English *you know*, but these are ascribed with pragmatic functions distinct from *tu vois* and *you see*. Erman (1987: 117/118), for

---

[1] Original quote in French: "[…] en tant que 'première source d'information objective et intellectuelle sur le monde extérieur' (Bat-Zeev Shyldkrot, 1989: 288)" (ibid.).

example, ascribes English *you see* with an argumentative 'terminating' function making "the addressee accept [the speaker's] ideas and explanations". The cognitive *you know*, on the other hand, is ascribed with an 'introductory' function making "the addressee accept parts of the information conveyed as common ground" (ibid., see also Schiffrin's, 1987 account of *y'know* as appealing to shared knowledge).

Second, there are cases such as in (4)-(6) where German does allow for parenthetical markers with *sehen* ('to see') (cf. (6a)), while cognitive parentheticals are less acceptable (cf. (6b)):

(4) The house isn't cleaned and I didn't go grocery shopping yet. **You see,** I still have a lot to do.

(5) La maison n'est pas propre et je n'ai pas encore fait les courses. **Tu vois**, j'ai du pain sur la planche, là.

(6) Das Haus muss noch geputzt werden und einkaufen war ich auch noch nicht.
   a. **Du siehst (also) / Wie du siehst,** ich hab echt viel zu tun.
   b. #**Du weißt / #Weißt du / ?Verstehst du,** ich hab echt viel zu tun.

I argue that the two sets of examples involve two different cognitive processes: one of explanation in (1) – (3) and one of justification or provision of evidence in (4) – (6). As (7) and (8) show, the causal explanation marker *because* is acceptable only in the first case, where loving guacamole is the explanation for going through the trouble of visiting three different stores. Such a relation cannot be applied to (8), where the unpleasantness of an uncleaned house and missing groceries are offered as a justification or evidence for the speaker still having a lot to do, instead. This, in turn, correlates with the paraphrase *this is evidence for the fact that*.[2]

(7) I went to three different stores to find the perfect avocado, **because / ? this is evidence for the fact that** I love guacamole.

(8) The house isn't cleaned and I didn't go grocery shopping yet, **#because / this is evidence for the fact that** I still have a lot to do.

It seems, then, that while parenthetical markers involving perception verbs can be used to express both processes in English and French, they are limited to the process of justification/evidence in German. A hint of a different cognitive status of German perception verbs used as discourse markers is provided by Günthner (2017). She studies the German *guck mal* ('look') and *weißt du* ('you know'), and while her verdict for the cognitive *weißt du* resembles Erman's description of *you know* ('projection of a knowledge transfer making the utterance part of the Common Ground' Günthner, 2017: 125), the visual *guck mal* in its discourse marker use is described as merely involving a shift in perception from a purely visual to the 'discourse world and thus the argumentation structure'.

In this paper, I will discuss data from English-German and French-German parallel corpora. The goal of this explorative approach to parenthetical discourse markers with the visual perception verbs *see/voir* is to find out, how the function of these markers is handled in German. In a first step, I will assess the German equivalents in the parallel data – does German make use of parenthetical markers at all, and if so, do they involve verbs of perception (*sehen*) or cognition (*wissen*, *verstehen*)? The second step consists of assessing possible discourse functions relating to the different cognitive processes – is there a pattern linking the different German equivalents to different discourse functions?

2 *You see/tu vois* and their German equivalents

2.1 Data and annotation criteria

The data is taken from two parallel corpora, Europarl7 and the OpenSubtitles2011 sub-corpus of OPUS2, accessed via SketchEngine. Both corpora consist of aligned transcriptions of spoken language, viz. political discourse data from the proceedings of the European Parliament in the case of Europarl7 and data from subtitles in movies and TV series in the case of OpenSubtitles2011.[3] Both corpora were searched for the two language pairs English–German and French–German each, with

---

[2] In a way, this is also a kind of an explanation: *I still have a lot to do, because the house isn't cleaned… .* However, the order of explanans and explanandum are inversed resulting in a different relation altogether – as shown by the fact that the paraphrase *this is evidence for the fact that* is not *un*acceptable in (7), but alters the sense in that way.

[3] This choice of corpora comes with two restrictions: first, it is often unclear which language is the source and which the translated language; and second, subtitles tend to involve shortened sentences in order to fit on the screen in the available time (cf. Müller & Volk, 2013: 2).

the English and French parenthetical markers *you see* and *tu vois* as the starting point.

In order to exclude matrix verb uses of these verbal constructions, the search request made use of the observation that parenthetical markers are "marked by "comma intonation" (pauses in speech, or actual commas in writing) that separates [the marker] from its anchor" (Brinton, 2008: 8). The positions of the markers (preposed or postposed) were not restricted in the search request. Since French is a language with strong verb inflection and the discourse in Europarl is of a formal register, the search request in this corpus includes the formal second person equivalents of the parenthetical *tu vois*, i.e. *vous voyez* and *voyez-vous*.)[4]

A random sample of 44 sets of data for each language in the Europarl7 and 48 in the OpenSubtitles2011, respectively, was annotated for simple criteria in line with the explorative nature of the investigation. The main focus lies on the (direct) German equivalent of the parenthetical markers in English or French. This involves the annotation of (i) the specific sequence of words (*verstehen Sie* in (9)), (ii) the lemma of that sequence (*verstehen*), and, most importantly, (iii) a categorization of these lemmas as a) perception verb, b) cognitive verb (as in (9)), c) particle/ connective, or d) no equivalent. Annotation further accounts for the presence of further discourse markers in the parentheticals' environment (e.g. the connective *but* in (9)).

(9) EN Sorry, but **you see**, we've gotta check up on everybody.
DE Tut mir Leid, aber **verstehen Sie**, wir müssen jeden überprüfen.
Ref: OPUS2; #176183352, en/1934/5990/4099372_1of1.xml.gz

2.2   Results: A general pattern

Among the 184 sets of parallel data, there are 37 different German linguistic expressions that can be identified as equivalent to their English and French counterparts *you see/tu vois* – form single word expressions such as *eigentlich* ('actually') to complex tag questions such as *weißt du, was ich meine* ('do you know what I mean'). Figure 1 shows the overall distribution of the German equivalents among the four categories described above:
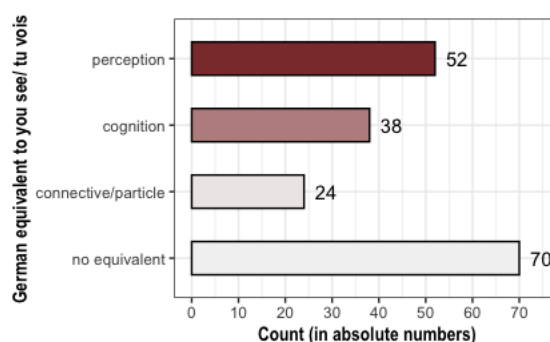

Figure 1: German equivalents to *you see/tu vois*

With 28.3% of the overall data, constructions involving perception verbs such as *siehste* or the phrase *wie Sie sehen* ('as you see') in (10) are more frequent than cognitive verbs such as *weißt du* or *verstehen Sie* in (9) above with only 20.7%. Most frequently, the German data does not contain any equivalent to the French or English parenthetical marker as in (11) (38%), whereas particles/connectives such as *eigentlich* ('actually') or *nämlich* ('namely') in (12) were found least frequently in the corpus data (13%).

(10) EN There is, **you see**, a clear risk that this is just procrastination.
DE […] es besteht, **wie Sie sehen**, eindeutig die Gefahr einer Verschleppung.
Ref: Europarl7; #32016758, /en/ep-08-05-07-014.xml

(11) FR J'étais occupé, **tu vois**.
DE lch war beschäftigt.
Ref: OPUS2; #228764636, fr/1931/8606/3505132_1of1.xml.gz

(12) FR Ce que nous sommes en train de faire, **voyez-vous**, c'est défendre les secteurs qui ne sont pas compétitifs […]
DE Was wir **nämlich** damit zurzeit erreichen, ist der Schutz und die Verteidigung von nicht wettbewerbsfähigen Wirtschafts-zweigen […]
Ref: Europarl7; #28994717, /fr/ep-06-10-11-016.xml

The comparison between the two languages of origin for the search request reveals similar patterns, cf. **Error! Reference source not found.**. Solely the category connective/particle differs considerably: whereas English *you see* is expressed by a connective or particle in its German equivalent in 19.6% of the time, examples as in (12) only make up 6.5% of French *tu vois*.
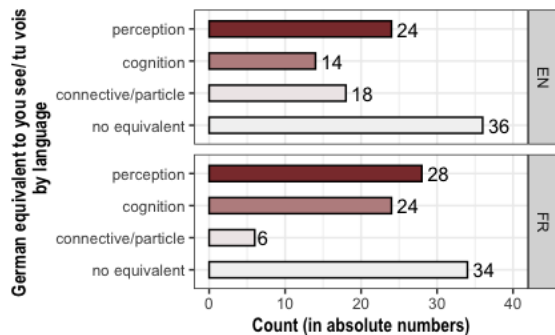

Figure 2: German equivalents by language

The distribution pattern changes completely when comparing the four categories by type of corpus instead, cf. Figure 3. The high number of German equivalents involving perception verbs predominantly relates to the Europarl corpora with 80.7% of the perception-verb-equivalents. The OpenSubtitles corpora seem to be responsible for most of the cognitive-verb-equivalents, instead (94.7%). The majority of German equivalents to *you see* and *tu vois* involving particles/connectives, in turn, correlates with Europarl again (91.7%).
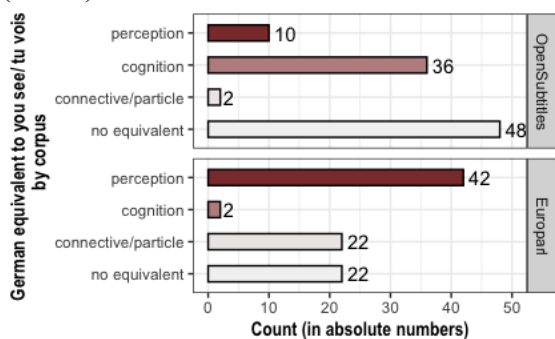

Figure 3: German equivalents by corpus

This considerable change in pattern is particularly interesting if we consider the type of discourse represented in the two kinds of corpora. Since Europarl comprises political discourse data from the proceedings of the European Parliament, the discourse can be said to be more argumentative in nature. This matches well with our assumption that in German visual parentheticals relate to the cognitive process of justification – the speakers use parenthetical markers such as *sehen Sie* or *Sie sehen also* to mark the provision of evidence for their argumentation. OpenSubtitles, on the other hand, comprises discourse that at least tries to imitate everyday interactions. We can thus expect a higher share of the cognitive process of explanation presumably correlating with cognitive parentheticals in German.

In a second step, we thus have to take a closer look at the possible discourse functions involved with visual and cognitive parentheticals in German.

3 Parentheticals of perception and cognition in German – different discourse functions?

There are three different types of discourse functions that are discussed in the literature on our constructions of departure, English *you see* and French *tu vois* (literature on German *siehst du* is – maybe unsurprisingly, considering the small amount of data and presumably limited discourse functions – as good as non-existent). The first can be entitled as 'interpersonal', i.e. "claim[ing] the addressee's attention (Quirk et al., 1985) or "keep[ing] control over the interaction, maintain[ing] contact with the interlocutor" ('Interpersonal monitoring', Crible & Degand, 2019: 27/35). The second can be summarized under the term 'segmentation', i.e. marking transitions between information units or arguments (Erman, 1987 on English *you see*, see also Bolly, 2012:10/11 on French *tu vois* as a 'ponctuant'). Finally, we have the 'explanation/justification' function as quoted from Erman (1987) above.

As I have argued above, however, I consider explanation and justification to be two different cognitive processes that – at least in German – seem to map with different parenthetical markers. For our purposes, these two should thus be considered as two separate functions that can be distinguished using a paraphrase test along the lines of (7) and (8) above: 'Explanation/Reason' with the paraphrase *because/the explanation for that is* vis-à-vis 'Justification/Evidence' with the paraphrase *this is evidence for the fact that* (or the French equivalent paraphrases for the French part of the data).

The other two functions, 'interpersonal' and 'segmentation' do not involve cognitive processes, but are entirely meta-discursive, relating to the discourse structure and interaction, instead. As such they cannot be identified via paraphrase tests and are more subtle in nature. Examples with question-answer-pairs as in (11) above, for example, seem to be cases where *you see/tu vois* simply marks the transition from question to answer. Other examples, as in (13), seem to mark the beginning of a new, bigger discourse segment,

while also "maintain[ing] the contact with the interlocutor":

(13) Let' s see ... where was I? Oh, yes! The master. He was kind, **you see**. He brought me to our mutual acquaintance, Father Karras. Not too well at the time.
Ref: OPUS2; #326900403, en/1990/4253/77639_1of1.xml.gz

Since, at this point, it is unclear how to operationalize a distinction between these two functions (and the focus of the exploration lies on the functions involving cognitive processes), I group them into one category 'Segmentation/Interpersonal'.

There is one challenge for this part of the exploration of the data, however: The annotation of the presence of discourse markers other than *you see/tu vois* reveals that exactly half of the data provide the combination of *you see/tu vois* with other markers, e.g. English *well*, *and then*, or *but* as in (9) above, cf Figure 4.
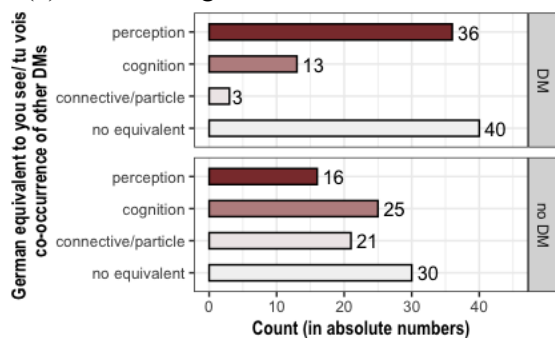

Figure 4: Co-occurrence of other discourse markers

This poses a challenge in so far as the presence of other discourse markers can block the application of the paraphrase test, cf. (14) and the failed attempt to paraphrase *you see* in (14'). Interestingly, omitting *you see* or any paraphrase altogether seems to be closest to the original meaning.

(14) A: Why, Captain John told me I could stay on my place as long as I wanted to. […]
B: Yeah, I know he did, Jeeter ... But **you see**, that land doesn't belong to us anymore.
Ref: OPUS2; #184020757, en/1941/25528/3671553_1of1.xml.gz

(14') I know he told you that you could stay. But **#because / #the explanation for that is that / #this is evidence for the fact that** / ∅ that land doesn't belong to us anymore.

For now, the annotation of discourse functions via the paraphrase test thus has to be limited to the 92 sets of data where *you see/tu vois* is the only discourse marker present. Unfortunately, this leaves us with only 16 instances of German visual

parentheticals and 25 cognitive parentheticals. Additionally, the represented languages and types of corpora become slightly skewed with 48 instances from OpenSubtitles compared to 44 from Europarl, and 52 with English *you see* as a point of departure compared to 40 with French *tu vois*.

Nevertheless, the paraphrase test reveals an interesting pattern in terms of cognitive processes and verb types used in German. As Figure 5 shows, German cognitive parentheticals are primarily used to express an explanation process (84%), whereas visual parentheticals primarily occur with the process of Justification/Evidence (62.5%). This latter process is overall least frequent which makes the strong relation with visual parentheticals in German all the more interesting. The observation that German equivalents in form of a connective or particle are used exclusively to express an Explanation/Reason process hardly seems surprising considering that these are mostly causal connectives and particles such as *denn* ('because') and *nämlich* ('namely') as in the French example in (12) above.
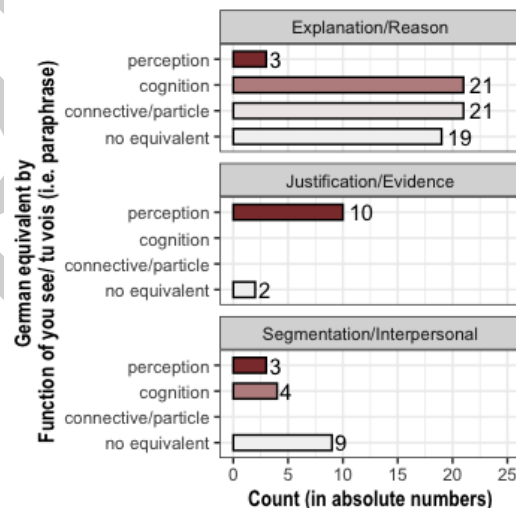

Figure 5: German equivalent by discourse function

4 Discussion

We set out to explore whether the corpus data reveal a pattern considering German parenthetical markers between perception and cognition. The claim was that the use of German visual vis-à-vis cognitive parentheticals as equivalent to *you see/tu vois* depends on the cognitive process involved – an Explanation/Reason that can be paraphrased by *because* or *the explanation for that is* goes in hand with German cognitive parentheticals such as *weißt du* or *verstehen Sie*, and a Justification/Evidence process that can be

paraphrased by *this is evidence for the fact that* goes in hand with visual parentheticals such as *siehst du* or *wie Sie sehen*.

A first clue of such a relationship can be observed from the distribution of German equivalents among the different types of corpora. As shown in Figure 3, the Europarl corpora are the origin for 80.7% of the perception-verb-equivalents, while 94.7% of the cognitive-verb-equivalents were found in the OpenSubtitles corpora. The explanation for this distribution was the varying types of discourse represented in the different corpora: the argumentative nature of political discourse represented in Europarl goes in hand with a process of Justification/Evidence, while the everyday discourse in OpenSubtitles would involve more Explanation/Reason processes. Of course, this neither means that there are no explanations in political discourse, nor that everyday conversation lacks justification. However, the high number of connectives and particles used in the German equivalents of the Europarl data (91.7% of this category) and the observation that these are causal in nature suggests that in political discourse the preferred way to express explanations in German are causal connectives and particles, while parenthetical markers are the preferred choice for this process in everyday discourse. A closer look at these different cognitive processes using the paraphrase test to distinguish the three discourse functions Explanation/Reason, Justification/Evidence, and Segmentation/Interpersonal supports these observations, as illustrated in Figure 5.

However, this difference between visual and cognitive parentheticals only concerns the German part of the data. We can thus derive that English and French parenthetical markers involving verbs of perception seem to be situated at different positions in the perception-cognition-continuum described by Viberg (2015, cited above) than their German counterparts: English *you see* and French *tu vois* cover the whole range from 'raw' visual perception over the visually inspired cognitive process of justification all the way to the complex cognitive process of explanation. The German *siehst du/wie du siehst*, on the other hand, only covers the first two functions, or, as Günthner (2017, cited above) put it for the imperative *guck mal* ('look'), merely accomplished the shift from actual visual perception to the abstract perception of argumentation structure (in the sense of 'Look,

this is the evidence for my argument!'). The shift to cognitive parentheticals such as *weißt/ verstehst du* in German when expressing the more complex cognitive process of explanation interestingly matches the Common Ground related functions ascribed to both German *weißt du* (cf. Günther 2017, cited above) and English *you know* (cf. Erman, 1987 and Schiffrin, 1987, cited above). In (15), for example, A's explanation that it's a surprise is not exactly presented as unexpected information, but can easily be accommodated (even without further context). This way of making "the addressee accept parts of the information conveyed as common ground" (Erman, 1987, cited above) is perfectly expressed by the German cognitive equivalent *verstehen Sie* ('(do) you understand').

(15) EN   A: I don't want them to see me arrive.
           B: Oh.
           A: It's a surprise, **you see**.
    DE   A: **Verstehen Sie**, eine Überraschung,

This raises the question whether, in this use as a marker of Explanation/Reason, English and French *you see/tu vois* and *you know/tu sais* are exchangeable. If we follow Brinton (2008) and many others in the assumption that the original semantics of verbs is bleached on their path towards parenthetical discourse markers, this could be the case. The 'persistence' (i.e. leftover meaning reflected in distributional constraints, cf. Hopper, 1991) in this case, however, might relate to a difference in what kind of information is added to the common ground: English *you see* might involve more objective information, whereas *you know* (in line with Günthner's, 2017 suggestion for German *weißt du*) could be used for (inter)subjective information instead.

Finally, the co-occurrence of parenthetical markers with other discourse markers provides interesting pointers for further research. Since the presence of other markers, especially connectives such as *and (then)* or *but*, impedes the application of the paraphrase test (cf. (14') above), I had to ignore half of the data for this part of the explorative study. As examples such as the following show, however, these cases might be particularly insightful for the as yet somewhat evasive function of Segmentation/Interpersonal, and for the analysis of the multifunctional contribution of discourse markers in general. Example (16) raises the question whether *you see*

simply complements the markers it co-occurs with: both *well* and *you see* seem to simply fulfill the same function, i.e. marking the transition from question to complex answer (potentially involving some hesitation as to where to begin). The altered version of (14) shown in (17), on the other hand, seems to provide the opposite case: it seems that the contrastive *but* and the parenthetical *you see* relate two different arguments – *but* marks the contrast between the inferences drawn from the first and second utterances and an implicit argument ('he told you that you could stay' → *you can stay*; 'this land isn't ours' → *you can't*), whereas *you see* marks the second utterance as an explanation for this implicit contrast-argument ('you can't stay here, because this land doesn't belong to us anymore').[5]

(16) A: What made you decide to become a lawyer?
    B: **Well, you see**, it's like this, Miss Roy. A white boy, he can take most any kind of job and improve himself. …
       Ref: OPUS2; #185696493, en/1942/37804/4037766_1of1.xml.gz

(17) I know he told you that you could stay. **But you see**, that land doesn't belong to us anymore.

To conclude, the explorative approach to parallel corpus data on English *you see* and French *tu vois* and its German equivalents not only provides us with interesting observations on parenthetical markers in the perception-cognition-continuum. It also points us towards important questions for future research: How can we operationalize discourse functions involving processes of different cognitive complexity (such as Justification/Evidence and Explanation/Reason) and those involving meta-discursive functions (such as Segmentation/Interpersonal)? What overlap do visual and cognitive parentheticals provide and what do they add to the Common Ground? Finally, the co-occurrence and interaction of parenthetical markers with other discourse markers prompts an analysis of the multifunctional contribution of discourse markers in general (for example via the two-dimensional model for discourse markers suggested by Crible and Degand, 2019) and the impact of inferences on discourse structure and the interpretation of discourse markers.

References

Bolly, Catherine. 2012. Du verbe de perception visuelle au marqueur parenthétique "tu vois": Grammaticalisation et changement linguistique. In: *Journal of French Language Studies*, 22(02), pages 143-164. (http://hdl.handle.net/2078.1/74646)

Brinton, Laurel J. 2008. *The comment clause in English*. Cambridge: Cambridge University Press.

Crible, Ludivine & Liesbeth Degand. 2019. Domains and functions: A two-dimensional account of discourse markers. *Discours. Revue de linguistique, psycholinguistique et informatique.* 24. (https://doi.org/10.4000/discours.9997)

Erman, Britt. 1987. *Pragmatic expressions in English: A study of 'you know', 'you see' and 'I mean' in face-to-face conversation.* Stockholm: Almqvist & Wiksell.

Günthner, Susanne. 2017. Diskursmarker in der Interaktion – Formen und Funktionen univerbierter *guck mal*- und *weißt du*-Konstruktionen. In Hardarik Blühdorn, Arnulf Deppermann, Henrieke Helmer & Thomas Spranz-Fogasy (eds.), *Diskursmarker im Deutschen: Reflexionen und Analysen*, 103–130. Göttingen: Verlag für Gesprächsforschung.

Hopper, Paul J. 1991. On Some Principles of Grammaticalization. In Elisabeth Traugott & Bernd Heine (eds.), *Approaches to Grammaticalization*, 17–35. John Benjamins Publishing.

Müller, Mathias & Martin Volk. 2013. Statistical machine translation of subtitles: From OpenSubtitles to TED. In Iryna Gurevych, Chris Biemann & Torsten Zesch (eds.), *Language processing and knowledge in the Web*, 132–138. Springer. (https://doi.org/10.1007/978-3-642-40722-2_14)

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language.* London: Longman Group Ltd.

Schiffrin, Deborah. 1987. *Discourse markers*. No. 5. Cambridge: Cambridge University Press.

Sweetser, Eve. 1990. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*, vol. 54. Cambridge University Press.

Viberg, Åke. 2015. Sensation, perception and cognition: Swedish in a typological-contrastive perspective." *Functions of language* 22.1, pages 96-131.

---

[5] I thank Merel C.J. Scholman for pointing that out in a fruitful discussion of this and similar corpus examples during the DiscAnn workshop.