Multiple Discourse Relations in English TED Talks and Their Translation into Lithuanian, Portuguese, and Turkish

Deniz Zeyrek[1], Giedrė Valūnaitė Oleškevičienė[2], Amália Mendes[3]

[1]Cognitive Science Dept., Grad. School of Informatics, Middle East Technical University,
[2]Institute of Humanities, Mykolas Romeris University,
[3]Center of Linguistics, University of Lisbon (CLUL),
dezeyrek@metu.edu.tr, gvalunaite@mruni.eu, amaliamendes@letras.ulisboa.pt

Abstract

This paper focuses on multiple discourse relations, which refer to more than one sense relation between a pair of discourse segments. It shows how they are realized in English texts and their translations into Lithuanian, Turkish, and Portuguese in TED Multilingual Bank, an annotated corpus of English TED transcripts and translations into multiple languages. The paper overviews the annotation procedure and shows the change and variation of multiple discourse relations in the translations, such as omitting the *and*-component of multiple relations. The cross-linguistically framed analysis reveals that while both senses of a multiple relation can be explicitly conveyed, the salient sense is generally rendered through an overt connective. Even when it is not overtly expressed, it remains inferable and annotated during the annotation stage. By describing the different discourse structures arising from multiple relations and highlighting the implicitated components in translation, the research contributes to the understanding of discourse, aims to raise the awareness of translators and translation educators, and bridge the gap between discourse analysis and translation.

Keywords: multilingual corpus, multiple discourse relations, translation, discourse connectives, implicitation

1. Introduction

Translation is the process of conveying the messages of the source language into the structure of the target language, while preserving the purpose and essence of the message. Translation studies can benefit from discourse analysis in "discovering patterns and systematicity in the choices made by a translator and for hypothesizing reasons behind these choices on the basis of detailed discourse analytic procedures" (House, 2015, p. 49). Hence, translation researchers can use the knowledge accumulating in discourse analysis to understand how text pieces are structured to maintain coherence.

Discourse relations (also called coherence relations or rhetorical relations) are one of the ways clauses or sentences are structured (Mann & Thompson, 1988; Knott & Sanders, 1998; Marcu, 2000; Asher and Lascarides, 2003, among others). They hold between clauses, groups of clauses, or sentences and are named after the senses they convey – comparison, contrast, contingency, elaboration. They are expressed by a range of linguistic devices, such as conjunctions (*and, but, so*), adverbials (*however, in addition*), or prepositional phrases (*in summary*). These words or word groups are called discourse operators, discourse markers, cue phrases, or discourse connectives (Fraser, 1999), the term we use in the current work. They express a two-place semantic relation relating text spans that have an abstract object interpretation (eventualities, propositions, facts), as depicted by Asher (2012), or are complete clauses, as argued by Pasch et al. (2003).

Discourse relations are among the fundamental notions that enable discourse and pragmatics researchers to understand how texts are organized beyond the sentence level. Lately, the basic blocks, or anchors of discourse relations, i.e., connectives, have been examined extensively, mainly focusing on single (*but, so, instead*) and complex connectives (*on the contrary*). However, connectives are also known to co-occur with other connectives. In English, an adverb (*otherwise, instead*) and a conjunction (*because, if, so*) or two adverbials (*previously, for example*) may co-occur, forming constructions referred to as multiple connectives (Webber et al., 2019).

Examples (1) - (2), both taken from British National Corpus (http://www.natcorp.ox.ac.uk), illustrate the phenomenon of multiple connectives. More generally, this situation is referred to as multiple discourse relations, a notion that refers to more than one sense relation that holds between a pair of discourse segments:

(1)   but I'm just not enough of a Facebook user. <u>So instead</u> I'm going to use data from a few kind souls around our company

(2)   Pamela, you are now in my power. <u>But if</u> you comply with my proposals, I will leave you.

In the first example, multiple relations are signaled by a conjunction and an adverb relating the exact text spans. The connective *so* signals the consequence of the fact that I am not using Facebook frequently, *instead* conveys how I'm going to replace my infrequent use of Facebook. In the second example, *but* raises the expectation of a contradiction, and the

125

contrary expectation is fulfilled immediately in the next segment by the entire conditional sentence. The connectives *but* and *if* relate the interpretations of different spans. The semantic relations that hold between the clauses, i.e. text spans that correspond to the arguments of connectives reflect different discourse structures associated with (1) and (2) as depicted in Figure 1 and 2.
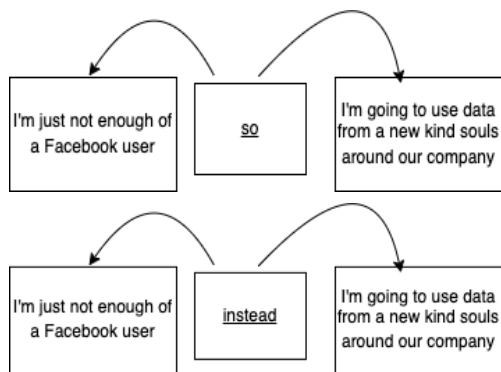

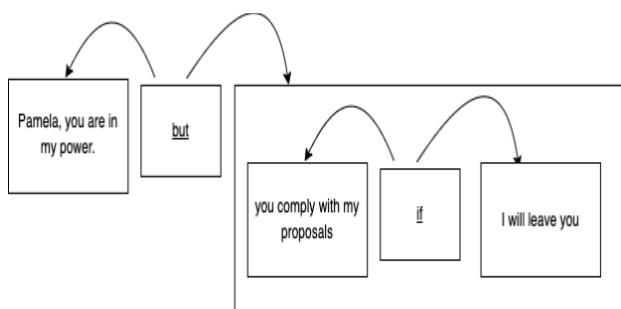Figure 1: Discourse structure of example (1)

Figure 2: Discourse structure of example (2)

As opposed to example (1), example (2) generates full embedding (Lee et al., 2006), where the relation anchored by the conjunction *if* is fully embedded within one argument of the conjunction *but*. Implicit in this approach is that semantics and pragmatics of discourse are derived compositionally from the structure exposed in the discourse relations between different parts of the text.

In the current work, we are concerned with multiple connectives that link the interpretation of exact text spans, as example (1) shows.[1] Examples like (2) fall out of our scope mainly because they derive discourse structures that need to be analysed separately. What also falls out of scope of multiple relations involves parallel connectives (*not only .. but also*, *on the one hand .. on the other hand*) since, in these cases, one text span presupposes the other, and both parts of the connective act together to relate the text spans, as described in the PDTB 2.0 annotation manual (Prasad et al., 2007).

Multiple connectives are challenging for translators. Word-for-word translations may lead to incorrect or

inappropriate results in target texts unless the context in which they occur is correctly interpreted. For example, the English connective sequence *but then* can function as a multiword connective conveying a single, concessive sense (see Fig. 3) though it can also function as a multiple connective conveying the contrastive/concessive sense followed by the temporal sense. This sequence will likely yield inappropriate translations if the human or machine translator misinterprets its meaning in the given context.
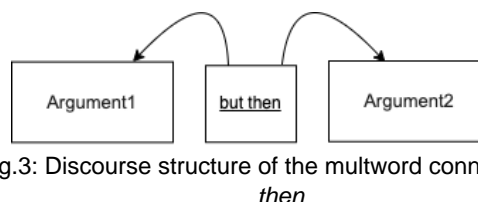

Fig.3: Discourse structure of the multword connective *but then*

Given the potential benefits of multiple connectives or, more generally, multiple discourse relations to discourse and translation research, we ask: (a) What is the occurrence of multiple discourse relations in the source language, English? (b) What is their variation and change in translated languages? Our data is drawn from TED-Multilingual Discourse Bank or TED-MDB's English, Lithuanian, Turkish, and Portuguese parts (Zeyrek et al., 2020).

The rest of the paper proceeds as follows: Section 2 aims to set the ground and summarizes the related work. It also describes the challenges of automatically extracting multiple connectives from raw texts with no discourse connective annotation. Section 3 presents the data and methodology. It explains the differences between different versions of TED-MDB and describes how multiple relations are spotted, annotated, and then checked for inconsistencies. Section 4 analyzes the annotated data quantitatively, and Section 5 discusses how multiple relations are translated to target texts with examples from the corpus. Finally, in Section 6, the paper is summarized, and some conclusions are drawn.

2. Background
2.1 Discourse Connectives

All languages have discourse connectives, but they differ in various ways, for example, in terms of the inventory and grammatical class of connectives, as shown as early as 1998 by Stede and Umbach (1998). The literature on connectives and discourse relations is rapidly increasing, but due to space constraints, only one source will be referenced in this discussion. Therefore, for further exploration of this topic, readers are encouraged to refer to Zufferey & Degand (2024), who provide an up-to-date account of the theories and various applications in different languages.

---

[1] The discourse structure in Fig. 1 also allows multiple relations that are totally or partly realized by pragmatic connectives (*well anyway, well if*) but not analysed in the current paper.

Despite the vast literature on discourse relations and discourse connectives, work on multiple relations and multiple connectives is scarce. In one of the earliest works, Cuenca and Marin (2009) showed the presence and co-occurrence of discourse markers in spoken Catalan and Spanish. Fraser (2013) delineated the combinations of specific and general contrastive discourse markers in English, such as *however in contrast*, *but yet*, *but still*.[2] Zeyrek (2014) dealt with the patterns of co-occurring contrastive/concessive discourse connectives in Turkish. Most recently, Cuenca and Crible (2019) described different degrees of integration of adjacent discourse markers in English.

In a different line of research, Webber (2016) and Rohde et al. (2015, 2018) suggested that multiple connectives are not required to infer multiple senses because speakers can infer multiple relations even without multiple connectives. They argued that factors like the lexical semantics of the adverbials and the properties of the passages that contain them influence the particular relations available in contexts with multiple connectives.

Given the idea that speakers do not need overt connectives to infer discourse senses, how multiple relations of the source text are inferred is a legitimate question. Looking into this issue in an annotated, multilingual corpus will enable us to reach a more complete picture of shallow discourse structure, revealing what the annotators infer from multiple relations in the presence or absence of explicit connectives in the source text and translations.

2.1. The Challenge of Automatically Extracting Multiple Connectives

Exploitation of data without annotations has several challenges if the goal is to discover multiple connectives. Notably, the methods based on collocation are unlikely to produce the desired result. Collocation is a standard method to discover multiword expressions (MWE) (Constant et al., 2017). However, it must deal with many issues, such as ambiguity, which is difficult for all NLP tasks. A quick experiment that retrieves collocating connectives from the raw texts of TED-MDB showed that the ambiguity that impacts the identification of multiple connectives in which we are interested can be derived from two sources: (a) usage ambiguity -- the system cannot decide whether an item participating in a sequence of connectives is serving as a connective or not, (b) multiword ambiguity -- a multiword sequence can be analysed as a single connective as in *but then* or a parallel connective (Webber et al., 2019).[3]

Usage ambiguity may derive from the lexical ambiguity of words, and it also leads to multiword ambiguity. For instance, the Turkish word *ancak* is ambiguous between 'only' and 'however.' It functions as a connective when it means 'however' in its context. Thus, the cluster *ancak sonra* 'only then' is a false positive token since *ancak* is not functioning as a connective but as the modifier of the adverbial 'then.'

Example (3) illustrates multiword ambiguity. Here, the Portuguese cluster *não so .. mas* 'not only .. but also,' a parallel connective, is a false positive.

(3) Pt Eles começaram a ver a sustentabilidade não só como uma coisa importante, mas como crucial para o êxito de_ o negócio .

'They began to see sustainability not only as an important thing but as crucial to the success of the business.'

Other false positives are derived from cases where two connectives reflect full embedding. An example from Lithuanian *ir jei* 'and if' is provided in (4).

(4) Lt Tai vyksta dėl mūsų pasirinkimo, o galėtų ir nevykti, ir jei pagalvoji, ką darome su šiais duomenimis, tai lyg paimtume teleskopą ir nukreipę į miestą žiūrėtume lyg į mokyklos valgyklą (..)'

'It's because of our choice, and it couldn't happen, and if you think about what we're doing with this data, it's like taking a telescope and pointing at the city, we're looking like a school canteen (..)'

For these reasons, manual annotation of multiple connectives and other types of multiple discourse relations is valuable, as it will pave the road toward better automatic systems.

3. The Current Work

The current work primarily focuses on analyzing the occurrence and variations of multiple discourse relations annotated in the English, Lithuanian, Portuguese, and Turkish segments of TED-MDB. Despite the limited scope of our data, our aim is to shed light on how multiple relations are translated and potentially open a new line of research. We hope that our observations will raise awareness, particularly by highlighting the implicitated and consistently inferred components in translation.

3.1 TED-MDB

TED-MDB is a multilingual discourse corpus annotated by following the rules and principles of the Penn Discourse Treebank 3.0 (PDTB) (Webber et al., 2019). It is a resource of six TED talks in English and translations into multiple languages, with annotations revealing the shallow discourse structure of texts. With a connective-based approach, it annotates how the underlying discourse relations are realized in texts. Thus, it annotates explicit relations, those conveyed by an overt connective, and nonexplicit ones, where a connective is absent. The labels used for nonexplicit relations involve Implicit, Alternative Lexicalization (AltLex), Entity Relation (EntRel), and

---

[2] Fraser's specific discourse markers include *on the other hand, instead, rather* and general ones involve *but, yet, still.*

[3] Our thanks to Mustafa Erolcan Er for running a preliminary experiment to automatically retrieve collocating connectives to spot candidate multiple connectives in raw texts.

No Relation (NoRel).[4] A sense label is assigned to each relation except EntRels and NoRels. The senses are assigned by selecting the most appropriate semantic category from the PDTB 3.0 sense hierarchy based on four first-level senses: Expansion, Temporal, Contingency, and Comparison. Each of these categories is specified further at a second level. The first-level senses have second-level and, in some cases, third-level sense categories encoding directionality. Briefly, Expansion refers to the elaboration relations between two text spans. The category Temporal subsumes time-related eventualities. Contingency relations encompass Cause and Condition relations and their further specifications. Comparison refers to the relations between two eventualities where differences are highlighted.

For implicit relations, the annotators insert a connective that best captures the discourse sense inferred. These are called "implicit connectives" and distinguished from relations cued by overt connectives, known as "explicit connectives" in the PDTB framework.

English, Portuguese, Lithuanian, and Turkish parts of TED-MDB have gone through several updates since the first release of the corpus. Multiple relations are systematically annotated in the most recent version of the PDTB (version 3.0), and the four language sets of TED-MDB have been updated by these new principles. Appendix 1 lists the distribution of discourse relation realization types across languages in TED-MDB.

The main extensions over TED-MDB are described in Özer et al. (2022). In the extended version, intra-sentential implicit relations and multiple relations are annotated in four languages by searching them in the circumstances determined by the PDTB 3.0 (these are listed in Table 1). Although a few multiple relations have already been annotated in the first version, more instances are spotted while annotating intra-sentential relations in the extended version. The extended version also involves the automatic alignment of the discourse relations of three target languages with the English part. A relation-linking approach is developed, where each argument of a discourse relation, its realization type, and its senses are matched. Relation linking through word alignment achieved a reasonable degree of precision, meaning the links it finds are highly likely to be an actual match.

In the current work, we rechecked the aligned dataset by examining it manually for alignment errors or inconsistencies.

3.2 Annotation of Multiple Relations

In the revised version of TED-MDB, multiple relations are searched in three circumstances introduced in the PDTB 3.0.[5]

Multiple relations are searched in instances where:

1. a relation that holds between two discourse segments is conveyed by a multiple explicit connective,

2. the explicit connective *and* conveys one relation between a pair of spans, annotators infer (and insert) a separate sense, as well,

3. there is an implicit relation between two spans, and annotators also infer (and insert) a separate sense.

Table 1: Circumstances where multiple relations are searched and annotated

3.2.2.1 Circumstance1 (Multiple Explicit Connectives): Each component of multiple explicit connectives is annotated separately with the argument spans they link and their respective senses.

3.2.2.2 Circumstance2 and Circumstance3: Multiple relations in these circumstances are annotated only in connection with the explicit or implicit conjunction *and* anchoring intra-sentential implicit relations.

To correctly identify multiple relations in circumstance2 and circumstance3, annotators are guided by specific questions. For example, multiple relations that fit circumstance2 are spotted by asking the questions shown just below excerpt (5):

(5) Now these initiatives create a more mobile workplace, <u>and</u> they reduce our real estate footprint ... (TED-MDB, Talk no. 1927)

Do you infer an implicit sense conveyed by 'and', such as causality or temporality? If so, annotate it separately by inserting an appropriate implicit connective such as 'so' or 'then'.

The questions below example (6) capture multiple relations of circumstance3.

(6) The petals unfurl, they open up, the telescope turns around. (TED-MDB, Talk no. 1976)

What is the implicit discourse relation that holds between adjacent clauses? Annotate each relation separately by inserting an implicit connective. If you infer another sense, annotate it as well with an appropriate implicit connective you will insert.

---

4. Analysis of the Corpus

4.2 Multiple Explicit Connectives

The analysis of the revised dataset shows that the least frequently occurring type of multiple relations is multiple explicit connectives (the first category in Table 1). There are 7 such tokens in English all involving the use of *and* and a separate discourse adverbial, with a corresponding number of 7 tokens in Portuguese, 6 in Lithuanian and the Turkish set. Examples from English and the matching connectives in translations are listed in Appendix 2. The table shows that the way the relation is conveyed differs as the translators may sometimes omit one of the components of multiple explicit connectives though they often translate both parts verbatim (also see the examples in Section 5).

4.3 Multiple Relations with Inferred Senses

Multiple relations annotated with one or more inferred senses, those spotted in circumstance2 and circumstance3 of Table 1, occur more frequently than multiple explicit connectives. In the English section of the corpus, 56 multiple relations with inferred senses are found. Appendix 3 presents these relations categorized by the connectives, where the implicit or explicit connective *and* and the inferred senses are counted separately.

To understand the change and variation in the translation of multiple relations with inferred senses, we checked how many relations of the source language are aligned with target texts. The analysis showed that Portuguese texts have the highest number of corresponding relations (35), while Turkish and Lithuanian translations have lower numbers (32 and 31, respectively).

Secondly, we checked the discourse relation realization types of the matching relations. The results are presented in confusion matrices in Tables 2 – 4. The tables show that translators vary regarding how they render English multiple relations with inferred senses; for example, they employ the well-known translation strategies of explicitation or implicitation.[6] The tables show that although all target languages resort to implicitation and explicitation, Portuguese ranks the highest in the implicitated and explicitated cases. For instance, according to Table 2, Portuguese translators implicitated 6 instances of the *and*-component out of the 16 cases aligned with English.

Tables 3 and 4 indicate that Turkish and Lithuanian translations tend to retain the *and*-component more often compared to Portuguese, suggesting a lower frequency of its implicitation.

En	Pt			
	AltLex	Explicit	Implicit	Total
Explicit	0	10	6	16
Implicit	1	9	9	19
Total	1	19	15	35

Table 2: How explicit-*ands* and accompanying implicit components, as annotated in English, are realized in Portuguese

En	Tr				
	AltLex	EntRel	Explicit	Implicit	NoRel	Total
Explicit	0	0	11	2	0	13
Implicit	2	1	5	10	1	19
Total	2	1	16	12	1	32

Table 3: How explicit-*ands* and accompanying implicit components, as annotated in English, are realized in Turkish

En	Lt		
	Explicit	Implicit	Total
Explicit	10	2	12
Implicit	2	17	19
Total	12	19	31

Table 4: How explicit-*ands* and accompanying implicit components, as annotated in English, are realized in Lithuanian

Thirdly, we investigated how translators addressed the inferred components that accompany *and*-relations annotated in English. Within our dataset, these inferred components encompass various senses, including Cause (expressed by 'as a result', 'so', 'consequently'), Purpose ('in order'), Temporality ('then'), and Level of detail ('in other words') (refer to Tables 5 – 7).

En	Tr					
	AltLex	EntRel	Explicit	Implicit	NoRel	Total
as a result	1	1	1	4	1	8
consequently	0	0	1	0	0	1
in order	1	0	0	0	0	1
in other words	0	0	1	0	0	1
so	0	0	0	1	0	1
then	0	0	1	3	0	4
therefore	0	0	1	0	0	1
Total	2	1	5	8	1	17

Table 5: How inferred components, as annotated in English, are realized in Turkish

The tables indicate that target languages can make the implicit component explicit. For example, Turkish renders 5 out of 17 inferred senses explicitly, and Portuguese renders 8 out of 16 (see Tables 5 and 6). Lithuanian translations demonstrate a higher degree of faithfulness to the original texts, explicitating few of

---

[6] In this work, implicitation refers to omitting a discourse connective present in the source text. In Tables 2-4, implicitation is assessed by the number of times an explicit relation of the source text is translated as an implicit relation. Explicitation is the reverse process, where a connective is used in translation, although it is absent in the source text.

the senses inferred in the English annotation process (refer to Table 7).

En	Pt			
	AltLex	Explicit	Implicit	Total
as a result	1	2	2	5
consequently	0	3	0	3
in order	0	1	0	1
in other words	0	0	0	0
so	0	0	1	1
then	0	2	3	5
therefore	0	0	1	1
Total	1	8	7	16

Table 6: How inferred components, as annotated in English, are realized in Portuguese

En	Lt		
	Explicit	Implicit	Total
as a result	2	3	5
consequently	0	1	1
in order	0	0	0
in other words	0	0	0
so	0	0	0
then	0	5	5
therefore	0	0	0
Total	2	9	11

Table 7: How inferred components, as annotated in English, are realized in Lithuanian

Tables 5 – 7 demonstrate that the implicit sense associated with *and*-relations in English annotations remains discernible in translations and is annotated accordingly during the annotation stage. This pattern is consistently observed across all translations, indicating that the senses inferred during the English annotation stage are also identified during their respective annotation stages. Further research on more significant amounts of data is needed but these initial observations imply that the more salient sense is either translated overtly or if not, it remains discernable and labeled with an appropriate implicit connective during the annotation stage.

5. Discussion

This section zooms into specific examples from the corpus to assess the change and variation in the translation of multiple relations.

5.1 Translating Multiple Explicit Connectives

Since the annotations mainly capture the connective *and* in multiple relations, this section focuses on its usage in the source text and possible implication in translated texts.

It is known that the connective *and* is highly prone to implicitation (Zufferey, 2016), and researchers have suggested that this is due to its being a weak conjunction (Asr and Demberg, 2012) or an underspecified discourse marker (Crible et al., 2019).

Whether the *and*-component of multiple connectives is kept in the translation or undergoes implicitation is

interesting, as it could contribute to the current understanding of the implication of *and*. However, it may also be the case that each connective of a multiple relation is kept or omitted independently of the other, and it is worth looking into the data with this perspective.

In the rest of this section, the examples are presented in each of the four languages if a target text is aligned with the source text. The annotated explicit connective is underlined, and the discourse realization type and the sense(s) are shown in parentheses.

Example (7) concerns *and so:* with *and,* the speaker signals a continuation; with *so*, a consequence. The consequence (or Cause:Result) sense is added to the discourse after signaling the continuation. So, readers interpret the text as follows: The fact that many of the author's early memories involved intricate daydreams is a result of the deep restlessness, a primal fear that they would fall prey to a life of routine and boredom.

(7) En There was a deep restlessness in me, a primal fear that I would fall prey to a life of routine and boredom. <u>And so</u> many of my early memories involved intricate daydreams where I would walk across borders, forage for berries, and meet all kinds of strange people living unconventional lives on the road. (Explicit; Expansion:Conjunction; Contingency:Cause)

Tr İçimde derin bir rahatsızlık var hayatın tek düzeliğine ve sıkkınlığa kurban düşeceğime dair ilkel bir korku. <u>Ve bu yüzden</u> çocukluk dönemi hatıralarımın çoğu sınırlarda yürüyüp, çilek peşinde koştuğum ve farklı farklı insanlarla karşılaştığım yollarda sıradışı bir hayat sürdüğüm karma karışık hayallerdi. (Explicit; Expansion:Conjunction; Contingency:Cause)

Lt Nenustygau vietoje, bijojau, kad tapsiu rutinos ir nuobodulio grobiu. <u>Todėl</u> daugumoje mano vaikystės prisiminimų įvairūs užsisvajojimai, kuriuose aš kertu sienas, ieškau uogų, sutinku visokiausius keistuolius - nesuvaržytus, gyvenančius kelyje. (Explicit; Contingency:Cause)

Pt Sentia uma profunda inquietação, um medo primordial de que seria vítima de uma vida de rotina e aborrecimento. <u>Por isso</u> muitas das minhas primeiras memórias envolviam sonhar acordada e de forma elaborada onde passaria fronteiras, a recolher bagas e a conhecer todo o tipo de pessoas estranhas, com vidas fora do convencional, pela estrada fora. (Explicit; Contingency:Cause)

Analysis of the translations of (7) reveals diverse strategies employed by translators. Some translators opt to directly translate the multiple explicit connective, while others choose to convey only the salient sense, such as Cause:Result, using an explicit connective. For instance, the Turkish translation maintains both the Expansion and Cause senses through equivalent multiple explicit connectives. However, the approaches differ in Lithuanian and Portuguese translations. In these languages, *and* is

implicitated, and the Cause sense is conveyed using a single explicit connective. Consequently, during the annotation stage, the Expansion sense is not inferred, yet annotators from both languages consistently infer the Cause sense. Despite these variations in translation, there is convergence among translators across different languages, as the more salient sense is always inferred. Example (8) is an instance of *and then*. Again, the connective signals a continuation of the discourse; then, the temporal relation is added. The interpretation is that we can see those planets after the star shade flies 50,000 kilometers from the telescope and is held right in its shadow. In this instance, the Portuguese and Lithuanian translations perfectly match the source text, capturing both discourse relations of the original text with equivalent multiple explicit connectives. However, in the Turkish translation, the relation is conveyed by a different connective type: a modified AltLex *ancak bu şekilde* 'only in this way', conveying a Manner sense. The annotator infers a Manner and an implicit Conjunction relation during the annotation stage. Some more cases like this exist in the corpus and reveal that the salient relation of the source text may be interpreted differently by the annotators of different languages due to a mismatching connective used in the translation.

(8) En (..) it [the star shade] has to fly 50,000 kilometers away from the telescope that has to be held right in its shadow, and then we can see those planets. (Multiple Explicit, Expansion:Conjunction; Temporal:Asynchronous)

Pt Esta sombra estelar tem cerca de metade do tamanho de um campo de futebol e tem que se distanciar 50 000 quilómetros do telescópio que tem que ser mantido na sua sombra, e então poderemos ver os planetas (Multiple Explicit, Expansion:Conjunction; Temporal:Asynchronous)

Lt (..) jis turi atsidurti tikslioje vietoje ir tada pamatysime tas planetas. (Multiple Explicit, Expansion:Conjunction; Temporal:Asynchronous)

Tr Bu yıldız gölgeleyici yaklaşık yarım futbol sahası büyüklüğünde ve gölgesi içinde tutulması gereken teleskoptan 50.000 kilometre uzakta uçması gerekiyor. (implicit = ve 'and') Ancak bu şekilde gezegenleri görebiliriz. (AltLex, Expansion:Manner)

In summary, in this section, our corpus analysis provides insights into the translation of multiple explicit connectives. We observe that the more salient meanings, such as Cause or Temporality, are generally preserved in the target text through explicit connectives, while the less salient sense, such as Expansion:Conjunction conveyed by *and* tends to be implicitated. That is, the *and*-component of a multiple explicit relation may be implicit in translation, but the more salient sense remains discernible via overt connectives. These observations suggest that multiple relations exhibit varying degrees of saliency. However, our analysis also identifies translation mismatches, which are inherent to the translation process. TED translators, often non-professionals,

may encounter challenges in conveying the multiple senses of the source text, possibly due to linguistic and contextual issues, leading to occasional discrepancies in the outcome.

5.2 Translating Multiple Relations with Inferred Senses

Having investigated how the original multiple explicit connectives (those spotted in circumstance1 of Table 1) are translated, this section focuses on how English multiple relations of the second and third circumstances are rendered in translation.

Example (9) involves an implicit intra-sentential relation; the English annotator infers multiple senses regardless (Conjunction, Temporal). The sentence is translated into Portuguese and Lithuanian verbatim. The annotators of these languages infer different senses: The Lithuanian annotator infers an Expansion and a Temporal relation that holds the clauses together. In the annotation stage, the connective *ir* is inserted to anchor the Expansion relation, the connective *tada* to anchor the Result sense, as required by the annotation guidelines. In Portuguese, however, only the Temporal sense is inferred - the relation is labeled with a single implicit connective (*depois* 'later'). Turkish translation differs from the others because the relation is translated with an overt cue (*ve* 'and'), and in the annotation stage, it is labeled with the sense of Asynchronous. Like the example (7) discussed above, this example indicates the salience of the Asynchronous sense because, in all the target texts, the relation is assigned the Asynchronous sense at the annotation stage, among other inferred senses, if any. The revised dataset has many examples where the target relation is not captured as a multiple relation. Nevertheless, annotators consistently identify the more prominent meaning annotated in the English multiple relation in our data.

(9) En (..) they open up (Imp1 = and, Imp2 = then) the telescope turns round (Implicit: Expansion:Conjunction; Temporal:Asynchronous)

Lt Žiedlapiai skleidžiasi, atsiveria (Imp1 = ir 'and' Imp2 = tada 'then') teleskopas apsisuka (Implicit; Expansion:Conjunction; Temporal:Asynchronous)

Pt (..) abrem -se (Imp = depois 'later') o telescópio vira -se (Implicit, Temporal:Asynchronous)

Tr Yapraklar açılıp genişliyor ve teleskop yön değiştiriyor (Explicit; Temporal:Asynchronous)

Finally, the intra-sentential relation in (10) is expressed through an explicit *and*-relation in English, and the annotator infers an Expansion:Conjunction sense and a separate Cause:Result sense as well. In translating this text, the explicit connective is kept in Portuguese and Lithuanian. In the annotation stage, the translations are labeled the same way as English. In the Turkish translation, the explicit connective is omitted. In the absence of *and*, only the implicit Result sense is inferred at the annotation stage. Once more, we interpret the varied annotations of this example not

as divergence but as convergence. This is because annotators from different languages consistently infer the more prominent Causal sense, even when it is not explicitly expressed in the target text.

(10) En It's a terrible shadow, <u>and</u> (Imp = as a result) we can't see planets. (Explicit; Expansion:Conjunction; Contingency:Cause)

Pt (..) uma sombra terrível <u>E</u> (Imp = por conseguinte) não conseguimos ver planetas (Explicit; Expansion:Conjunction; Contingency:Cause)

Lt Šešėlis didžiulis. <u>Ir</u> (Imp = todel) planetų mes nematome (Explicit; Expansion:Conjuntion; Contingency:Cause)

Tr Bu kötü bir gölge. (Imp = böylece 'thus') Gezegenleri göremeyiz. (Implicit; Contingency:Cause).

6. Summary and Conclusion

In summary, the analysis of multiple relations in the source text and their translation to multiple languages revealed the following:

- The analysis of multiple explicit relations that involve *and* revealed that the source text is often translated by keeping both components and if not, the more salient sense is inferrable by the annotator. These findings underscore the presence of varying degrees of saliency in multiple relations. Generally, it is the more salient relation that is explicitly conveyed, while the less salient one, that is, the expansion sense of *and*, can be implicitated.
- To further understand the issues surrounding multiple relations, we investigated *and*-relations where additional senses are inferred. In these cases, whether the explicit *and*-component is kept in translation or undergoes implicitation, our observation holds: In both of these instances, annotators of different languages often converge in inferring the salient sense in the target text corresponding to that annotated in English.

- Finally, inappropriate translations or human error in the annotation stage cannot be totally overridden. These create noise in the data and should be analyzed with caution.

The present study shows the use of a parallel, aligned dataset in investigating a specific discourse phenomenon. It sheds light on how multiple relations are treated in translation and invites translation researchers to consider these constructions in different languages. For NLP researchers, it emphasizes the challenges of automatically extracting multiple connectives, and given manual annotation costs, it highlights the need to develop discourse parsers that handle them as well as other connective types. Finally, the research also has implications for pedagogy since it increases the awareness of translators and translation teachers.

Despite these conclusions, the work is not without its limitations. The annotated multiple relations are limited to those where the connectives link the same spans and only the multiple relations associated with the conjunction *and*. The results are limited by the overall corpus size and four language sets. Multiple relations need to be readdressed in future investigations by drawing upon different genres and more amounts of data in different target languages.

In future research, we aim to experiment with the automatic extraction of multiple explicit connectives in both English and translated languages using parallel corpora. This endeavor will enhance our understanding of shallow discourse structure from the view of multiple relations cross-linguistically and contribute to the development of more robust computational tools for discourse and translation.

7. Appendices

Appendix 1: The distribution of discourse realization types across languages in TED-MDB (Özer et al., 2022)

Language Explicit	Explicit	Implicit	Alex	EntRel	NoRel	Total
English	289 (40%)	254 (36%)	46 (6%)	78 (11%)	49 (7%)	716
German	240 (43%)	214 (38%)	17 (3%)	59 (11%)	30 (5%)	560
Lithuanian	377 (46%)	315 (38%)	18 (2%)	79 (10%)	32 (4%)	821
Polish	218 (37,5%)	195 (33,5%)	11 (2%)	104 (18%)	52 (9%)	580
Portuguese	269 (40%)	311 (46%)	29 (4%)	38 (6%)	33 (5%)	680
Russian	237 (42%)	221 (39%)	20 (4%)	57 (10%)	30 (5%)	565
Turkish	315 (41%)	264 (35%)	60 (8%)	70 (9%)	51 (7%)	760
Total	1945	1774	201	485	277	4682

Appendix 2: Multiple explicit connectives in English and their correspondences in target texts. Connectives that are implicitated are enclosed within braces.

En DRID	En	Pt	Tr	Lt
DR169	and at the same time	e ao mesmo tempo	ve aynı zamanda	ir tuo pačiu
		'and at the same time'	'and at the same time'	'and at the same time'
DR25	and so	então	böylece	taigi
		'then / so'	'thus'	'thus'
DR30	and then	depois	sonra da	ir tik tada
		'after'	'then'	'and only then'
DR112	and then	e então	ancak bu şekilde	ir tada
		'and then'	'only in this way'	'and then'
DR80	and so	e por isso	ve (sonuçta)	dėl to
		'and due to this'	'and (consequently)'	'therefore'
DR120	and so	(consequentemente)	Non-aligned	Non-aligned
		'consequently'	--	--
DR42	and so	por isso	ve bu yüzden	todėl
		'due to this'	'and due to this'	'therefore'

Appendix 3: English Multiple relations in Circumstance2 and Circumstance3 with their explicit and implicit components

	Discourse Connective									
Type	And	and	as a result	consequently	in order	in other words	so	then	therefore	Total
Explicit	2	18	0	0	0	0	0	0	0	20
Implicit	0	8	10	4	1	1	1	10	1	36
Total	2	26	10	4	1	1	1	10	1	56

8. Bibliographical References

Asher, N. (2012). *Reference to abstract objects in discourse.* Vol. 50. Springer Science & Business Media.

Asr, F. T., & Demberg, V. (2012). Implicitness of discourse relations. In Proceedings of COLING 2012 (pp. 2669-2684).

Crible, L., Abuczki, Á., Burkšaitienė, N., Furkó, P., Nedoluzhko, A., Rackevičienė, S., Valunaite Oleskeviciene, G., & Zikánová, Š. (2019). Functions and translations of discourse markers in TED Talks: A parallel corpus study of underspecification in five languages. *Journal of Pragmatics, 142,* 139-155.

Constant, M., Eryiğit, G., Monti, J., Van Der Plas, L., Ramisch, C., Rosner, M., & Todirascu, A. (2017). Multiword expression processing: A survey. *Computational Linguistics, 43*(4), 837-892.

Cuenca, M. J., & Crible, L. (2019). Co-occurrence of discourse markers in English: From juxtaposition to composition. *Journal of Pragmatics, 140,* 171-184.

Cuenca, M. J., Marin, M. J (2009). Co-occurrence of discourse markers in Catalan and Spanish oral narrative. *Journal of Pragmatics*, 41(5):899–914.

Zufferey, S., & Degand, L. (2024). *Connectives and Discourse Relations.* Cambridge University Press.

Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics* 31, no. 7 (1999): 931–952.

Fraser, B. (2013). Combinations of contrastive discourse markers in English. *International Review of Pragmatics*, 5(2), 318–340.

House, J. (2015). Global English, discourse, and translation. Linking constructions in English and German popular science texts. *Target.* International Journal of Translation Studies 27(3): 370–386.

Knott, A., Sanders, T. (1998). The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of Pragmatics* 30.2 (1998): 135–175.

Lee, A., Prasad, R., Joshi, A., Dinesh, N., & Webber, B. (2006, December). Complexity of dependencies in discourse: are dependencies in discourse more complex than in syntax?. In *5th International Workshop on Treebanks and Linguistic Theories.*

Mann, W. C., Thompson, S. A. Rhetorical structure theory: Toward a functional theory of text organization. *Text-interdisciplinary Journal for the Study of Discourse* 8, no. 3 (1988): 243–281.

Marcu, D. (2000). *The theory and practice of discourse parsing and summarization*. MIT Press.

Özer, Sibel, Murathan Kurfalı, Deniz Zeyrek, Amália Mendes, and Giedrė V. Oleškevičienė. 2022. Linking discourse-level information and induction of bilingual discourse connective lexicons. Semantic Web, Vol. Pre-press, pp. 1-22.

Pasch, R., Brauße, U., Breindl, E., & Waßner, U. H. (2003). *Handbuch der deutschen Konnektoren: linguistische Grundlagen der Beschreibung und syntaktische Merkmale der deutschen Satzverknüpfer* (Konjunktionen, Satzadverbien und Partikeln) (Vol. 2). Walter de Gruyter.

Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A.K., Robaldo, L., & Webber, B.L. (2007). The Penn Discourse Treebank 2.0 Annotation Manual.

Rohde, H., Dickinson, A., Clark, C., Louis, A., Webber, B., (2015). Recovering discourse relations: Varying influence of discourse adverbials. In Proceedings of the First Workshop on Linking Computational Models of Lexical, Sentential and Discourse-level Semantics (pp. 22–31).

Rohde, H., Johnson, A., Schneider, N., & Webber, B. (2018). Discourse coherence: Concurrent explicit and implicit relations. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 2257-2267).

Stede M, Umbach C. (1998). DiMLex: A lexicon of discourse markers for text generation and understanding. In Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 2 (pp. 1238-1242). Association for Computational Linguistics.

Webber, B., Prasad, R., Lee, A., & Joshi, A. (2016). A discourse-annotated corpus of conjoined VPs. In Proceedings of the 10th Linguistics Annotation Workshop (pp. 22–31). Berlin: Association for *Computational* Linguistics.

Webber, B., Knott, A., & Joshi, A. (2001). Multiple discourse connectives in a lexicalized grammar for discourse. *Computing Meaning: Volume 2*, pp. 229–245.

Webber, B., Prasad, R., & Lee, A. (2019, May). Ambiguity in explicit discourse connectives. In *Proceedings of the 13th International Conference on Computational semantics-Long papers* (pp. 134–141).

Zeyrek, D. (2014). *On the distribution of contrastive-concessive discourse connectives ama (but/yet) and fakat (but) in written Turkish*. In P. Suihkonen and L.J. Whaley, editors, On Diversity and Complexity of Languages Spoken in Europe and North and Central Asia.

Zeyrek, D., Mendes, A., & Kurfali, M. (2018). Multilingual extension of PDTB-style annotation: The case of TED multilingual discourse bank. In *Proceedings of the 11th Language Resources and Evaluation Conference-LREC'2018* (pp. 1913-1919). European Language Resources Association.

Zeyrek, D., Mendes, A., Grishina, Y., Kurfalı, M., Gibbon, S., & Ogrodniczuk, M. (2020). TED Multilingual Discourse Bank (TED-MDB): a parallel corpus annotated in the PDTB style. *Language Resources and Evaluation*, *54*, 587-613.

Zufferey S. (2016). Discourse connectives across languages. Factors influencing their explicit or implicit translation. *Languages in Contrast* Vol. 16(2): 264–279.