

Toward a Multilingual Connective Database: Aligning German/French Concessive Connectives

Sophia Rauh¹, Karolina Zaczynska¹, and Peter Bourgonje^{1,2}

¹Applied CL Discourse Research Lab, Department of Linguistics, University of Potsdam

firstname.lastname@uni-potsdam.de

²Language Science and Technology, Saarland University

lastname@coli.uni-saarland.de

Abstract

We report on experiments to align discourse connectives from two language-specific connective lexicons (German and French) by their relation sense. In this case study, we focus on *concessive* connectives, and align them using a parallel corpus. The ultimate goal is to arrive at bi- (or multi-)lingual connective lexicons, that at the same time provide insights on the “semantic space” that connectives cover in different languages.

1 Introduction

A typical way to establish coherence in a text is through the use of *discourse connectives*. Such markers (single words or – potentially discontinuous – phrases) convey a specific relation; *contrast* (e.g., “but”), *contingency* (e.g., “if...then”) or *cause* (e.g., “therefore”) that links propositions in the text. They can be ambiguous in two ways, and can either signal a discourse relation between two propositions (1) or sentential reading (2).

- (1) It would have made a dreadfully ugly child, *but* it makes rather a handsome pig. (Carroll, 1893)
- (2) “I beg your pardon?” said the Mouse, frowning, *but* very politely. (Carroll, 1893)

In addition, certain connectives can express multiple senses. In (3), *once* signals a temporal relation, whereas in (4), it signals a conditional relation.

- (3) *Once* it gets there, a company can do with it what it wishes. (wsj_0989 (Marcus et al., 1993))
- (4) Normally, *once* the underlying investment is suspended from trading, the options on those investments also don’t trade. (wsj_1962 (Marcus et al., 1993))

Discourse relations can also be realized implicitly and expressed, for example, by syntactic parallelism, layout, but explicit discourse markers are considered important indicators of coherence relations as explored in various frameworks like Rhetorical Structure Theory (RST) (Mann and Thompson, 1988) and Segmented Discourse Representation Theory (SDRT) (Asher and Lascarides, 2005). After attempts to exhaustively list such markers (Knott and Dale, 1994), specific discourse connective or discourse marker lexicons started to emerge, with the first documented lexicon being for German (Stede and Umbach, 1998), and following ones for French (Roze et al., 2012), Italian (Feltracco et al., 2016), Czech (Mírovský et al., 2016) and several other languages.

Once language-specific lexicons are created and augmented with semantic information, parallels can be drawn based on the *distribution* of relation senses¹ across languages. In addition, since connectives can pose challenges to translators and (L2) language learners, having a layer over language-specific lexicons that aligns entries across languages can be a useful resource. Earlier work in this direction has been carried out by Bourgonje et al. (2017). We use a similar approach, but work on French and German, and base our work on LexConn (Roze et al., 2012), DiMLex (Stede and Umbach, 1998) and a parallel corpus. The main contribution of this paper is to present the results of a case study on aligning French and German connectives that can signal a *concessive* relation. Our code is made publicly available².

In Section 2 we summarize related work on connective lexicons. Section 3 explains the corpus and alignment procedure. Section 4 presents the results, and Section 5 sums up our main findings.

¹Inventories of relation senses for a specific paradigm or theory are, presumably, language-independent.

²https://github.com/SophiaRauh/fr_de_connectives_alignment

2 Related Work

Several language-specific lexicons are available online. In addition to the ones mentioned in Section 1, lexicons exist for English (Das et al., 2018), Dutch (Bourgonje et al., 2018), Bangla (Das et al., 2020), Portuguese (Mendes and Lejeune, 2016), Nigerian Pidgin (Marchal et al., 2021) and Turkish (Zeyrek and Başıbüyük, 2019). These lexicons are conveniently bundled on the online platform Connective-Lex³ (Stede et al., 2019). While this platform already allows multi-lingual comparison of connective groups (grouping by part-of-speech tag or relation sense), with our contribution we aim to expand this multi-lingual aspect to individual connectives.

The lexicons differ slightly with regard to their take on connectives (for example, what syntactic classes to include, and how to encode morphological variation). A comprehensive discussion is offered by (Danlos et al., 2018). We like to note that it is exactly the kind of cross-lingual investigation of connectives we are reporting on in this paper that allows subtle differences to surface, and enables refinement of the understanding and definition of connectives.

3 Method & Data

3.1 Lexicons and Parallel Corpus

Our starting point is the list of entries from DiMLex and LexConn, both of which are available online⁴. We follow Bourgonje et al. (2017) by focussing on concessive connectives for this case study. While the discourse senses of the German connective lexicon are based on the Penn Discourse Treebank (PDTB) (Webber et al., 2019) senses, the French discourse relations are an extended version of SDRT. Both include the relation *concession*, though interestingly, German concessive connectives frequently align with *violation* in French, which is equivalent to the PDTB sense *exception*.

We used the Europarl parallel corpus (Koehn, 2005), as the translations are curated, which avoids the risk of including automatically created low-quality translations. The French part of the corpus consists of 63.2 million tokens and the German part consists of 54.6 million tokens. For the word

³<http://connective-lex.info/> (URLs were all last accessed on 2023-05-06.)

⁴<http://www.linguist.univ-paris-diderot.fr/~croze/>,
<https://github.com/discourse-lab/dimlex>

alignment, our data was tokenized and converted to lower case.

3.2 (Semi-) Automated Word Alignment Procedure

For the word alignment we used eflomal⁵, which is based on efmara (Östling and Tiedemann, 2016). The alignments are saved in “Pharaoh” format, i.e. for the (pre-tokenized) input

```
schwarzes Haus || maison noire  
the representation “0-1 1-0” is returned, indicating that the first (0-indexed) token in the source is aligned to the second token in the target, and the second token in the source is aligned to the first token in the target. NULL alignments are not present in the output.
```

Once these alignments were calculated for the entire corpus, we used both the German DiMLex connectives and the French LexConn connectives as seed lists to extract the probability that a certain connective is aligned to a word or phrase in the target language.⁶ This process is straightforward for single-word connectives. For multi-word (phrasal) connectives, the alignment probabilities are obtained by concatenating the single-word alignments that constitute the phrase. The results were stored in a JSON file and are further processed in a semi-automated way:

1) If contractions of prepositions and articles occur at the end of a phrase, they are replaced with the preposition only, since the articles are not part of the connective. For example, the contracted German word *zur* (“to the”) is replaced by *zu* (“to”) and the French contraction *aux* (“in the”) is substituted with *à* (“in”).

2) Since connectives are frequently (sub-)clause initial, hence alignments may include punctuation, punctuation is removed, i.e., “*weil*” (“because”) becomes “*weil*”.

3) If tokens were NULL-aligned, we included an empty string as alignment, as this influences alignment probabilities. Words that were aligned to punctuation marks were aggregated with the empty string placeholder.

⁵<https://github.com/robertostling/eflomal>

⁶Since the values are extracted from the entire corpus, they are not probabilities, but actual counts converted into floating point values by dividing by the total frequency. We use the term probability throughout the rest of this paper, though, since we interpret this number as the probability that a certain word or phrase in the source language is aligned to a certain word or phrase in the target language.

This slightly modified version of the extracted word alignments was stored in a dictionary, in which we then proceeded to look up connectives from source to target language.

3.3 (Semi-) Manual Filtering

Looking up connectives in our dictionary resulted in several incorrect or irrelevant target words or phrases. Many of these could be discarded in a semi-automated way.

First, for some instances, the alignment probabilities to reasonable candidates were very low. For example, after the above mentioned adjustments, *dabei* (“thereby/at that”) aligned to an empty string in 34% of cases, and to *en*, *il*, *à* and *ce* in 7, 5, 3 and 2% of cases, respectively. This might be due to eflomal alignment errors, or could be related to the frequency of sentential instances (e.g., example (2) in Section 1) far outweighing the frequency of discourse reading instances (e.g., example (1) in Section 1) for some connectives.

In (5), for example, *dabei* does not have a connective reading and it is translated with *dans ce processus* (“in this process”).

(5) Inwieweit wird das Europäische Parlament *dabei* eine Rolle spielen können?

Dans quelle mesure le Parlement européen pourra-t-il jouer un rôle *dans ce processus*?

To what extent will the European Parliament be able to play a role *in this (process)*?

Using the adjusted alignment probabilities, we filter out all words and phrases below a certain threshold. Due to the concatenation process to arrive at phrase alignment probabilities, we found that working with two different threshold values (one for single word connectives, one for phrasal connectives) worked best. In addition, we use a combination of relative and absolute thresholds. First, all single word connectives with a probability below 2.1%, and all phrasal connectives with a probability below 1.4%, were discarded. Because some very low-frequent connectives can have a relatively high probability, we furthermore discarded connectives below an absolute count in our corpus (20 for single words, 10 for phrases).

Second, results for phrasal connectives were often only partially relevant. For example, for the French connective *alors même que* (“even though”), the German phrase *obwohl die* (“although the”) was among the candidates, whereas the relevant

German connective would be only *obwohl* (“although”). These only partially relevant alignments could often be filtered out on syntactic grounds, by looking for prepositions, articles and pronouns. In addition, phrasal connectives led to incomplete target phrases. For the French connective *c’est pourquoi* (“that is why”), we found *c’. . . pourquoi* among the alignments. If we found the complete phrase among the alignment results as well, these incomplete alignments were removed from the list of candidates. Some phrasal connectives truly are discontinuous (e.g., *entweder. . . oder* (“either. . . or”)), while for others, the connective was not discontinuous but the correct/relevant alignment was just not in the set of results. One example is the 4-token connective *soit dit en passant* (“by the way/incidentally”), for which only *soit dit . . . passant* was among our results. This processing of phrasal connectives therefore had to be done in a manual way.

3.4 Augmentation

The combination of semi-automated filtering and manual curation of the results described above mainly deleted irrelevant candidates, and completed some partially correct ones. Since word alignments are extracted from parallel sentences (hence do not go beyond sentence boundary), we constructed sentence tri-grams and also extracted word alignments from those. This procedure lead to further completion of candidates.

Furthermore, this manual augmentation step involved weeding out non-connective, or non-concessive candidates. For example, the concessive connective *entgegen* (“contrary to”) was aligned to *contre* (“against”). Looking at the sentences revealed that *entgegen* does not have a connective reading when aligned to *contre*, which is also not a connective.

(6) Das Volk hat das Recht, innerhalb der Grenzen des Gesetzes zu demonstrieren, wenn es das Gefühl hat, dass die Regierung *entgegen* ihrer Interessen handelt.

La population est autorisée à manifester dans les limites de la législation lorsqu’elle estime que le gouvernement agit *contre* ses intérêts. The population has the right to demonstrate within the limits of the law when it feels that the government is acting *against* its interests.

Some candidates were excluded on these grounds. Finally, other candidates were deleted,

modified or completed (for missing particles) based on intuition. After this final curation of the candidates, we arrived at a list of aligned French connectives for the German seed list, and vice versa. We projected the final list of the target language connectives back onto the source once more, to see if we would get any additional results. In principle, this procedure could be repeated until no more new instances are found. Due to the amount of manual labour involved in the process though, we stopped after 3 “turns” (from French to German, back to French, and then back to German again).

4 Results & Discussion

Recall that we start with all connectives that have *concession* as their second-level sense in the PDTB3 Sense Hierarchy. The final alignments are included in Appendix A, Tables 1 and 2, where “-” indicates an empty alignment. The parentheses in the left column contain the absolute occurrence of the connectives in the corpus, whereas those in the right column indicate the relative occurrence of the aligned connectives. To get an overview of the distribution and the degree of ambiguity (i.e., different senses that groups of connectives can express), we include Figures 2 and 3 in Appendix A. The diagrams show which discourse relations align with which based on the connectives of the final alignment. For comparison, the SDRT senses of the French connectives are mapped to PDTB3 senses using the mapping included in Figure 1 in Appendix A.

Since many connectives can express multiple senses, Figures 2 and 3 also include second-level senses other than just *concessive*; we group all connectives by the set of senses they can express. Generally, Figure 3 looks much more straightforward; the set of connectives that can (only) signal concession map to a set in German that also exclusively signals concession, and the ambiguous sets map to each other relatively neatly. Figure 2 is much less straightforward. The set that exclusively signals concession in German maps to a much wider range of senses in French. A case in point is “dennoch”, which can only signal *concession*, which is aligned to “néanmoins” and “cependant” (*exception*), “pourtant” (*exception* or *concession*) and “mais” (*contrast* or *exception*). It is interesting to further look into whether particular corpus examples of “dennoch” also carry some aspect of the different senses of the aligned connectives in French.

For example, “dennoch” might be relatively ambiguous, as its “semantic space” (for lack of a better description) is covered by several different connectives in French. These semantic spaces could surface through clusters of connectives, which can be explored in this bi-lingual setup. The German connectives “allerdings, dennoch, doch, gleichwohl, jedoch” seem to constitute one potential example of such a cluster, and map to “cependant, néanmoins, pourtant, mais”, and “obgleich, obwohl, wenn auch, wenngleich” mostly map to “bien que, même si, alors que”. While the PDTB senses are already such clusters in themselves, our approach might lead to a more fine-grained classification or grouping of individual connectives’ meanings. Furthermore, interestingly, there seem to be asymmetries in the mapping: The German “aber” frequently maps only to “mais” and “cependant”. In reverse, however, “mais” maps to a larger set of German connectives (“aber, sondern, doch, jedoch”). While the reason for some of these asymmetries might just be low frequency, both “aber” and “mais” are fairly common connectives, indicating that this might not just be an artefact of the data we used.

In terms of future work, we plan to include connective disambiguation modules to separate connective instances from their sentential interpretation surface forms. For German, a connective classifier has been developed (Bourgonje and Stede, 2018). To the best of our knowledge, no such (pre-trained) classifier is available for French, so for this language, we consider the use of annotation projection (Sluyter-Gäthje et al., 2020).

Furthermore, since the performance of our word aligner is critical for downstream processing, it would also be interesting to evaluate this module in isolation by creating a gold set based on our data.

5 Conclusion

We present work on aligning *concessive* connectives in German and French, using word-alignments extracted from a parallel corpus. Our approach is semi-automated and the code is made available on GitHub. We provide some first insights on how particular relation sense groups are covered by the two languages. In addition to validating the mono-lingual connective lexicons, we hope that this contributes to our ultimate goal of providing insights on how discourse relation senses are covered in different languages through explicit markers (i.e., discourse connectives).

References

- Nicholas Asher and Alex Lascarides. 2005. Logics of conversation. In *Studies in Natural Language Processing*.
- Peter Bourgonje, Yulia Grishina, and Manfred Stede. 2017. Toward a bilingual lexical database on connectives: Exploiting a German/Italian parallel corpus. In *Proceedings of the Fourth Italian Conference on Computational Linguistics*, Rome, Italy.
- Peter Bourgonje, Jet Hoek, Jacqueline Evers-Vermeul, Gisela Redeker, Ted Sanders, and Manfred Stede. 2018. Constructing a Lexicon of Dutch Discourse Connectives. *Computational Linguistics in the Netherlands Journal*, 8:163–175.
- Peter Bourgonje and Manfred Stede. 2018. [Identifying explicit discourse connectives in German](#). In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 327–331, Melbourne, Australia. Association for Computational Linguistics.
- Lewis Carroll. 1893. *Alice's adventures in Wonderland*. T. Y. Crowell & co, New York, Boston.
- Laurence Danlos, Katerina Rysova, Magdalena Rysova, and Manfred Stede. 2018. [Primary and secondary discourse connectives: definitions and lexicons](#). *Dialogue and Discourse*, 9(1):50–78.
- Debopam Das, Tatjana Scheffler, Peter Bourgonje, and Manfred Stede. 2018. [Constructing a lexicon of English discourse connectives](#). In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 360–365, Melbourne, Australia. Association for Computational Linguistics.
- Debopam Das, Manfred Stede, Soumya Sankar Ghosh, and Lahari Chatterjee. 2020. [DiMLex-bangla: A lexicon of bangla discourse connectives](#). In *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC'20)*, pages 1097–1102, Marseille, France. European Language Resources Association (ELRA).
- Anna Feltracco, Elisabetta Jezek, Bernardo Magnini, and Manfred Stede. 2016. Lico: A lexicon of Italian connectives. In *Proceedings of the 3rd Italian Conference on Computational Linguistics (CLiC-it)*, pages 141–145, Napoli, Italy.
- Alistair Knott and Robert Dale. 1994. Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18:35–62.
- Philipp Koehn. 2005. [Europarl: A parallel corpus for statistical machine translation](#). In *Proceedings of Machine Translation Summit X: Papers*, pages 79–86, Phuket, Thailand.
- William Mann and Sandra Thompson. 1988. Rhetorical Structure Theory: Towards a functional theory of text organization. *TEXT*, 8:243–281.
- Marian Marchal, Merel Scholman, and Vera Demberg. 2021. [Semi-automatic discourse annotation in a low-resource language: Developing a connective lexicon for Nigerian Pidgin](#). In *Proceedings of the 2nd Workshop on Computational Approaches to Discourse*, pages 84–94, Punta Cana, Dominican Republic and Online. Association for Computational Linguistics.
- Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. [Building a large annotated corpus of English: The Penn Treebank](#). *Computational Linguistics*, 19(2):313–330.
- Amália Mendes and Pierre Lejeune. 2016. LDM-PT. A Portuguese Lexicon of Discourse Markers. In *Conference Handbook of TextLink – Structuring Discourse in Multilingual Europe Second Action Conference*, pages 89–92, Budapest, Hungary.
- Jiří Mírovský, Pavlína Jínová, Magdaléna Rysová, and Lucie Poláková. 2016. [Designing CzeDLex – a lexicon of Czech discourse connectives](#). In *Proceedings of the 30th Pacific Asia Conference on Language, Information and Computation: Posters*, pages 449–457, Seoul, South Korea.
- Charlotte Roze, Laurence Danlos, and Philippe Muller. 2012. [LEXCONN: A French lexicon of discourse connectives](#). *Discours. Revue de linguistique, psycholinguistique et informatique*, 10.
- Henny Sluyter-Gäthje, Peter Bourgonje, and Manfred Stede. 2020. [Shallow discourse parsing for under-resourced languages: Combining machine translation and annotation projection](#). In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC'20)*, pages 1044–1050, Marseille, France. European Language Resources Association (ELRA).
- Manfred Stede, Tatjana Scheffler, and Amalia Mendes. 2019. [Connective-lex: A web-based multilingual lexical resource for connectives](#). *Discours. Revue de linguistique, psycholinguistique et informatique*.
- Manfred Stede and Carla Umbach. 1998. [DiMLex: A lexicon of discourse markers for text generation and understanding](#). In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 2*, pages 1238–1242, Montreal, Quebec, Canada. Association for Computational Linguistics.
- Bonnie Webber, Rashmi Prasad, Alan Lee, and Aravind Joshi. 2019. [The Penn Discourse Treebank 3.0 Annotation Manual](#).
- Deniz Zeyrek and Kezban Başibüyük. 2019. [TCL - a lexicon of Turkish discourse connectives](#). In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 73–81, Florence, Italy. Association for Computational Linguistics.
- Robert Östling and Jörg Tiedemann. 2016. [Efficient word alignment with markov chain monte carlo](#). *The Prague Bulletin of Mathematical Linguistics*, 106.

A Appendix

```
{
  "alternation": ["EXPANSION:Disjunction"],
  "background": ["TEMPORAL:Synchronous"],
  "background-inverse": ["TEMPORAL:Asynchronous:Succession"],
  "concession": ["COMPARISON:Concession"],
  "condition": ["CONTINGENCY:Condition"],
  "consequence": ["CONTINGENCY:Condition"],
  "continuation": ["EXPANSION:Conjunction"],
  "contrast": ["COMPARISON:Contrast"],
  "detachment": ["EXPANSION:Exception"],
  "digression": ["EXPANSION:Conjunction"],
  "elaboration": ["EXPANSION:Level-of-detail"],
  "evidence": ["EXPANSION:Conjunction"],
  "explanation": ["EXPANSION:Manner",
    "EXPANSION:Level-of-detail",
    "CONTINGENCY:Cause:Reason"],
  "explanation*": ["CONTINGENCY:Cause+belief"],
  "flashback": ["TEMPORAL:Asynchronous:Succession"],
  "goal": ["CONTINGENCY:Purpose"],
  "narration": ["TEMPORAL:Asynchronous:Precedence",
    "EXPANSION:Conjunction"],
  "parallel": ["COMPARISON:Similarity"],
  "rephrasing": ["EXPANSION:Equivalence"],
  "result": ["CONTINGENCY:Cause:Result"],
  "result*": ["CONTINGENCY:Cause:Result+belief",
    "CONTINGENCY:Cause:Result+speechact"],
  "summary": ["EXPANSION:Level-of-detail:Arg2-as-detail"],
  "temploc": ["TEMPORAL"],
  "violation": ["EXPANSION:Exception"]
}
```

Figure 1: SDRT to PDTB Sense Mapping

DE Connective (frequency)	FR Connective(s)
aber (98898)	mais (0.65), - (0.09), cependant (0.04)
abgesehen davon (553)	cela dit (0.06), - (0.05), par ailleurs (0.03), ceci dit (0.02)
allerdings (14935)	cependant (0.18), mais (0.16), - (0.08), néanmoins (0.06), pourtant (0.02)
dennoch (7920)	néanmoins (0.19), cependant (0.13), pourtant (0.09), mais (0.08), - (0.07)
dessen ungeachtet (187)	néanmoins (0.12)
doch (30068)	mais (0.38), - (0.2), cependant (0.03), pourtant (0.03)
gleichwohl (1714)	cependant (0.12), néanmoins (0.12), mais (0.09), - (0.08), pourtant (0.07)
immerhin (1023)	- (0.26), après tout (0.11), pourtant (0.05), quand même (0.03), tout de même (0.02), au moins (0.02)
jedoch (43525)	mais (0.26), cependant (0.16), - (0.09), néanmoins (0.05), pourtant (0.03)
nebenbei gesagt (59)	soit dit en passant (0.24)
nichtsdestotrotz (618)	néanmoins (0.38), cependant (0.09), malgré tout (0.02)
nichtsdestoweniger (213)	néanmoins (0.4)
obgleich (1612)	bien que (0.18), même si (0.15), - (0.09), alors que (0.05), mais (0.05), bien qu' (0.05), même s' (0.03), alors qu' (0.02)
obwohl (10904)	bien que (0.19), même si (0.12), bien qu' (0.06), - (0.06), alors que (0.06), alors qu' (0.03), même s' (0.02)
trotzdem (3585)	néanmoins (0.18), pourtant (0.1), cependant (0.09), - (0.08), mais (0.05), malgré tout (0.04), quand même (0.03)
wenn auch (1849)	même si (0.09), bien que (0.06), - (0.06), mais (0.05), bien qu' (0.03), quoique (0.03), même s' (0.01)
wenngleich (1463)	même si (0.23), bien que (0.15), mais (0.06), - (0.05), même s' (0.05), bien qu' (0.04)

Table 1: German to French Connective Alignments

FR Connective (frequency)	DE Connective(s)
alors même que (542)	während (0.14), obwohl (0.11), - (0.03)
alors qu' (3466)	obwohl (0.14), während (0.12), - (0.09), als (0.03)
alors que (10341)	während (0.22), obwohl (0.09), - (0.07), da (0.03)
après tout (1870)	schließlich (0.3), - (0.09), immerhin (0.08), denn (0.03), doch (0.02), nämlich (0.02)
bien qu' (3413)	obwohl (0.25), - (0.07), obgleich (0.03), zwar (0.02), wenn auch (0.02)
bien que (9678)	obwohl (0.25), - (0.06), obgleich (0.03), wenngleich (0.03)
ceci dit (466)	- (0.14), abgesehen davon (0.03)
cela dit (1163)	- (0.14), allerdings (0.05), aber (0.05), davon abgesehen (0.03), jedoch (0.03), dennoch (0.03), abgesehen davon (0.03), doch (0.02)
cependant (19138)	jedoch (0.36), aber (0.22), allerdings (0.14), - (0.09), doch (0.05), dennoch (0.05)
en dépit du fait que (159)	obwohl (0.17)
mais (142830)	aber (0.46), sondern (0.2), doch (0.09), jedoch (0.08), - (0.07)
malgré le fait qu' (85)	obwohl (0.28)
malgré le fait que (259)	obwohl (0.19)
malgré que (47)	obwohl (0.43)
malgré tout (1145)	trotzdem (0.14), dennoch (0.12), - (0.07), doch (0.05), jedoch (0.03)
même s' (2097)	obwohl (0.13), - (0.05), wenngleich (0.03), wenn auch (0.03), obgleich (0.02)
même si (9593)	obwohl (0.16), wenngleich (0.04), - (0.04), obgleich (0.03), zwar (0.03), wenn auch (0.02)
néanmoins (8521)	doch (0.23), dennoch (0.17), aber (0.16), allerdings (0.11), - (0.09), trotzdem (0.07), doch (0.05), nichtsdestotrotz (0.03), gleichwohl (0.02)
pourtant (5890)	jedoch (0.18), - (0.16), doch (0.14), aber (0.12), dennoch (0.11), allerdings (0.05), trotzdem (0.05), obwohl (0.02)
quand même (1480)	doch (0.15), - (0.13), trotzdem (0.08), dennoch (0.07), immerhin (0.03), auch (0.02)
quoique (413)	- (0.17), obwohl (0.15), aber (0.11), wenngleich (0.07)
soit dit en passant (277)	- (0.05), nebenbei gesagt (0.04)
tout de même (1628)	doch (0.16), - (0.12), immerhin (0.05), dennoch (0.04), trotzdem (0.04)

Table 2: French to German Connective Alignments

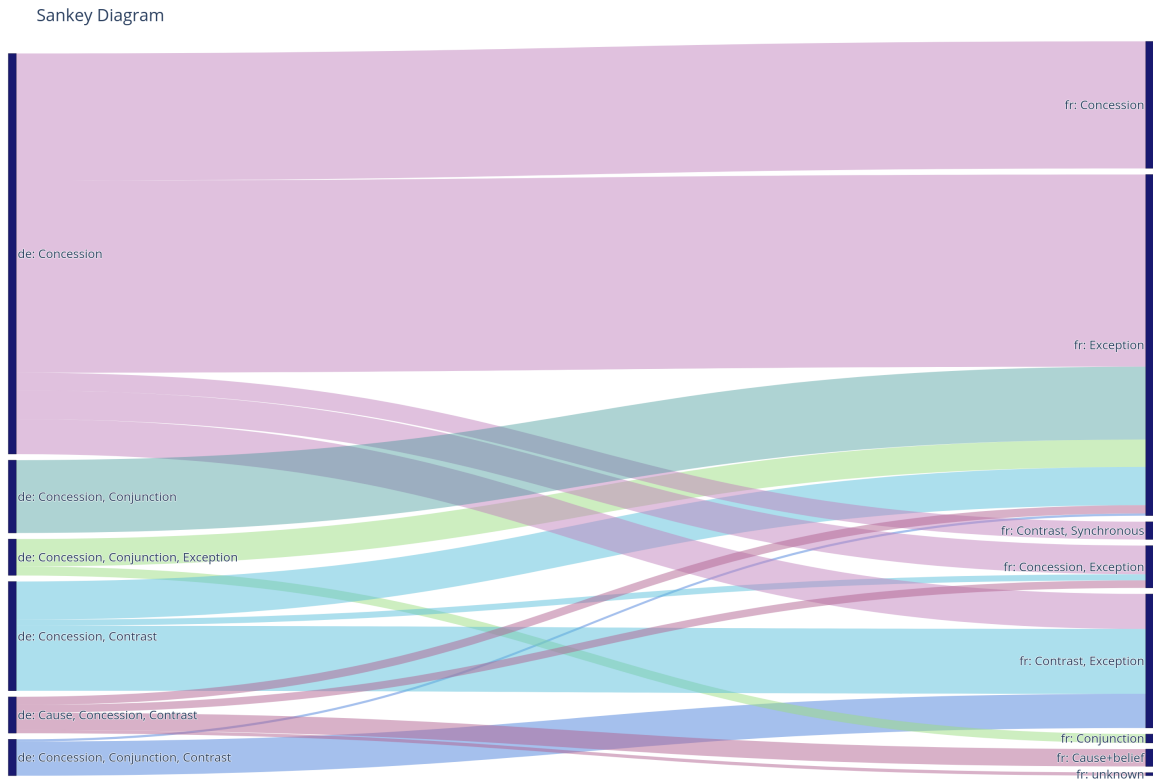


Figure 2: Mappings of German to French Connectives

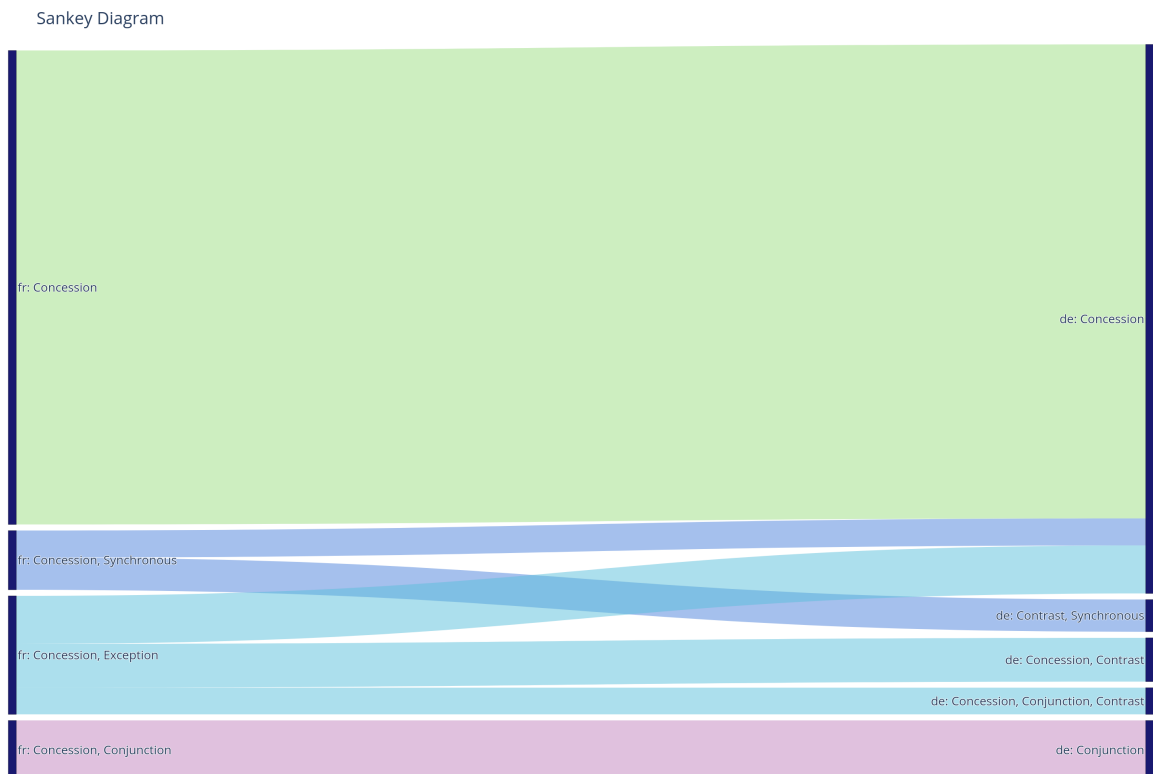


Figure 3: Mappings of French to German Connectives