KWIC view on Constructed Action (CA) and its Collocates in German Sign Language (DGS) – Possibilities and Limitations

Sabrina Wähl
University Hamburg
Gorch-Fock-Wall 7, 20354 Hamburg, Germany
sabrina.waehl@uni-hamburg.de

Abstract

Constructed action (CA) is a phenomenon that is used in signed discourse to show the actions of a referent (cf. Cormier et al., 2015; for DGS cf. Fischer and Kollien, 2010). To achieve this, the signer adopts the role of the referent. Most studies use retellings as their data basis (e.g. Herrmann and Pendzich, 2018; Cormier et al., 2015). Consequently, there is less research on CA and its use in data that is not influenced by stimuli. Though there is a considerable number of studies on CA, the phenomenon is still not well understood. One possible way to understand this multifaceted phenomenon better is to analyse collocations in conversations. In spoken language lexicography concordance lines – also known as keyword in context (KWIC) – have proven to be a useful tool in the analysis of collocations. The data used in this study are *Free conversations* in the Public DGS Corpus.
This paper explores the possibilities and limitations of concordance lines as a tool to analyse collocational behaviour of CA. It also presents preliminary results regarding CA and its collocates which may be explored further in the future.

Keywords: keyword in context, constructed action, corpus

1. Introduction

There are different ways to express referents in sign languages by telling or showing. Constructed action (CA) is a way to express referents by showing. The term refers to constructions that depict the actions, behaviour, thoughts and utterances of a referent (e.g. Cormier et al., 2015; for DGS see Fischer and Kollien, 2010). This phenomenon was described from an early stage in sign language research (e.g. Friedman, 1975 or Mandel, 1977 for American Sign Language – ASL) though it was labeled very differently in various approaches. Different terms are among others: *personal transfer* (Cuxac, 2000), *body classifiers* (Suppalla, 1986) and *performatives* (Winston, 1991). However, the most frequently used terms are, *role shift* (e.g. Lillo-Martin, 2012; Herrmann and Steinbach, 2010) and *CA* (e.g. Metzger, 1995; Cormier et al., 2015). Quite often CA is regarded as gestural (cf. Liddell and Metzger, 1998; Emmorey, 1999; Ferrara and Johnston, 2014). It is expressed by non-manuals like eye gaze, facial expression or shifts of the torso, as well as the hands. Signers may also use *body partitioning* (Dudis, 2004), i.e. different articulators may represent different referents as shown in figure 1. In this example the signer articulates the one-handed DGS sign with the meaning 'to look at' with both hands. Thus, the signer maps one referent, i.e. the signer's partner, onto one hand and a second referent, i.e. the signer, onto the other hand. Both look with a frown (expressed by the signers face) at something on a website[1].

[1] Transcript, 00:12:22:12.

Figure 1: 'to look at sth. together'

There is a considerable amount of research on CA and the non-manuals marking CA, but it is also interesting to further investigate the manual activity during a CA and the preceding signs. One possible way to approach this is using concordance lines, also known as *keyword in context* (KWIC). These display signs or words that precede or follow a specific keyword. Collocational behaviour can reveal a lot about the use of a sign. This leads to the following research questions:

- What insights can be gained about CA by using a KWIC view for analysis?
- How useful is such a view of the data and where are its limitations?

2. Data

To answer these research questions, I used conversational data of German Sign Language (DGS;

Proceedings of the 12th Workshop on the Representation and Processing of Sign Languages, pages 484–490
16 May 2026.

Deutsche Gebärdensprache) collected by the DGS-Korpus project as a data basis. The DGS corpus contains approximately 560 hours of signed discourse (Konrad et al., 2024). Participants were filmed in couples and presented with different elicitation tasks such as *Free conversation*, *Discussion* or story retellings (e.g. *Frog Story*, *Sylvester and Tweety*). The data basis for the present research was the task *Free conversation* in the Public DGS Corpus (Konrad et al., 2020). The Public DGS Corpus is a publicly available selection[2] of annotated language material. In total, 34 transcripts of these *Free conversations* are available in the Public DGS Corpus, which is a total of 68 participants. For this task, the participants were asked to have a conversation about a topic of their choice. The session's moderator was not present during the *Free conversation*. Within the DGS corpus, this task is the one that is the least influenced by stimuli. As there already is a lot of research on CA in story retellings but little research on formats that are less influenced by stimuli, this task was chosen for the analysis.

The annotations in the DGS corpus include glosses, mouthing and translations (cf. Konrad et al., 2022). iLex – an integrated annotation tool and lexical database (Hanke and Storz, 2008) – was used to annotate the present data set. *Productive signs* are glossed as $PROD which is a collective gloss including various productive signs. The DGS Corpus project defines productive signs as follows:

> Unlike in vocal languages, signing discourse is not only a succession of lexical units, but an interplay between lexical signs (predominantly denoting something: *telling*) and productive signs (predominantly illustrating the intended meaning: *showing*). These productive signs – also known as classifier or polymorphemic signs – are fully iconically motivated. They have a general meaning to which each iconic value of its components contributes (Konrad et al., 2022, 12).

The glosses for gestures start with $GEST (cf. Konrad et al., 2022). They are defined as follows:

> Manual activity that is neither a conventional nor a productive sign is likely to be a gesture when this activity is also used in the hearing majority with same or similar functions (Konrad et al., 2022, 16).

Unlike other corpus projects, such as the Auslan (Australian Sign Language)[3] corpus, CA is not explicitly annotated in the gloss tier in the DGS corpus, it was annotated with glosses such as $PROD or $GEST instead. DGS corpus transcripts include, among others, separate tiers for glosses, mouthing, translations and comments. Annotators had the option to note supplementary information such as CA in a tier for comments. Some of the annotators also chose to make the distinction between CA and *constructed dialogue* (CD)[4]. As this was not obligatory, annotation of CA is therefore incomplete. Still, there is a considerable amount of annotations that can be used for preliminary analysis. In this paper, a KWIC view was used to conduct such an analysis which is described in the following sections.

3. KWIC

For the analysis I used an SQL[5] query that showed all annotations of CA combined with a KWIC (see figure 2). The query showed 486 occurrences of CA annotation from 47 signers across 27 transcripts. It showed:

1. which signs annotated in the gloss tier co-occur with the CA annotation and
2. which signs precede the CA annotations.

I chose to look at preceding and aligning signs only, as I was interested in seeing which signs are used to indicate a referent if it is expressed by a manual sign at all. Initially the query was designed so that it showed up to 5 signs preceding the CA annotations. I could not identify any relevant patterns within that range and narrowed it to 3 signs that were uttered before the CA. In the following sections, detectable patterns which appeared within the range of up to 2 preceding signs are discussed.

3.1. Top 5 signs aligned with CA comment

For the signs aligning with CA annotations, I analysed which signs occurred most often. In this context, aligned means that there are CA annotations in a comment tier and parallel to that gloss annotations in the gloss tier. The top 5 signs (i.e. the 5 signs or group of signs with the most occurrences) that are annotated aligned with CA are productive signs, gestures, referential signs, i.e. indexes and the signs for 'I' (I1 – index finger,I2 – flat

[2]For the selection criteria, see Salden and Konrad (2014).

[3]For the Auslan Annotation guidelines, see Johnston (2024).

[4]In this paper, CD is treated as subcategory of CA as it denotes CAs that construct communicative actions.

[5]SQL (Structured Query Language) is a standard query language designed for interacting with databases.

Figure 2: KWIC view: CA and preceding signs

Sign aligned to CA comment	Number of Tokens
$GEST	91
$PROD	71
I1/ I2	39
$INDEX	39
TO-LOOK	34

Table 1: Top 5 signs aligning with CA annotations

Figure 3: 'to look at somebody with discomfort'

B-handshape) and signs that denote 'to look' (see table 1).

Of the the top 5, gestures – a group of types all starting with $GEST[6] – occur the most with 91 tokens. This finding could be considered in support of the claim that CA is gestural. Productive signs make up the group that is second most aligned with CA comments. The gloss $PROD contains a large number of classifier constructions. Dudis (2004) (among others) has described that classifier constructions and CA tend to co-occur whenever body partitioning takes place. Again, the results from the query could be seen as further evidence of patterns that have been observed before.

Third most common in *Free conversations* are referential expressions, i.e. I1/ I2 and $INDEX. Here a limitation of the KWIC view shows: The frequency of use is apparent but it is not possible in the KWIC to distinguish whether the signs for 'I' actually means me – the signer – or refers to a third person. I additionally analysed the use of I1/ I2 within CA further with regard to: **Is I1/ I2 referencing the signer or a third person?** 30 tokens were referring to the signer. Hence, these are first person uses of I1/ I2 within CA with the signer being the referent. This construction can be interpreted as: "I was like ..."[7]. The remaining 9 tokens referred to third person referents whose actions are represented. The frequency of I1/ I2 is probably an effect caused by the elicitation task which allowed for narratives of the signer's own experiences. Also, the use of I1 and I2 shows that there seems to be a need to clarify that the signer is the referent.

The last group of signs that aligned the most with the CA annotations is a group of signs that express 'to look at sth.' (for examples see, figures 1 or 3[8]). The group consists of signs using the V-handshape with different movement patterns (TO-LOOK-AT1,TO-LOOK-AT2, TO-LOOK-AT3, TO-LOOK-AWAY-AND-BACK1, TO-SEE-EACH-OTHER) and signs using the 5-handshape indicating a larger number of people (TO-LOOK-AT7, see figure 4[9]).

Hou (2022) described the usage of LOOK for ASL and stated: "While there is no unique marker or a

[6]This is a cluster of types. Accordingly, there is no DOI link whenever I refer to this cluster. For an overview of all types included in the Public DGS Corpus, see the types list.

[7]Lillo-Martin (2012) describes similar constructions and compares them to the use of "be like" in spoken English.

[8]Transcript, 00:02:35:08.

[9]Transcript, 00:00:11:49.

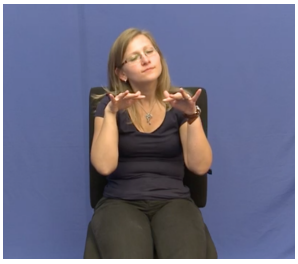

Figure 4: 'audience watches attentively a presentation'

unique group of markers that signals constructed action, LOOK seems to be a common trigger for signaling it in ASL and even other signed languages." The concordance shows that TO-LOOK indeed occurs with a certain frequency together with CA in the DGS data. At this point, it would be interesting to analyse more tokens that denote 'to look at sth.' in order to confirm that these are signs primarily used in CA. The query used for this paper would not be useful for this kind of analysis as annotation of CA is incomplete. Therefore, it would be necessary to annotate the given sign types for CA first. However, the result of the query points toward a similar use of TO-LOOK in DGS and ASL.

3.2. Top 5 signs immediately preceding CA comment

Sign immediately preceding the CA comment	Number of Tokens
I1/ I2	75
$GEST	41
$INDEX	41
$PROD	37
signs denoting persons	29

Table 2: Top 5 signs immediately preceding CA annotations

The top 5 signs that precede CA again include indexes and I1 and I2, but the signs for 'I' occur the most (see table 2). Accordingly, DGS signers tend to use I1/ I2 as a preceding sign in *Free conversations*.
At this point the question arises: **Do the signs for 'I' introduce the following CA?** Again, the limitations of the KWIC view show here. It presents a string of glosses, but one cannot determine whether the preceding sign relates to the following CA or is part of another clause-like unit. I addressed this issue by analysing those 75 co-occurrences. Based on this analysis, I have determined that most occurrences – i.e. 67 tokens – serve as an introduction indicating that the signer is depicted in the subsequent CA. Another possibility to account for that shortcoming of the KWIC view is to have annotations for clause-like units that could be displayed together with the concordance lines indicating boundaries. The DGS corpus only provides translations that could be used as an approximation. This approach and its limitations have been discussed in the past. There is also a risk that the translation influences the analysis, i.e. its structures will be analyzed rather than the signed sequence itself which is the main reason why I decided against this approach.

$GEST and $PROD also belong to the top 5 though the number of tokens drops visibly. The number of tokens of $INDEX is relatively stable. Signs that denote a person also occur, which I grouped together into one category as there are 29 signs denoting all in all 20 different concepts such as 'deaf person', 'father', 'teacher' or 'boss'. The sign that was used the most was DEAF1A. This sign can be used with different senses, two of them being:

1. term for a person who is deaf
2. to be deaf (in a cultural or medical sense)[10].

The intended sense in the given context cannot be detected from the gloss alone. Thus *word-sense disambiguation* is necessary to determine the actual meaning of the token. Accordingly, this is also a limitation of the KWIC. It gives you an overview over types but the more fine grained differences in use are not necessarily reflected.
Taking the 67 tokens of I1/ I2 introducing CA and the signs denoting a person together, there are 96 tokens that introduce the referent of a following CA. That is 19,75% of the tokens preceding the CA annotations (based on the 486 occurrences – see section 3). This shows that DGS signers use manual referent introduction quite regularly in *Free conversations*. This aligns with research on referent shift strategies which have shown that "point-to-chest" (i.e. I1/ I2) and lexical signs are relevant strategies to mark a reference entity before the signer shifts (cf. Kocab et al., 2015; Stamp and Sandler, 2021).

3.3. Other patterns

So far we have looked at signs that align with the CA annotations and signs that immediately precede them. Interestingly, a pattern of two signs (on one occasion 3 signs) and a subsequent CA is also becoming apparent. This pattern is a sequence of a sign for a person followed by a sign of communication followed by a *constructed dialogue*.

[10] See Dictionary entry 8 in Digitales Wörterbuch DGS (DW-DGS) (Langer et al.), own translation.

As mentioned above, CD is the term for all communicative actions that are constructed in signed discourse. Examples for this pattern are:

- I1 TO-LET-KNOW1A 'I let someone know': Constructed dialogue
- BOSS1B SAY1 'The boss said': Constructed dialogue

This sequence of signs introduces the CD. This pattern occurs 14 times in the data set used for this paper. To further confirm this pattern, it would be necessary to expand the data set by adding another elicitation task. Alternatively, a further analysis that takes the communicative signs as a starting point could be conducted for further confirmation.

4. Discussion

To sum up the preceding sections, it can be said that KWIC provides an overview of a CA's neighbouring signs. There are several limitations of the approach presented in this paper which are:

- glosses are only an approximation to the actual meaning of the used signs,
- the concordance lines currently do not show boundaries of clause-like units.

These shortcomings can be dealt with by performing additional analyses, such as word-sense disambiguation, which would in any case be necessary for a lexicographic analysis.

So far, there is one issue that may affect the concordances described above: overall lexical frequency in a corpus. You can find a certain number of high frequency words or signs in any corpus. For the DGS corpus, the top 6 signs are[11]:

1. $INDEX (44,304 tokens)
2. I1 (37,181 tokens)
3. $GEST-OFF (29,442 tokens)[12]
4. I2 (12,788 tokens)
5. $GEST (12,027 tokens)
6. $PROD (10,467 tokens).

Hence, it is not surprising that these highly frequent signs appear as collocations in the KWIC. To address this issue one could calculate the mutual information score that measures the relation between two variables. Another possibility is to exclude the high frequency signs from the query which might help to identify patterns with less frequent signs, but also does not show the whole picture. I chose a mixed approach that combines quantitative and qualitative elements by describing the data and interpreting them. For the analysis, I revisited the concordance lines and analysed the range of uses of the signs discussed below. Thus the following ranges include signs from the third preceding sign up to aligned signs in the KWIC (for a schematic representation, see figure 5).

Figure 5: Schematic representation of the KWIC view with signs represented from 3rd preceding sign to signs aligned to CA annotations

The number of $INDEX tokens in the different positions (i.e. third preceding sign, second preceding sign, first preceding sign, aligned sign) is relatively steady and ranges between 37 and 43 occurrences. Because of this little variance of tokens, I conclude that this collocational pattern is likely due to the high frequency of $INDEX.
$GEST tokens preceding the CA annotations range between 29 and 41 tokens. The token number for $GEST and CA annotations aligned is 91 and doubles in comparison. This is similar for $PROD with preceding signs ranging between 27 and 31 tokens and 75 tokens that are aligned. Accordingly, a tendency that $GEST and $PROD are more often used in CA can be observed which is probably not only due to their high frequency in the corpus. I1 and I2 are used more often before CA than $GEST and $PROD. Token counts for I1 and I2 for preceding signs range from 58 to 75 – with 75 being the signs immediately preceding the CA annotations. The use of I1 and I2 drops (39 occurrences) when aligned with the CA annotations which indicates a tendency that signers may prefer constructions where signs denoting 'I' are preceding the CA.
The signs for the meaning 'to look' are also highly frequent (e.g. TO-LOOK-AT1 with 901 tokens throughout the DGS corpus) but are not as numerous as the signs discussed in the previous paragraph. Hence, high frequency effects might be less strong. Also it is interesting that the pattern is only relevant for the signs aligned to the CA annotations, which leads me to the conclusion that this indeed points to a preferred use of these signs within CA.

For the preceding signs in general, I found that 19.75% of the 486 occurrences are used to introduce the CA's referent. However, within that group I1/ I2 are the most frequent. Nevertheless it is interesting to observe that signers quite frequently intro-

[11] As of 18.02.2026.

[12] This gesture is often called *palm up open hand* (PUOH) in literature.

duce the following referent. This facilitates referent identification for the addressee. Non-manual markers or simply a prototypical representation could be used to mark the referent, but one could argue that prototypical representations do not work if the signer is the depicted referent as each person is unique. This might therefore be the reason for the more frequent use of I1 and I2 directly preceding a CA. However, in order to present valid results for DGS with regard to this, a consequent and complete annotation of CA would be needed throughout the DGS corpus or at least throughout a sub-corpus.

5. Conclusion

A KWIC view can be used as a means to explore collocational patterns of CA as it provides an overview and can point to patterns of interest for further analysis. Future studies could complete annotations for CA and clause-like units which may allow for a more detailed statistical analysis. Clearly, one has to keep in mind that certain signs appear with a higher frequency and take that into account. Concordance lines are a useful tool within a set of methods, but as a standalone method its limitations are reached fairly quickly.

6. Acknowledgements

I would like to thank all the participants of the DGS-Korpus project for taking part in the project and thus making such valuable signed data available for research. I also want to thank the DGS-Korpus project for the support and their work. This publication has been produced in the context of the joint research funding of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the German Academies of Sciences and Humanities.

7. Bibliographical References

Kearsy Cormier, Sandra Smith, and Zed Sevcikova-Sehyr. 2015. Rethinking Constructed Action. *Sign Language & Linguistics*, 18(2):167–204.

Christian Cuxac. 2000. Iconicity of Sign Languages. In M. Martin Taylor, Françoise Néel, and Don Bouwhuis, editors, *The Structure of Multimodal Dialogue II*, pages 321–334. John Benjamins Publishing Company.

Paul G. Dudis. 2004. Body Partitioning and Real-Space Blends. *Cognitive Linguistics*, 15(2):223–238.

Karen Emmorey. 1999. Do Signers gesture? In Lynn Messing and Ruth Campbell, editors, *Gesture, Speech, and Sign*, pages 133–158. Oxford University Press.

Lindsay Ferrara and Trevor Johnston. 2014. Elaborating Who's What: A Study of Constructed Action and Clause Structure in Auslan (Australian Sign Language). *Australian Journal of Linguistics*, 34(2):193–215.

Renate Fischer and Simon Kollien. 2010. Gibt es Constructed Action in Deutscher Gebärdensprache und in Deutsch (in der Textsorte Bedeutungserklärung)? *DAS ZEICHEN*, 24(86):502–510.

Lynn A. Friedman. 1975. Space, Time, and Person Reference in American Sign Language. *Language*, 51(4):940–961.

Thomas Hanke and Jakob Storz. 2008. iLex – a Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In *Proceedings of the LREC2008 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, pages 64–67, Marrakech, Morocco. European Language Resources Association (ELRA).

Annika Herrmann and Nina-Kristin Pendzich. 2018. Between Narrator and Protagonist in Fables of German Sign Language. In Annika Hübl and Markus Steinbach, editors, *Linguistik Aktuell/Linguistics Today*, volume 247, pages 275–308. John Benjamins Publishing Company.

Annika Herrmann and Markus Steinbach. 2010. Eine neue Perspektive auf Role Shift in Deutscher Gebärdensprache (DGS) – Perspektivwechsel als nichtmanuelles Kongruenzphänomen. *DAS ZEICHEN*, 24(84):112–118.

Lynn Hou. 2022. Looking for Multi-Word Expressions in American Sign Language. *Cognitive Linguistics*, 33:291–337.

Trevor Johnston. 2024. Auslan Corpus Annotation Guidelines.

Annemarie Kocab, Jennie Pyers, and Ann Senghas. 2015. Referential Shift in Nicaraguan Sign Language: A Transition from lexical to spatial Devices. *Frontiers in Psychology*, 5:1–13.

Reiner Konrad, Thomas Hanke, Amy Isard, Marc Schulder, Lutz König, Julian Bleicken, and Oliver

Böse. 2024. Corpus à la carte – improving Access to the Public DGS Corpus. In *Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources*, pages 390–399, Torino, Italy. ELRA Language Resources Association (ELRA) and the International Committee on Computational Linguistics (ICCL).

Reiner Konrad, Thomas Hanke, Gabriele Langer, Dolly Blanck, Julian Bleicken, Ilona Hofmann, Olga Jeziorski, Lutz König, Susanne König, Rie Nishio, Anja Regen, Uta Salden, Sven Wagner, Satu Worseck, Oliver Böse, Elena Jahn, and Marc Schulder. 2020. MEINE DGS – annotiert. Öffentliches Korpus der Deutschen Gebärdensprache, 3. Release / MY DGS – annotated. Public Corpus of German Sign Language, 3rd release.

Reiner Konrad, Thomas Hanke, Gabriele Langer, Susanne König, Lutz König, Rie Nishio, and Anja Regen. 2022. Öffentliches DGS-Korpus: Annotationskonventionen / Public DGS Corpus: Annotation Conventions. Project Note AP03-2018-01, DGS-Korpus project, IDGS, Hamburg University, Hamburg, Germany.

Gabriele Langer, Anke Müller, Sabrina Wähl, Felicitas Otte, Lea Sepke, and Thomas Hanke. Digitales Wörterbuch der Deutschen Gebärdensprache. Das korpusbasierte Wörterbuch DGS – Deutsch.

Scott K. Liddell and Melanie Metzger. 1998. Gesture in Sign Language Discourse. *Journal of Pragmatics*, 30(6):657–697.

Diane Lillo-Martin. 2012. Utterance Reports and Constructed Action. In Roland Pfau, Markus Steinbach, and Bencie Woll, editors, *Sign Language*, pages 365–387. De Gruyter Mouton.

Mark A. Mandel. 1977. Iconic Devices in American Sign Language. In Lynn A. Friedman, editor, *On the Other Hand: New Perspectives on American Sign Language*, Language, Thought, and Culture, pages 57–108. Academic Press.

Melanie Metzger. 1995. Constructed Dialogue and Constructed Action in American Sign Language. In Ceil Lucas, editor, *Sociolinguistics in Deaf Communities*, pages 255–271. Gallaudet University Press.

Uta Salden and Reiner Konrad. 2014. Auswahl von Aufnahmen für das Teilkorpus. Project Note AP06-2013-01v02, DGS-Korpus project, IDGS, Hamburg University, Hamburg, Germany.

Rose Stamp and Wendy Sandler. 2021. The Emergence of Referential Shift Devices in three young Sign Languages. *Lingua*, 257:103070.

Ted Suppalla. 1986. The Classifier System in American Sign Language. In Colette G. Craig, editor, *Typological Studies in Language*, volume 7, pages 181–214. John Benjamins Publishing Company.

Elizabeth A. Winston. 1991. Spatial Referencing and Cohesion in an American Sign Language Text. *Sign Language Studies*, (73):397–410.