Discourse-Based Annotation of Relative Clause Constructions in Turkish Sign Language (TID): A Case Study

Okan Kubus
Universität Hamburg
Hamburg, Germany
E-mail: okankubus@gmail.com

Abstract

The functions of relative clause constructions (RCC) should be ideally analyzed at the discourse level, since the occurrence of RCCs can be explained by looking at interlocutors' use of grammatical and intonational means (cf. Fox and Thompson, 1990). To date, RCCs in sign language have been analyzed at the syntactic level with a special focus on cross-linguistic comparisons (see e.g. Pfau and Steinbach, 2005; Branchini and Donati, 2009). However, to our knowledge, there is no systematic corpus-based analysis of RCCs in sign languages so far. Since the elements of RCCs are mostly non-manual markers, it is often unclear how to capture and tag these elements together with the functions of RCCs. This question is discussed in light of corpus-based data from Turkish Sign Language. Following Biber et al. (2007), the corpus-based analysis of RCCs in TID follows the "top-down" approach. In spite of modality-specific issues, the steps in the process of annotation and identification of RCCs in TID fairly resemble this approach. The advantage of using these multiple steps is that the procedure not only captures the discourse functions of the RCCs but also identifies different strategies for creating RCCs based on linguistic forms.

Keywords: relative clause constructions, Turkish Sign Language, prosody, non-manual elements

1. Introduction

The first study on RCCs as regarding sign languages was the Liddell (1978) study on ASL. Analyses on RCCs in German Sign Language (DGS) (Pfau and Steinbach, 2005) and in Italian Sign Language (LIS) (Branchini and Donati, 2009 among others) have also been put forward. Analysis of the variation among sign languages by Perniss et al. (2007) indicates that there may be non-manual markings on RCCs in common over these three sign languages, i.e. raised eyebrows. However, the aforementioned researchers emphasize that the syntactic contributions do not necessarily have to be the same: the manual markers can vary. For example, Pfau and Steinbach (2005) show that RCCs in DGS might have unique syntactic properties as compared to RCCs in the other sign languages that have been studied so far.

Indeed, the functions of relative clause constructions (RCC) should be ideally analyzed at the discourse level, since the occurrence of RCCs can be explained by looking at interlocutors' use of grammatical and intonational means (cf. Fox and Thompson, 1990). To date, RCCs in sign language have been analyzed at the syntactic level with a special focus on cross-linguistic comparisons (see e.g. Pfau and Steinbach, 2005; Branchini and Donati, 2009). However, to our knowledge, there is no systematic corpus-based analysis investigating discourse functions of RCCs in sign languages to date.

At the same time, corpus-based sign language studies have been conducted mostly at the lexical or morpho-syntactic levels. For example, at the lexical level, Johnston (2013) investigated pointing signs using corpus data in Auslan. Bank et al. (2013) describe mouthing and mouth gestures in NGT using various tiers including mouth (Dutch word that is mouthed), mouth type (mouthing or mouth gesture), mouth lemma (dictionary version of lemma) and mouth spreading (progressive or regressive spreading occurrences). At the morpho-syntactic level, Branchini et al. (2013) have discussed WH-duplication patterns in LIS by looking at occurrences of WH-signs in the LIS corpus. This paper aims towards a different approach: How it is possible to look at the bigger picture to identify a specific linguistic unit and its interconnection throughout a text through a corpus study.

Biber et al. (2007) state that corpus linguistic studies are in fact a type of discourse analysis because they cover the investigation of the functions of the linguistic forms within a particular context. Specifically, Biber et al. (2007) state corpus linguistic studies are generally considered to be a type of discourse analysis because they describe the use linguistic forms in context (p. 2). According to Biber et al., corpus studies take one of two perspectives: (i) looking at the distribution and functions of surface linguistic features and (ii) investigating the internal organization of texts. The researchers point out that corpus studies have, surprisingly, not attempted to combine these two perspectives. This study is an attempt to combine these perspectives, notwithstanding the confronted difficulties.

Following in the steps of Biber et al, the corpus-based analysis of RCCs in TİD follows the so-called "top-down" approach. In spite of issues specific to modality, there is an urgent need to develop a similar approach to investigate RCCs in sign languages. The advantage of using such an approach is that the procedure not only captures the discourse functions of RCCs but also identifies different strategies for creating RCCs based on their linguistic forms. Non-manual elements that have no independent linguistic function should be ideally covered by the "top-down" approach. This paper provides the details of these steps. The advantage of using these multiple steps is that the

procedure not only captures the discourse functions of the RCCs but also identifies different strategies for creating RCCs based on their linguistic forms.

2. Corpus study for RCCs in TİD

The data collection for the ongoing dissertation project (Kubus in progress) was conducted in two ways: (i) data obtained via elicitation and (ii) video clips shared publicly (in internet). The aim is here to obtain naturalistic, spontaneous data collected for the purpose of observing the nature of relativization.

Data elicitation (retelling stories) was conducted with three TİD signers (one native, two near-native signers). However, the data collected provided nine potential relative clauses. The amount of relative clauses thus fell short of expectations for the systematic analysis of RCCs. Obviously, there was a clear need for more relative clause samples in order to examine a wider variety of relative clause strategies that would allow for generalizations. Therefore, in addition to data obtained via elicitation, sixteen video clips, covering a wider range of potential RCCs, were selected for the annotation. The video clips are predominantly monologues signed by eleven participants (six female and five male). The entire data collection comprises of a total of twenty-one video clips consisting of approximately 3 hours of film. The sign language corpus on Turkish Signs is annotated using iLex ("integrated Lexicon"; Hanke, 2002). An annotation sample is given in Figure 1.

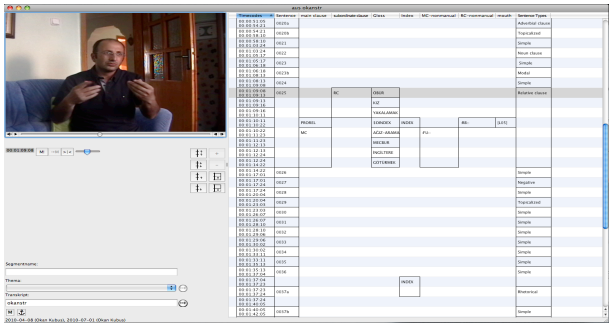

Figure 1: Data annotation (iLex)

The small-scale corpus in the ongoing dissertation project includes thirteen tiers (Table 1). Only one tier, labelled "chunks", is a structure tier and the tier "token" is a token tier. The other tokens are text tokens. Only the "chunk type" tier is subordinated under the "chunks" tier[1].

[1] Hanke and Storz (2010; p. 65) list different types of tiers. In the following, I present the three tiers, which are most often used in the ongoing dissertation project: (i) *token* tier, (ii) *structure* tier, and (iii) *text* tier.

Label	Function
Chunks	ID of each chunk
MC	The boundaries of matrix clause
RC	The boundaries of relative clause
Token	Glosses of both main clauses MC and subordinate clauses RC
Index	Marking index or other relative elements
NMM-MC	Non-manual markers for matrix clause (general)
NMM-RC1	Non-manual markers for relative clause part 1 (head movements)
NMM-RC2	Non-manual markers for relative clause part 2 (eyebrow)
NMM-RC3	Non-manual markers for relative clause part 3 (squint)
Mouth	Mouthings/ Mouth gestures specifying RC
Chunk Type	List of sentence types (e.g. declarative, interrogative, etc.)
Tr	Turkish translation equivalents of relative clauses
Eng	English translation equivalents of relative clauses

Table 1: The list of the tiers

2.1 "Top-Down" approaches in corpora study

Corpus linguistics covers various approaches with various goals for linguistic and especially discourse analyses (cf. Conrad, 2002). Conrad summarizes four corpus linguistics approaches for discourse analyses in spoken languages: (i) *Investigating characteristics associated with the use of a language feature* (p. 78), (ii) *Examining the realization of a particular function of language* (p. 81), (iii) *Characterizing a variety of languages* (p. 83) and (iv) *Mapping the occurrence of a language feature through a text* (p. 84). In the next paragraphs, each approach is described, and an argument is provided as to whether such an approach suits the current study.

According to Conrad (2002), the first approach is much more focused on a language feature, for example a word or a phrase or else a grammatical structure. In the ongoing dissertation project it is obvious that it is sought for RCC. However, due to the modality-specific properties, it turns out to be quite challenging to seek for a possible RCC in a specified corpus, since there is no previous research on this topic. Furthermore, there are no clearly spell-out words or phrases that can specify or hint such constructions. Rather, RCC seems to rely mostly on prosodic constituents of the sign language.

The second approach focuses on *a function of language and determines how it is realized in discourse* (Conrad, 2002; p. 81). For example, Biber et al. (1998) have investigated six characteristics: register, pronoun vs.

noun forms, given vs. new information status, type of reference, type of expression for anaphoric reference and the distance relationships among the characteristics. One of the findings in the ongoing dissertation project was that the type of referring expression and given/new information status relied on each other (as cited in Conrad, 2002). The present study follows this approach more by investigating RCCs and their functions in TİD. However, the challenge regarding sign language corpora which is mentioned in the previous paragraph persists. How this issue can be resolved will be explained in the next sections with the steps that are followed in the study.

In the third approach, the primary focus is *the language variety* (ibid, p. 83). For instance, Biber (1988) has developed a methodology called "multi-dimensional (MD) analysis" which includes a big scale of corpora with an automated analysis of linguistic features in more than two variables: for instance, various texts, text types, styles and/or registers (see also Biber, 1993). In this approach, multivariate statistical techniques are essential. In the ongoing dissertation project, three main discourse modes (cf. Smith, 2003) are investigated. However, since the primary focus is on RCC, it seems difficult to follow this approach with one linguistic feature variable in three different conditions. The quantity of the data and its uneven distribution over three modes makes it difficult to conduct statistical analyses. Rather, proportional (descriptive) and qualitative analysis are emphasized here.

The last approach is ... *one or more features are tracked through an entire text to determine how the features contribute to some aspect of the discourse development, such as its rhetorical organization*... (Conrad, 2002; p. 84). Indeed, this approach is closer to the approach in the ongoing dissertation project, with an exception: I am only focusing on RCC in TİD, and not on other linguistic elements. Such an approach is often related to the "top-down" approach.

2.2 The process of annotation in the "top-down" approach

The analysis and approach used in the ongoing dissertation project is inspired by the work of Biber and his colleagues. Even though there are some differences between the approach they define and the approach in the ongoing dissertation project, the core idea of the "top-down" approach is followed. It is essential to understand the structure of the RCCs in discourse analysis. In the ongoing dissertation project, not all of the signs were annotated. Rather, only the chunks that cover potential relative clauses are annotated in a detailed manner. Since this study is based on empirical research on relativization strategies, it would be too time-consuming if each segment was transcribed in a similarly detailed manner. Therefore, it is more practically efficient to follow the "top-down" approach, i.e. to specify first the possible relative clauses in TİD and then to annotate each of them.

The corpus-based approach in the ongoing dissertation project entails seven steps. First, the boundaries of discourse chunks are defined. Second, the possible sentence types included in these chunks are listed and the chunks with potential relative clauses are flagged. Then, tokens/types are constructed for each chunk, which includes possible relative clauses. Before the definition of the boundaries of each relative and matrix clause, the accompanying non-manual markers are defined. The sixth step is to translate the chunks covering the relative clauses into English and Turkish. The final step is to determine the referents in the RCC and its familiarity status within the text (i.e. if the referents have already been introduced to the text or not.).

2.2.1. Step 1: The determination of the boundaries of discourse chunks

The discourse units are narrowed down to smaller units, based on various non-manual and manual cues. Besides the prosodic cues, the meaningful smaller units are also based on semantic intuitions. It is preferred to label these smaller units as discourse chunks, because each chunk includes one or more sentences or clauses, which means that their definitions are open to discussion. The next step is to mark those chunks covering possible RCCs in order to investigate them more deeply.

2.2.2. Step 2: Selecting the chunks which include potential RCCs

RCCs in TİD are usually realized with specific non-manual markers such as raised eyebrows, tensed eyes and cheeks, some head movements and body lean. Tokens are marked with one of non-manual markers which may indicate RCCs.

Specifically, three criteria for marking RCCs in TİD are used: (i) the token includes two clauses, (ii) one clause is dependent on another clause in the selected token (iii) the token is realized with one of specific non-manual markers.

2.2.3. Step 3: Token/type constructions for the flagged discourse chunks

Only the discourse chunks which might include the potential RCCs are annotated. The entries for tokens and types are adapted from the transcription process used in Technical Sign Lexicon Projects (cf. Konrad, 2010), under the auspices of the Institute of German Sign Language and Communication of the Deaf (IDGS). According to Konrad (2010; pp. 28-29), this transcription is based on the distinction between tokens and types, i.e. each token refers to a distinctive type. In other words, *types should be uniquely or consistently identified*.

2.2.4. Step 4: Defining non-manual markers

The next step after annotating the tokens is to annotate non-manual markers for both relative clauses and matrix clauses. The cross-linguistic analyses of relative clauses in signed languages indicate that non-manual markers in relative clauses are generally accompanied by brow raise, tensed eyes/squint, and head movements if needed. Therefore, three tiers are constructed for annotating non-manual markers: (i) eyebrow movements, (ii) tensed

eyes/cheeks and (iii) head/body movements.
Common categorizations for eyebrow movements are (i) brow raise, (ii) neutral brow and (iii) furrowed brows (cf. Wilbur, 2000). Both brow raise and furrowed eyebrow raise are indicated by 'br' and 'fb' respectively and any other eyebrow movement assumes a neutral eyebrow code. Other non-manual markers are also involved, such as: tensed lips (i.e. ASL: Liddell, 1978), tensed eyes (i.e. LSC, Mosella, 2010), tensed cheeks (i.e. LIS, Branchini and Donati, 2009) and squint (i.e. Dachkovsky and Sandler, 2009). It is assumed that these four facial expressions resemble each other and I categorize them as squint which is coded as 'sq.' In addition, some head and torso movements may accompany relative clauses, even though they are not strong indicators. In order to mark these indicators, the third tier represents head and torso movements which include head tilt (back) 'ht', head nod (forward) 'hn', head shake 'hs', and body lean 'bl' (cf. Wilbur, 2000).
Non-manual expressions are not restricted to relative clauses. Different non-manual markers in matrix clauses may be observed as well. These markers may give a clue about sharp boundaries between relative clauses and matrix clauses (cf. Dachkovsky and Sandler, 2009). Also, these non-manual markers occurring in matrix clauses can be independent from the indication of relative clauses (e.g. negation, question). Therefore, another tier is constructed for the investigation of facial, head and torso movements in matrix clauses.
Furthermore, lower face movements may be significant for the realization of relative clauses. For instance, in TİD tensed lips and the mouthings 'o' and 'bu' are frequently observed. These are also coded separately.

2.2.5. Step 5: Defining boundaries of RCCs

After specifying the non-manual markers, the boundaries of relative and matrix clauses need to be specified as well. Boundaries are primarily based on non-manual markers such as brow raise and squint.

2.2.6. Step 6: Translation equivalents of potential RCCs

Turkish translation equivalents and Turkish glosses of Turkish Sign Language, as well as English glosses and English translation equivalents, are provided in a separate tier. Translation equivalents of some RCCs may not represent potential TİD RCCs exactly because of possible cross-language/cross-modal differences in syntactic constructions.

2.2.7. Step 7: Discourse analysis of RCCs

The referents that are used in RCCs are determined and interconnections between the referents are checked. This helps to understand the function of RCCs. This study focuses on the function of RCCs in various discourse modes from a linguistic point of view, in the framework of the Segmented Discourse Representation Theory (SDRT: Asher and Lascarides, 2003).
Aksu-Koç and Erguvanlı-Taylan (1998; p. 277, inspired by Fox and Thompson, 1990) specify two different references to the expressions (i) *head* and (ii) *modifying clause*. According to them, head can either be introduced into discourse for the first time, or else introduced again in the sense of the familiarity status of information. The information in a modifying clause can be realized in three different forms. If the modifying clause is made for clarifying the ambiguous content of the head, the clause has an *identification* function. If the content of the modifying clause has already been introduced earlier and is once again introduced into the discourse, it has been *re-identified*. Conversely, some modifying clauses may function as tools to express supplementary information about the head. Such clauses are regarded as *characterizing* modifying clauses. Using this categorization, in the ongoing dissertation project each head and modifying clause in flagged discourse chunks with potential RCCs in TİD is identified with underlying properties.

3. Advantages and disadvantages of the "top-down" approach

The annotation process in this dissertation project favors the "top-down" approach. This process has both advantageous and disadvantageous sides. The first advantage is that the "top-down" approach is primarily based on a specific research question and can focus on the findings and annotations that are related to this goal. The second advantage of this approach is the fact that it does not tokenize data which may not be related to the specific goal. The third advantage is that this approach allows deduction, i.e. from wider linguistic units to narrower units. For instance, this study looks at the discourse text first and divides it into possible discourse chunks and phonological utterances (cf. Sandler and Lillo-Martin, 2006). It also goes further into intonational phrases, phonological phrases and even prosodic words (i.e. here tokens). In addition, after deduction, it allows an inductive approach as well, e.g. in the ongoing dissertation project tokens may give a clue about the syntactic constructions.
However, this approach has disadvantages as well. If all discourse chunks are not treated equally, there is a danger of missing potential samples. For instance, in the ongoing dissertation project not all discourse chunks are glossed in terms of tokens/types and therefore other possible relative clauses may potentially be overlooked. In order to avoid such loss, each discourse type has been labelled with respect to its sentence types, as far as possible. This strategy may make up for the first disadvantage. The second drawback is that there is a need for a native signer with meta-linguistic awareness so that he/she may decide which chunks may include potential data related to the specific research aim.

4. Conclusion

Due to modality-specific properties, the "top down" approach can be seen as challenging to use for corpora in signed languages. No matter how large the corpus is, in order to understand the function of a linguistic element, a "top down" approach can assist in the obtaining of a bigger picture of the discourse development. As mentioned before, RCCs in TİD do not necessarily have a linguistic and manual form. Rather, RCCs in TİD mostly rely on prosodic constituents, which can vary. Starting from a text and dividing into smaller units with the help of non-manual expressions as well as semantic

intuitions made the analysis of RCCs in TİD possible.

The approach developed for this study might have some drawbacks and may benefit from further refinements; however, this approach might shine a light on the investigation of linguistic forms in signed languages which might not have a manual form, such as yes/no questions, topicalization and RCCs.

5. References

Aksu-Koç, A. and Erguvanlı-Taylan, E. (1998). The functions of relative clauses in narrative discourse. Lars Johanson (Ed.) T*he Mainz Meeting Proceedings of the Seventh International Conference on Turkish Linguistics Wiesbaden,* Harrasowitz Verlag, pp. 271--284.

Asher, N. and Lascarides. A. (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.

Bank, R. ; Crasborn, O. and van Hout, R. (2013). Alignment of two languages: The spreading of mouthings in Sign Language of the Netherlands. *International Journal of Bilingualism*. Published online before print, May 3, 2013, doi: 10.1177/1367006913484991.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. (1993). Representatives in corpus design. *Literary and Linguistic Computing*, 8, pp. 243--257.

Biber, D., Connor, U. and Upton, T.A. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure*. Amsterdam: John Benjamins.

Biber, D.; Conrad, S. and Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Branchini, C.; Cardinaletti, A.; Cecchetto, C.; Donati, C. and Geraci, C. (2013). WH-duplication in Italian Sign Language (LIS). *Sign language & linguistics*, 16(2), pp. 157--188.

Branchini, C. and Donati, C. (2009). Relatively different: Italian Sign Language relative clauses in a typological perspective." In A. Liptàk (Ed.), *Correlatives cross-linguistically* Amsterdam: Benjamins, pp. 157--191.

Conrad, S. (2002). Corpus linguistics approaches for discourse analysis. *Annual Review of Applied Linguistics*, 22, pp. 75—95.

Dachkovsky, S. and Sandler, W. (2009). Visual intonation in the prosody of a sign language. *Language and Speech,* 52(2/3), pp. 287--314.

Fox, B. and Thompson, S. (1990) A discourse explanation of the grammar of relative clauses in English conversation. *Language* 66, pp. 297--317.

Hanke, T. (2002). iLex - A tool for sign language lexicography and corpus analysis. In M. González Rodriguez, & C. Paz Suarez Araujo (Eds.): *Proceedings of the third International Conference on Language Resources and Evaluation*, Las Palmas de Gran Canaria, Spain. Paris: ELRA, pp. 923--926.

Hanke, T. and Storz, J. (2010). iLex – A database tool for integrating sign language corpus linguistics and sign language lexicography. In Crasborn, Onno et al. (Eds.): LREC 2008. *6th International Conference on Language Resources and Evaluation. Workshop Proceedings. W25. 3rd Workshop on the Representation and Processing of Sign Languages*. Sunday 1st June 2008, Marrakech – Morocco. Paris: ELRA, pp. 64--67.

Johnston, T. (2013). Formational and functional characteristics of pointing signs in a corpus of Auslan (Australian sign language): are the data sufficient to posit a grammatical class of 'pronouns' in Auslan? *Corpus Linguistics and Linguistic Theory*, 9(1), pp. 109-159.

Konrad, R. (2010). Die Erstellung von Fachgebärdenlexika am Institut für Deutsche Gebärdensprache (IDGS) der Universität Hamburg (1993-2010). Universität Hamburg. URL: http://www.sign-lang.uni-hamburg.de/projekte/mfl/Konrad_2010_FachgebLexika.pdf (last accessed on January 06 2012).

Kubus, O. (in progress) Relative Clause Constructions in Turkish Sign Language (TİD) Ph.D. dissertation, Universität Hamburg.

Liddell, S. (1978). Non-manual signals and relative clauses in ASL. In P. Siple (Ed.), *Understanding language through sign language research*. New York: Academic Press.

Mosella Sanz, M. (2011). The position of fronted and postposed relative clauses in Catalan Sign Language. *Paper presented at Venice FEAST Colloquium, Formal and Experimental Advances in Sign language Theory*, Venice, June 20-22, 2011

Perniss, P.; Pfau, R.; M. Steinbach (2007). Can't you see the difference? Sources of variation in sign language structure. In P. Perniss, R. Pfau & M. Steinbach (Eds.), *Visible variation: Comparative studies on sign language structure*. Berlin: Mouton de Gruyter, pp. 1--34.

Pfau, R. and Steinbach, M. (2005). Relative clauses in German Sign Language: Extraposition and reconstruction. In L. Bateman and C. Ussery (Eds.), *Proceedings of the North East Linguistic Society (NELS 35)*, Vol. 2. Amherst, MA: GLSA, pp. 507--521.

Sandler, W. and Lillo-Martin, D. (2006). *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press.

Smith, C. S. (2003). *Modes of Discourse*. Cambridge University Press.

Wilbur, R. B. (2000). Phonological and prosodic layering of nonmanuals in American Sign Language. In K. Emmorey & H. Lane (Eds.), *The Signs of Language Revisited: An Anthology to Honor Ursula Bellugi and Edward Klima*. Mahwah, NJ: Lawrence Erlbaum.