REPRESENTATION OF SIGNED LANGUAGE WITH CONCEPTUAL GRAPHS: A NEW RESEARCH TOOL

SYLVIE LECLERC, SERGE DE MAISONNEUVE
Education Sciences Department, Université du Québec à Rimouski
300 Allée des Ursulines, Rimouski, Québec, G5L 3A1, Canada

and

COLETTE DUBUISSON, CHRISTOPHER MILLER
Linguistics Department, Université du Québec à Montréal
CP8888, suc. A, Montréal, Québec, H3C 3P8, Canada

ABSTRACT

Signed languages are the primary means used by Deaf people to communicate among themselves. They are characterised by the encoding of linguistically significant behaviours by several different articulators such as hands, mouth, head, eyebrows, etc. There is currently no available tool for the analysis of large textual corpora of signed languages. The problem of representation is central to most recent developments in cognitive science. In the study of language, as in the study of other cognitive processes, production of a representation is determining for subsequent analysis and treatment. The need to compute and manipulate is a justification for the development of a representation. Corpus-based analysis of signed language is in need of such a representation. Different notation systems are used depending upon the level of analysis[1]. If a transcription consisted of no more than a linear sequence of such glosses, currently available tools for the analysis of large textual corpora could be used without major modifications. However, a pervasive structural characteristic of signed languages is the simultaneous encoding of linguistically significant behaviours by several different articulators, a characteristic that has no clear counterpart in oral languages.

Two important requirements for a research tool aimed at natural signed language processing are discussed. The first concerns the representation of temporal relations such as sequential, total or partial simultaneity of linguistically significant behaviours. The second is the flexibility the system must exhibit to be adapted to particular researcher's needs. We describe an implementation of conceptual graphs[2] as a representation for Langue des Signes Québécoise (LSQ) the signed language most widely used in Québec. We implement conceptual graphs expressing signs with SAAC, an interactive system for conceptual analysis[3], which constructs an adapted frame-based representation of conceptual graphs. Production rules are proposed as tools for the goals of description and exploration in corpus-based signed language analysis.

REFERENCES

1. S. Prillwitz and H. Zienert, In *Current Trends in European Sign Language Research*, eds S. Prillwitz & T.Vollhaber (1990) Hamburg: Signum-Press (p.355-79).
2. J.F. Sowa, *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley, 1984.
3. S. Leclerc and S. Maisonneuve, *Proceedings of the First International Conference on Conceptual Structures Theory and Applications*. Université Laval, Québec, Canada, (1993) 161-170.

Poster Proceedings of AI '94
Armidale, Australia
p. 65-72

REPRESENTATION OF SIGNED LANGUAGE WITH CONCEPTUAL GRAPHS: A NEW RESEARCH TOOL

SYLVIE LECLERC, SERGE DE MAISONNEUVE
Education Sciences Department, Université du Québec à Rimouski
300 Allée des Ursulines, Rimouski, Québec, G5L 3A1, Canada

and

COLETTE DUBUISSON, CHRISTOPHER MILLER
Linguistics Department, Université du Québec à Montréal
C.P. 8888, Succ A, Montréal, Québec, H3C 3P8, Canada

ABSTRACT

Signed languages are the primary means used by Deaf people to communicate among themselves. They are characterised by the encoding of linguistically significant behaviours by several different articulators such as hands, mouth, head, eyebrows, etc. There is currently no available tool for the analysis of large textual corpora of signed languages. Two important requirements for a research tool aimed at natural signed language processing are discussed. The first concerns the representation of temporal relations such as sequential, total or partial simultaneity of linguistically significant behaviours. The second is the flexibility the system must exhibit to be adapted to particular researcher's needs. This paper describes how an implementation of Conceptual Graphs formalism[1] is used to represent signed language. Rule-based reasoning is proposed as tool for the goals of description and exploration involved in corpus-based signed language analysis.

1. Introduction

The problem of representation is central to most recent developments in cognitive science. In the study of language, as in the study of other cognitive processes, production of a representation is determining for subsequent analysis and treatment. The need to compute and manipulate is a justification for the development of a representation[2]. Corpus-based analysis of signed language is in need of such a representation. Until the present time, signed languages have not been an object of study for computational linguistics. There are problems specific to signed languages that can explain the reasons behind this fact. Our aim is to point out specific problems associated with the study of signed languages and advantages that are sought by taking a computational approach. We describe, therefore, an implementation of conceptual graphs as a representation for Langue des Signes Québécoise (LSQ) the signed language most wideley used in Québec. We exemplify how production rules provide tools for description of a corpus in LSQ.

2. Requirements for the study of signed language

Different notation systems are used depending upon the level of analysis. Phonological transcription, in particular, has seen a proliferation of different notation systems[3-4-5-6], among others. Such systems are very effective tools for phonological analysis but, given that the units of analysis at the phonological level are much smaller and numerous than those at the morphological and syntactic levels, the use of such transcription systems in the context of large corpora quickly becomes too detailed and unwieldy. At the level of morphosyntactic transcription, the solution adopted by most if not all researchers has been the use of glosses. Certain informal conventions have arisen to allow for additional information about the form of signs to be expressed in the form of annotations on the glosses. If a transcription consisted of no more than a linear sequence of such glosses, currently available tools for the analysis of large textual corpora could be used without major modifications. However, a pervasive structural characteristic of signed languages is the simultaneous encoding of linguistically significant behaviours by several different articulators, a characteristic that has no clear counterpart in oral languages. For example, an interrogative sentence could be signed with several components occurring simultaneously. While the non-dominant hand points toward the interlocutor ($INDEX_2$), the dominant hand signs the verb ($_{(nc2)}GIVE_1$), the word *quoi* 'what' is mouthed and, at the same time, the head is tilted back ($h\uparrow$) and the eyebrows raised (RE)[1]. The code (nc2) indicates the spatial locus at or toward which the first part of the sign is articulated (i.e., neutral level, central sector of the horizontal signing space, and two degrees of distance from the body, that is, at the length of the forearm). The code 1, following the gloss, indicates first person. This sentence would be transcribed as in (1).

(1) $\underline{\qquad\qquad\qquad h\uparrow+RE}$

quoi

$_{(nc2)}GIVE_1$

$INDEX_2$__________

The use of an underline with non-manual behaviours serves to indicate the scope of the behaviour with respect to the signs that make up the utterance. Similarly, an extended line following a gloss indicates that the sign is held during the other behaviours with which it co-occurs.

Special requirements are necessary for the automated processing of large corpora involving multiple simultaneous channels as illustrated above. An adequate system must allow the researcher to account for total or partial simultaneity of behaviours in distinct channels; at the same time, sequential aspects of the text, equally important, must also be recognised. The different aspects of a sign's form and meaning can be described in terms of a set of co-occurring parameters tailored to the researcher's needs. Any adequate system must therefore be flexible enough to allow the researcher to specify the parameters relevant to a particular analysis.

3. Conceptual Graphs as representations of signs

Conceptual graphs have been proposed as a formalism for knowledge representation whose expressiveness is attuned to the representation of natural languages. According to

[1] Although, at a more general level of analysis, these behaviors all occur simultaneously, at a more detailed level, the onset of head and eyebrow movements precedes the manual and oral behaviors. These differences in onset time are measured at approximately 0.1 second.

Lehmann[7], conceptual graphs form one of the eight major families of semantic networks systems research (see Sowa[1-8-9] for characteristics of the conceptual graphs formalism). Apart from a variety of natural language applications, conceptual graphs have been used for the representation of feature-related information[10]. The problems posed by the goal of processing signed language have interesting similarities with the problems associated with the representation of feature-related information. In each case, feature-related information is relative to the goals of the user. Inferencing about features is as important for the description of part of a mechanical device as for the description of part of a discourse process.

We represent signs in conceptual graphs by equating parameters with conceptual relations. Signs and values are represented as concepts. For example,

[Sign: #] -
  $(\text{PARAMETER}_1) \rightarrow [\text{Value}_1]$
  $(\text{PARAMETER}_n) \rightarrow [\text{Value}_n]$.

Let us consider the next three utterances. Their meaning is: Peter and Mary met (2). He said to her that he will give me something (3). Do you know what he will give me? (4).

(2)
Pierre
MARY $\text{INDEX}_{(nr2)}$ P $\text{INDEX}_{(nc2)}$ $_{(nr2)}\text{CL-1\^{}MEET}_{(nr2/nc2)}$
$_{(nc2)}\text{CL-1\^{}MEET}_{(nr2/nc2)}$

(3)
donne
$_{(nc2)}\text{SAY}_{(nr2)}$ $_{(nc2)}\text{GIVE}_1$ SOMETHING

(4)
______________ h↑+RE
quoi
$_{(nc2)}\text{GIVE}_1$
INDEX_2______________

Sentence (2) contains five signs. The last sign, normally glossed as MEET, is nevertheless composed of two pronominal classifier morphemes, each of which is initially produced at the spatial locus representing one of the two individuals referred to in the discourse. In LSQ, mouthing of French words often co-occurs with signs. Some signs have obligatory mouthing, with others the mouthing is optional. Here we find three signs with mouthing: P (an initialised sign in finger spelling space) with obligatory mouthing of *Pierre*; GIVE with *donne*; and GIVE with *quoi*. Cases where the gloss and the French word that is mouthed are not semantically connected are examples of one type of simultaneous articulation of lexical units. Below is the representation of these signs in conceptual graphs of the type *composite individual*. The conceptual relations represent the following parameters: gloss (GLSS), realised by hand (HAND), precede (PRCD), mouthed (ORAL), oriented (ORNT), depart position (DPRT), arrival position (ARRV), morpheme (FORM), head movement (HEAD), eyebrows movement (EYBR), finish[2] (FNSH).

[Sentence:
[[Sign: #1] -

---

[2] Based on Allen's temporal relations[11], conceptual relation *finish* expresses that an event begins then is followed by a second event, the end of these two events is simultaneous.

(GLSS)→[Person: Mary]
(HAND)→[Dominant]
(PRCD)→[Sign: #2]-
    (GLSS)→[Index]
    (HAND)→[Dominant]→(ORNT)→[Nr2]
    (PRCD)→[Sign: #3]-
        (GLSS)→[Person: P]
        (HAND)→[Dominant]
        (ORAL)→[Pierre]
        (PRCD)→[Sign: #4]-
            (GLSS)→[Index]
            (HAND)→ [Dominant]→(ORNT)→[Nc2]
            (PRCD)→[Sign: #5]-
                (GLSS)→[Meet]
                (HAND)→[Dominant]-
                    (DPRT)→[Nr2]
                    (ARRV)→[Nc2Nr2]
                    (FORM)→[Cl1]
                (HAND)→[NonDominant]-
                    (DPRT)→[Nc2]
                    (ARRV)→[Nc2Nr2]
                    (FORM)→[Cl1]]

→(PRCD)→[Sentence:
[[Sign: #6] -
    (GLSS)→[Say]
    (HAND)→[Dominant]-
        (DPRT)→[Nc2]
        (ARRV)→[Nr2]
    (PRCD)→[Sign: #7]-
        (GLSS)→[Give: #12]
        (HAND)→[Dominant]-
            (DPRT)→[Nc2]
            (ARRV)→[Signer]
        (ORAL)→[Donne]
        (PRCD)→[Sign: #8]-
            (GLSS)→[Something]
            (HAND)→[Dominant]]

→(PRCD)→[Sentence:
[[Sign: #9] -
    (GLSS)→[Index]
    (HAND)→[NonDominant]→(ORNT)→[Interlocutor]
    (HEAD)→[ChinUp]
    (EYBR)→[RaisedEyebrows]
    (FNSH)←[Sign: #10]-
        (GLSS)→[Give: #13]
        (HAND)→[Dominant]-
            (DPRT)→[Nc2]
            (DPRT)→[Signer]
        (ORAL)→[Quoi]].

We implement conceptual graphs expressing signs with SAAC, an interactive system for conceptual analysis developed by de Maisonneuve. The SAAC system constructs an adapted frame-based representation of conceptual graphs. An essential requirement[12] of this implementation is the declaration of an ontology by means of a network of concept types and relation types. A particular conceptual graph is then represented as an instantiation of types. Provided that the ontology is a superset of the parameters used by a researcher, he can specify values for those parameters relevant to a particular analysis when declaring specific conceptual graphs.

Essentially, there are two processes involved in computer-based corpus analysis. The first one is organised relative to descriptive goals. The researcher seeks to augment textual elements with categories and relations pertinent to the description of the language. The second process is centred on exploration goals. The researcher wishes to test hypotheses about the distribution of different morphosyntactic patterns. For example, an exploratory rule could identify what is mouthed in each occurrence of a sign glossed Give, across different corpora. For each goal, rules express the decision process. We will exemplify the first process by a descriptive rule for inferencing the conceptual relation *agent*.

4. Rules as tools for description

The following rule is part of a first approximation of what characterises the linguistic behaviours associated with the expression of the conceptual relation *agent* in LSQ.

• Rule 1: **If** a sign, whose gloss is of the type Act, is produced from a specific position, and that sign is preceded[3] by a sign, whose gloss is of the type Index, oriented in a specific position adjacent[4] to a sign, whose gloss is of the type Animate, and the specific position of these two signs is the same, **then** the position of the sign, whose gloss is of the type Index, is the grammatical marker of a concept Agent, and this concept Agent is linked to the concept Animate, and this concept Agent is linked to the concept Act.

Rule 1 is formalised[9] as:

```
IF: [Sign: *t]-
      (GLSS)→[Animate: *u]
      (PRCD)→[Sign: *v]-
               (GLSS)→[Index]
               (HAND)→ [Hand]→(ORNT)→[Position: *w]
               (PRCD)→[Sign: *x]-
                        (GLSS)→[Act: *y]
                        (HAND)→[Hand]→(DPRT)→[?w]
THEN: [?t]-
      (GLSS)→[?u]→(LINK)→[?z]
      (PRCD)→[?v]-
               (HAND)→ [Hand]→(ORNT)→[?w]→(GRAM)→[Agent: *z]
```

---

[3] At a first level of analysis, we specify the relation *precede* to include an earlier sign within the same sentence. At a further level, we take the domain of the relation to include up to four sentences preceding the sentence in which the sign of the type ACT is present, the upper limit of four sentences being set as an arbitrary hypothesis.

[4] If an index oriented elsewhere than the signer or the interlocutor is *adjacent* to a noun, that is, it immediately precedes or follows the noun or is signed simultaneously with the noun, then the index is interpreted as a determiner and a location fixer.

```
        (PRCD)→[?x]-
                (GLSS)→[?y]→(LINK)→[?z].
```

The application of this rule on the preceding graphs will produce the next graphs. There are five patterns that can unify with the rule. The new parts of the graphs in the first sentence that are asserted by the rule are printed in bold.

```
[Sentence:
[[Sign: #1] -
   (GLSS)→[Person: Mary]→ (LINK) → [Agent:#50]
   (HAND)→[Dominant]
   (PRCD)→[Sign: #2]-
           (GLSS)→[Index]
           (HAND)→[Dominant]→(ORNT)→[Nr2]→ (GRAM) ← [Agent:#50]
           (PRCD)→[Sign: #3]-
                   (GLSS)→[Person: P] -
                                (LINK) → [Agent:#51]
                                (LINK) → [Agent:#52]
                                (LINK) → [Agent:#53]
                                (LINK) → [Agent:#54]
                   (HAND)→[Dominant]
                   (ORAL)→[Pierre]
                   (PRCD)→[Sign: #4]-
                           (GLSS)→[Index]
                           (HAND)→ [Dominant]→(ORNT)→[Nc2] -
                                                  (GRAM) ← [Agent:#51]
                                                  (GRAM) ← [Agent:#52]
                                                  (GRAM) ← [Agent:#53]
                                                  (GRAM) ← [Agent:#54]
                           (PRCD)→[Sign: #5]-
                              (GLSS)→[Meet]→ (LINK) → [Agent:{#50,#51}]
                              (HAND)→[Dominant]-
                                                  (DPRT)→[Nr2]
                                                  (ARRV)→[Nc2Nr2]
                                                  (FORM)→[Cl1]
                              (HAND)→[NonDominant]-
                                                  (DPRT)→[Nc2]
                                                  (ARRV)→[Nc2Nr2]
                                                  (FORM)→[Cl1]]
  →(PRCD)→[Sentence: ...
```

The conceptual relation (GRAM)13 plays a fundamental role in our text analysis strategy. This relation unites grammatical markers specific to the conceptual elements relevant to the semantic analysis of the text. The relation (GRAM) is defined as follows:

relation GRAM (x,y) **is**

[LinguisticEntity:*x]→(LINK)→[Gram]→(LINK)→[ConceptualEntity:*y]

The relation (GRAM) expresses the relation between a conceptual entity and a linguistic entity. The linguistic entity is recognised as a grammatical marker, a concept of the type Gram, associated with the presence of the conceptual entity in the sentence. The relation (GRAM) thus makes it possible to explicitly associate the syntactic and semantic levels that are relevant to a particular analysis. It comes into play in the definition of all conceptual relations for which the analyst desires to formalise the link between syntax and semantics. In the example that we develop here, the conceptual relation (AGNT) incorporates the conceptual relation (GRAM) in its definition.

```
relation AGNT (x,y) is
[Agent] -
    (LINK)→[Animate: *x]
    (LINK)→[Act: *y]
    (GRAM)→[PronomPosition]
```

In this definition of the conceptual relation (AGNT), the concept [PronomPosition] is linked by the (GRAM) relation to the concept [Agent]. The concept [PronomPosition] is of the type LinguisticEntity, it groups together the positions that manifest a reference to a person. The concept [Agent] is linked to the concept [Animate: *x]. As in all definitions of dyadic conceptual relations, there are two concepts that correspond to two different points of view on a same object. The animate is an agent with respect to the action.

These two definitions permit the *contraction*1 of the preceding graphs. The contracted form of the graph corresponding to the first sentence includes the five relations agent that have been inferred by the rule.

```
[Sentence:
[[Sign: #1] -
  (GLSS)→[Person: Mary]→ (AGNT) → [Meet:#14]
  (HAND)→[Dominant]
  (PRCD)→[Sign: #2]-
          (GLSS)→[Index]
          (HAND)→[Dominant]→(ORNT)→[Nr2]
          (PRCD)→[Sign: #3]-
                  (GLSS)→[Person: P] -
                                (AGNT) → [Meet:#14]
                                (AGNT) → [Say:#15]
                                (AGNT) → [Give: #12]
                                (AGNT) → [Give: #13]
                  (HAND)→[Dominant]
                  (ORAL)→[Pierre]
                  (PRCD)→[Sign: #4]-
                          (GLSS)→[Index]
                          (HAND)→ [Dominant]→(ORNT)→[Nc2]
                          (PRCD)→[Sign: #5]-
                                  (GLSS)→[Meet:#14]
                                  (HAND)→[Dominant]-
                                              (DPRT)→[Nr2]
                                              (ARRV)→[Nc2Nr2]
                                              (FORM)→[Cl1]
                                  (HAND)→[NonDominant]-
                                              (DPRT)→[Nc2]
                                              (ARRV)→[Nc2Nr2]
                                              (FORM)→[Cl1]].
```

5. Conclusion

This paper describes an ongoing research effort oriented toward the development of corpus-based signed language analysis. It has been proposed that conceptual graphs can provide a representation adapted to specific problems associated with signed languages, particularly the simultaneous encoding of linguistically significant behaviours by different articulators. An example of the representation of utterances from written notation of signs to conceptual graphs has been provided. Production rules, along with formation rules usually applied to conceptual graphs, have been proposed as tools for the goals of description and exploration in corpus-based signed language analysis. The feasability of the approach has been demonstrated by applying a rule inferencing the conceptual relation *agent* to utterances represented by conceptual graphs.

6. References

1. J.F. Sowa, *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley, 1984.
2. P. Suppes, M. Pavel and J.-Cl. Falmagne, *Ann. Rev. of Psychology*, ***45*** (1994) 517-44.
3. W. Stokoe, *Studies in Linguistics, Occasional Papers. 8,* Buffalo: University of Buffalo, 1960.
4. W. Stokoe, D. Casterline, D. and Croneberg, *A Dictionary of American Sign Language on Linguistic Principles*. Silver Spring, Md: Linstok Press (1976 edition).
5. S. Prillwitz and H. Zienert, In *Current Trends in European Sign Language Research,* eds S. Prillwitz & T.Vollhaber (1990) Hamburg: Signum-Press (p.355-79).
6. P. Jouison, In *Current Trends in European Sign Language Research,* eds S. Prillwitz & T.Vollhaber (1990) Hamburg: Signum-Press (p.337-53)
7. F. Lehmann, *Computers Math. Applic.* **23**, (1992) No. 2-5, 1-50.
8. J.F. Sowa, *Computers Math. Applic. 23*,(1992) No. 2-5, 75-93.
9. J.F. Sowa, In *Conceptual Graphs for Knowledge Representation,* eds G. W. Mineau, B. Moulin and J. F. Sowa (1993), Berlin: Springer-Verlag.1-35.
10. R.E. Billo, M. Henderson and R. Rucker, *Elsevier Computers in Industry 13*,(1989) 195-214.
11. J.F. Allen, *Artificial Intelligence 23 (2)*(1984) 124-54.
12. J.F. Sowa and E.C. Way, *IBM J. RES. DEVELOP. 30:1*, (1986) 57-69.
13. S. Leclerc and S. Maisonneuve, *Proceedings of the First International Conference on Conceptual Structures Theory and Applications*. Université Laval, Québec, Canada, (1993) 161-170.