

## Determining aspects of text difficulty for the Sign Language of the Netherlands (NGT) Functional Assessment instrument

Annieck van den Broek-Laven, E. Boers-Visker & B. van den Bogaerde  
Hogeschool Utrecht, University of Applied Sciences

In this paper we describe our work in progress on the development of a set of criteria to predict text difficulty in Sign Language of the Netherlands (NGT). These texts are used in a four year bachelor program, which is being brought in line with the Common European Framework of Reference for Languages (Council of Europe, 2001). Production and interaction proficiency are assessed through the NGT Functional Assessment instrument, adapted from the Sign Language Proficiency Interview (Caccamise & Samar, 2009). With this test we were able to determine that after one year of NGT-study students produce NGT at CEFR-level A2, after two years they sign at level B1, and after four years they are proficient in NGT on CEFR-level B2. As a result of that we were able to identify NGT texts that were matched to the level of students at certain stages in their studies with a CEFR-level. These texts were then analysed for *sign familiarity*, *morpheme-sign rate*, *use of space* and *use of non-manual signals*. All of these elements appear to be relevant for the determination of a good alignment between the difficulty of NGT signed texts and the targeted CEFR level, although only the morpheme-sign rate appears to be a decisive indicator.

**Key words:** CEFR, Sign Language, Comprehension, Text difficulty, Higher Education

## Introduction

At the Hogeschool Utrecht, University of Applied Sciences in the Institute for Sign, Language & Deaf Studies (ISLD) we educate teachers and interpreters in Sign Language of The Netherlands (NGT) in a four year bachelor program. The program was founded in 1997, and up to 2010 a grammar-oriented approach was used to teach students sign language. New insights in the field of language pedagogy, the introduction of the Common European Framework of Reference for Languages (CEFR, Council of Europe, 2001) and changes in the internal organisation of our university led to the decision to change our teaching approach and align it to the CEFR. The development of the new curriculum (see also Efsli (2013a) and implementation of the new didactics took place between 2010 and 2014.

The Council of Europe (2011) describes the CEFR as follows:

The CEFR provides a common basis for the explicit description of objectives, content and methods in foreign language education and adopts an action-oriented approach, describing language learning outcomes in terms of language use. There are three main dimensions: language activities, the domains in which they occur, and the competences on which we draw when we engage in them. The CEFR for spoken languages presents four types of language activities: reception (listening and reading), production (spoken and written), interaction (spoken and written) and mediation (translating and interpreting).

For signed languages the terms *spoken* production and interaction are replaced by *signed* production and interaction. Since signed languages have no written form, reading and writing are substituted by watching/producing video-recorded texts in signed language (e.g. see ATERK, 2013; Leeson & Byrne-Dunne, 2009). The CEFR distinguishes the following language domains: public, personal, educational and professional. There are six common reference language levels (A1, A2, B1, B2, C1 and C2) and 'can-do' descriptors to define the learner's proficiency at each level. These common reference levels were used to design our NGT curriculum, the materials and the assessment procedures (see also McNamara, 2000; Efsli 2013b).

In our program the student-teachers and student-interpreters follow the same NGT language courses during the first two years of the program. After these two years, their respective curricula differ in the professional content courses during which further NGT acquisition takes place as well as in the extensive practice periods. For student-teachers the sign language instruction corresponds to 60 ECTS (European Credit Transfer System) that are equivalent to 1,680 hours of study, of which 420 hours consist of in-class sign language

contact. For student-interpreters the NGT instruction corresponds to 75 ECTS, equivalent to 2,100 hours of study, including 525 hours of in-class sign language contact.

Although the CEFR does not dictate any language teaching methodology (Council of Europe, 2001, p. 142; Pearson Longman, n.d.), the can-do statements indicate a communication-oriented approach. We decided to alter our didactics accordingly. Besides introducing communication-oriented or action-oriented tasks, we made some additional alterations. One of these alterations was the decision to avoid using written Dutch as much as possible as language of instruction and to predominantly use the target language NGT in our in-class and online contacts.

Together with the new didactic approach, we introduced a new assessment method for sign language production and interaction: the NGT Functional Assessment. As a result of that we can determine the CEFR-level of NGT production/interaction fluency of our students at different stages of their studies. This was the starting point for reviewing the CEFR-level of our NGT materials. Although the CEFR gives an indication on what a text can look like at a certain level, there are no strict rules or described features for written or signed texts at a particular level. Even more, as far as we know, no research has as yet been conducted for signed languages on the determination of text difficulty, and certainly not in relation to establishing CEFR language levels. By selecting texts that were understood well by students on a particular production/interaction level, we estimated the CEFR-level of these texts. The texts were then analysed to develop criteria in order to predict the CEFR-level of a certain text.

In the following sections we will describe the adaptation and translation of the Sign Language Proficiency Interview (SLPI) into the NGT Functional Assessment (NFA), the assessment procedures and the implementation of the NFA in our curricula. Subsequently we will describe several textual aspects which we took into consideration to determine whether or not and how these influenced NGT text difficulty, in order to be able to select matching NGT texts to be used in comprehension tests. For this study we formulated the following research question: Can specific textual features be used to predict the difficulty of the NGT signed texts at each CEFR level?

### **The SLPI and the NFA**

The Sign Language Proficiency Interview for American Sign Language (SLPI-ASL) is described by the National Technical Institute for the Deaf (2014) as an interview tool to assess a person's skills in American Sign Language that is

widely used across the United States and Canada (Caccamise & Samar, 2009). The SLPI was originally adapted from the Language/Oral Proficiency Interview (L/OPI), an interview technique for assessing spoken language communication skills (Newell, Caccamise, Boardman, & Holcomb, 1983). The SLPI website is hosted by the National Technical Institute for the Deaf (NTID) at the Rochester Institute of Technology (RIT) in the USA. The Coordinator of NTID's Office of ASL Training and Evaluation together with an expert linguist of ISLD initially translated and adapted the SLPI to a Dutch version for NGT, the NGT Functional Assessment (NFA).

The NFA procedure is as follows. The candidate is interviewed by an interviewer for 20 minutes. The interview is recorded with both the interviewer and the candidate in view. The interviewer uses SLPI standard interview techniques (Caccamise & Newell, 2010), in which s/he is trained to elicit the candidate's best possible NGT. The candidate is asked questions about different topics such as study, work, family and hobbies. Subsequently the interview is rated by two independent raters. The raters use a rater-sheet to determine at which level the candidate performs, taking *form* and *function* into consideration. In determining the *function* level, the rater measures the ability of the candidate to take part in a conversation. *Form* focuses on (1) vocabulary knowledge, (2) signing rate, (3) fluency, (4) grammar, and (5) comprehension (Caccamise & Newell, 1999). Rating takes approximately 40 minutes per rater per interview, including discussion when necessary. In Figure 1 we present the NFA procedure.

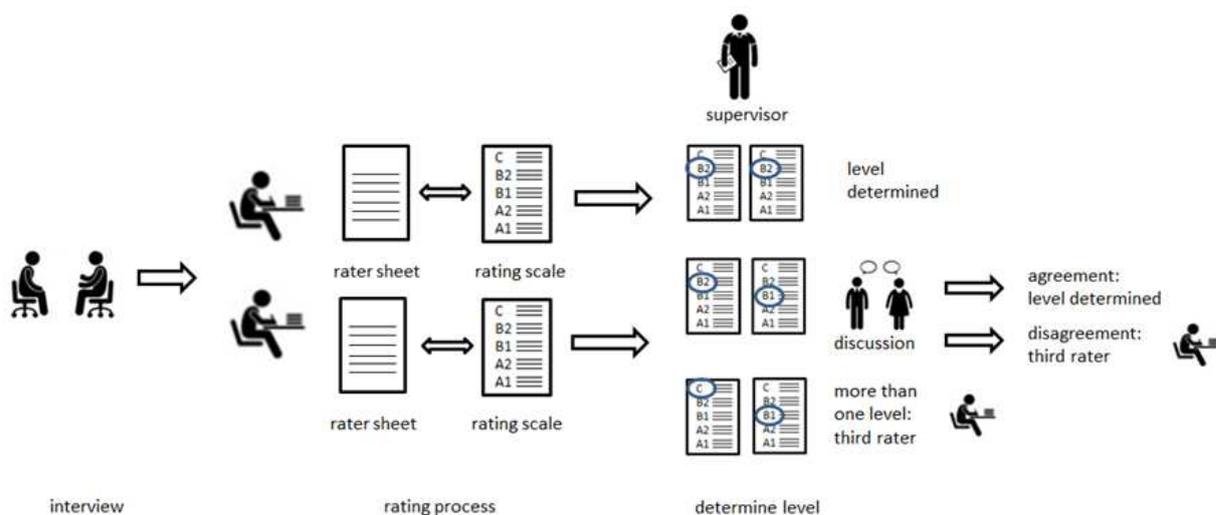


Figure 1. Schematic presentation of NGT Functional Assessment procedures

Raters mark examples of a particular aspect/achievement of the candidate on the rater sheet. These notes are compared to the descriptors of the rating scale.

To reach a certain level, the candidate must demonstrate proficiency for *all* descriptors of that level.

The raters send their results to a supervisor, who compares the outcomes. If the raters agree, the level is determined as such. If they disagree, they are to discuss their argumentations to come to an agreement about the allocation of the CEFR level. In case raters differ more than one level, or when they fail to reach agreement after discussion, the supervisor will assign a third rater. The third rater's decision is final.

In the next section we will describe the SLPI to NFA adaptation process in more detail.

### **The SLPI-NFA adaptation process**

During the process of the NFA-adaption, all documents needed for the instrument, e.g. the descriptors, the rater sheet and the rating scale, were translated from written English into written Dutch. During this process, all documents translated into Dutch and adapted for NGT were subsequently translated back into English for the SLPI-ASL coordinator to be checked and discussed.

In order to adapt the SLPI as an instrument to assess NGT, we had to take into account (1) the fact that the language levels used in the USA and Canada differ from the CEFR-levels and (2) the descriptors to mark grammar needed to be adapted from ASL to NGT.

First we focused on the conversion of the ASL-proficiency levels to CEFR-levels. The SLPI distinguishes eleven levels of ASL-proficiency where the CEFR only has six levels. By comparing the descriptors of the SLPI-levels and the descriptors used to indicate CEFR-levels for signed languages as described by Leeson & Byrne-Dunne (2009) (see appendix A), the eleven SLPI-levels were aligned with five CEFR-levels: A1 (comparable to SLPI Novice), A2 (comparable to Survival), B1 (comparable to Intermediate), B2 (comparable to Advanced), C1 and C2 (Superior).

As levels C1 and C2 are not fully described for signed languages, these are merged, awaiting the outcome of a PRO-sign project as funded by the European Centre of Modern Languages (ECML, 2011, 2014), which focuses on developing descriptors for the C levels for signed languages for professional purposes. Parallel to the SLPI-NFA adaptation process a Dutch national team of researchers and professionals working in the field of sign language teaching prepared the Dutch version of the CEFR for signed languages (ATERK, 2013).

This publication includes descriptors for CEFR sign language levels A1 up to B2 and a few for levels C1/C2.

Second, we focused on the grammatical differences between ASL and NGT. We examined which ASL grammatical features listed on the rating sheet also occurred in NGT, which features did not and so could be removed from the rater sheet and rater scale, and which grammatical features that occur in NGT but not in ASL were missing and needed to be added.

With the full set of documents belonging to the SLPI translated and adapted to NGT, we had two training sessions by the NTID expert coordinator, a 28-hour training period in December 2011 followed by a 21-hour training in August 2012. All NGT teachers and some Interpreting teachers (n=13) were trained in taking the interviews, in scoring, and in assigning a CEFR level to the candidate's performance. The team consists of deaf and hearing teachers who use NGT as a first or second language. During these training sessions, the documents were continuously adapted based on the teachers' experiences and feedback. At the end of the courses all participants were accredited by NTID for employment as NFA interviewers and/or raters. Following consultation with all the trained NFA teachers the NFA set of documents were finalised in July 2013.

In April 2013 we conducted a sample survey, interviewing and rating ten percent of our students in the first year of their studies. We found that these first year students (N=6) scored CEFR-level A2. Also it was decided that in line with the other languages taught at our university, the students should be proficient in NGT at level B2 at the end of their studies. After the second year of their studies they are therefore expected to have reached level B1. Since September 2013 the NFA was officially adopted as a summative test in our bachelor program. Figure 2 depicts the NGT modules in our curriculum, as well as the expected CEFR-level at the three points in their studies.

|                 | year 1 |   |   |   | year 2 |   |   |   | year 3                                    |   | year 4  |   |
|-----------------|--------|---|---|---|--------|---|---|---|---|---|---|---|
|                 | A      | B | C   | D | E      | F | G | H   | Action<br>embedded<br>language<br>contact | I | J   | Action<br>embedded<br>language<br>contact |
| NFA-assessments |        |   | <br>A2 |   |        |   |   | <br>B1 |   |   | <br>B2 |   |

**Figure 2.** NGT modules and expected CEFR-level

In the next section we will discuss the rationale for our study on the difficulty of NGT texts related to CEFR levels, and the method and results of our ongoing study.

### **Comprehension tests**

Since the end-level of NGT courses (C, H, J in Figure 2) were aligned with a CEFR level (Figure 2), we needed to be sure that the materials we offered the students were a properly aligned to the targeted CEFR language level. There was an urgent need for a set of criteria that could predict NGT text difficulty, aligned to the CEFR. Texts for comprehension thus far were selected or designed intuitively by NGT teachers and proved to be appropriate or not for a certain level during their use. The development of the CEFR for signed languages (ATERK, 2013) offered a stronger basis for selecting signed texts, as the can-do statements for sign languages formed a solid starting point. Also, themes that should be addressed in the different sign language courses were already linked to the CEFR.

New materials that were developed for new comprehension tests were produced, transcribed, and analysed for this study. The results of this study are described below, after a short theoretical description.

### **Comprehension in signed languages**

As signed languages have no written form, comprehension consists of understanding a signer either in real-life (face to face) or via receptive sign language exercises, in the form of recorded materials (Leeson & Byrne-Dunne, 2009) (see Appendix). These two forms do differ, as a signer in real-life can be viewed in three-dimensions, where a film is two-dimensional. Even though watching and understanding recorded signed texts cannot directly be compared to reading and understanding a written text, they are similar in that both forms can be (re-)styled and altered during production, can be viewed again and again, without alteration(s) in the content, in contrast to spoken or signed face to face production. In our bachelor program NGT film clips are being used for NGT comprehension assessment and self-study. In the following we will refer to these recorded films as signed texts.

### **Criteria for determination of difficulty**

There are many aspects of signed languages that could influence text difficulty (see e.g. Nilsson, 2010; Winston & Monikowski, 2000). In order to determine which specific aspects should be taken into account to establish the difficulty of signed text, first we will discuss difficulty of texts for comprehension in spoken

languages. In this we view listening and reading as forms of comprehension, which will be compared to comprehension in signed languages. Secondly we review certain elements specific for signed languages that could influence NGT text difficulty.

### *Sign familiarity*

Crossley, Allen, and McNamara (2011) describe how readability studies have introduced different readability formula (Bamford, 1984; Brown, 1998; Carrell, 1987; Greenfield, 2004). In these traditional formulas, difficulty of texts is measured by word length and sentence length. Carrell (1987) put forward that for L2 texts reading, more accurate readability formulas are needed to ensure a good match to the language level of a L2-learner. These factors should include coherence (Gernsbacher, 1997; McNamara, Kintsch, Butler-Songer, & Kintsch, 1996) and the percentage of frequent words (Andringa & Hacquebord, 2000; Staphorsius, 1994). Leroy and Kauchak (2014) however, state that there is little evidence that readability formula outcomes relate to text understanding, as it might be that too strong a reliance is put on word and sentence length. Therefore they evaluated word familiarity as stand-in for word difficulty. They found that difficulty indeed correlated with word familiarity and not with word length.

Until recently, little research has been done on the frequency of signs in different sign languages. Only a few studies had been published (Cormier, Schembri, Fenlon, Rentelis & Reynolds, 2011; Johnston, 2012; McKee & Kennedy, 2006; Morford & MacFarlane, 2003). However, Johnston (2012) mentions that a part of the data from these studies probably is unreliable, as the data sets used were highly controlled compilations of the sign corpora used for these studies. Sign-frequency lists cannot be generated from these studies. Fenlon, Schembri, Rentelis, Vinson, and Cormier (2014) describe the first study for British Sign Language (BSL) on lexical frequency, based entirely on spontaneous conversational data. They found that the frequency of specific lexical items appears to differ depending on the type of texts that are used in the dataset.

Researching BSL, Vinson, Cormier, Denmark, Schembri, and Vigliocco (2008) state that although familiarity appears to be dissociable from objective counts of lexical frequency (Colombo, Pasini, & Balota, 2006), the strong correlation between the two allows *sign familiarity* to be used as a proxy for frequency until such time as objective measures of sign frequency become available.

There has been quite a lot of research on the relation between text understanding and the percentage of known vocabulary (e.g. Nation, 2008;

Vermeer, 2006). Nation (2008) puts forward that for any reader at least 98% of the vocabulary should be known in order to fully understand a text and its message. In addition to this, Van Zeeland and Schmitt (2013) found that for L2-learners, at least 95% coverage is needed in order to understand a listening task. As Marschark et al. (2005) found that students process spoken, written and signed information in the same way, we can assume that the percentages above also apply to signed texts.

In the ten NGT courses in our program, prescribed lexicon-lists are provided to the students to be studied, and thus we can determine quite reliably whether or not students should know certain vocabulary. However, as students learn a substantial amount of other vocabulary during sign language classes and in other language activities and situations as well, in this study we assume that the percentage from the courses' lexicon-lists to occur in the texts can be lower than 95% to make the text fully understandable for students. In this study we hope to find a more accurate percentage as a guide to selecting appropriate texts.

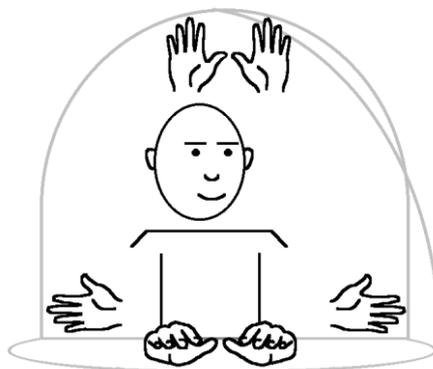
#### *Morpheme-sign rate*

In signed languages it is hard to determine length of signs, first because the parameters (handshape, location and movement) of signs are basically different from phonemes in words. Second, because these parameters are usually produced simultaneously, whereas phonemes are produced sequentially. Third, even though various phonetic and phonological theories for signed languages have been introduced (Brentari, 1998; Crasborn 2001; Liddell, 2003; Sandler 1989; Sandler & Lillo-Martin, 2006; Stokoe, 1960), Johnson and Liddell (2010) conclude that while the evidence for segmenting signs is compelling, each of the proposed systems of segmentation has significant problems. As there is no agreement yet on a comprehensive phonetic system that fully explains the formation of parameters and thus the structure of signs, and because of simultaneity issues on phonetic, phonological and morphological levels, no reliable standard exists yet to measure sign length.

What can be determined is how many morphemes a sign or a sentence consists of – if we can determine what the definition of a signed sentence is. As there also is no overall agreement yet on how to unequivocally define a sentence (Crasborn, 2008), we decided to determine the *morpheme-sign rate*: the number of morphemes per sign, measured over a minute of discourse. This m/s-rate is expected to yield information about the morphological complexity in a text.

### *Use of space*

Signed languages use space for grammar in a way that spoken languages do not (Stokoe, Casterline, & Croneberg, 1965). 'By the signing space we mean the space in front, next and above the body of the signer, in which sign language is articulated, and which extends as far as the stretched arms can reach' (Nijen Twilhaar, & van den Bogaerde, in press) (Figure 3).



**Figure 3.** The signing space. © Van Tricht. Used with permission from Nijen Twilhaar, 2009.

The signing space can be exploited referentially (Sandler & Lillo-Martin, 2006, pp. 24-25). The physical location of the signer is the location used for signs that refer to first person (e.g. point to oneself). The presence of an addressee or third person during discourse establishes a location that can be used for reference (e.g. point to addressee). When reference is made to someone/something not present, a locus in space must be associated with that referent (nominal establishment). Once a locus is assigned to a referent X, pronominal signs can then be (re-)directed to that locus to refer back to X. Verb agreement follows pronouns in making use of referential loci in space. A verb agreeing both with subject and object uses the location of the subject as the beginning location and the locus of the object as the end location (Sandler & Lillo-Martin, 2006, p. 27). Another typical feature of sign languages is the use of classifiers, i.e. a handshape that together with a place, orientation and movement of the hand(s) can form a verb of motion or location. A classifier that is incorporated in the verb give can indirectly represent the form of the handled object (the manner in which that object is held, e.g. I-CL<sub>thick-book</sub>-GIVE-YOU). Classifiers can also be used for nominal establishment and anaphoric reference.

In this study space is analysed for anaphoric reference via loci, redirection to these loci, agreement verbs and locationally used classifiers, to see if these separate elements influence text difficulty.

### *Non-manual signals*

Non-manual parts of a sign are for instance the facial expression, the movement or posture of the head of the upper body and mouth movements. Examples of non-manual grammatical markers (NMGM) are e.g. the marker for sentence type, for negation and affirmation or topicalisation (Nijen Twilhaar et al., in press). Van den Bogaerde et al. (2011) found an increased complexity of NMGMs used in NGT signed texts at CEFR-level A1 and A2. In this study we not only focus on NMGM's, but on non-manual signals in general.

## Method

In order to establish a set of criteria to predict NGT text difficulty, first a series of texts was selected from the assessments already in use. These texts had been shown to match the student's level at particular moments of their studies. In total we analysed 13 NGT signed texts, signed by Deaf (native) signers, aged 11-55, as can be seen in Table 1. Signers were classified according to whether or not they used the autocue for Dutch keywords or written Dutch texts, and if so, if they were experienced translators.

**Table 1.** Meta-data for analysed NGT signed texts

| Text # | Signer # | Text level | Signer age | Text topic       | Autocue (Y)es/(N)o | Experienced translator (Y)es/(N)o |
|--------|----------|------------|------------|------------------|--------------------|-----------------------------------|
| 1      | 1        | A1         | 50-55      | Calendar         | N                  |                                   |
| 2      | 1        | A1         | 50-55      | Calendar         | N                  |                                   |
| 3      | 2        | A2         | 25-30      | Holidays         | Y                  | N                                 |
| 4      | 2        | A2         | 25-30      | Holidays         | Y                  | N                                 |
| 5      | 2        | A2         | 25-30      | Holidays         | Y                  | N                                 |
| 6      | 2        | A2         | 25-30      | Holidays         | Y                  | N                                 |
| 7      | 3        | B1         | 50-55      | Book review      | Y                  | Y                                 |
| 8      | 4        | B1         | 20-25      | Children's story | Y                  | Y                                 |

---

|    |   |    |       |                  |   |   |
|----|---|----|-------|------------------|---|---|
| 9  | 4 | B1 | 20-25 | Children's story | Y | Y |
| 10 | 5 | B1 | 20-25 | School trip      | Y | N |
| 11 | 7 | B2 | 20-25 | Fairytale        | Y | Y |
| 12 | 7 | B2 | 20-25 | Fairytale        | Y | Y |
| 13 | 8 | B2 | 11    | Holidays         | N |   |

---

Before each recording we asked the (native) adult signers to intuitively sign at the right speed for the relevant level. We also asked the adult signers to structure their text so that it would be a coherent (often chronological) text. We did not give any instruction on the use of grammar, as we were looking for authentically signed texts, with a natural rhythm. Some signers preferred using an autocue with keywords or written texts in Dutch, others signed from memory.

The NGT texts were transcribed using the software ELAN (The Language Archive, n.d.). Of each text, a sample of 60 seconds was selected arbitrarily, though the selection was led by the clear introduction of a (new) theme or a clear start of a new sentence.

In the transcript we coded signs, morphemes, oral and spoken components, non-manual (grammatical) markers and the use of space. Here, the use of space has been specified in anaphoric reference via loci, redirection to these loci, agreement verbs and nominal establishment, to see if these separate elements influence text difficulty. These modality specific elements do not occur in spoken languages and therefore are new to L2-learners of a signed language (e.g. Johnston & Schembri, 2007). During the transcription process a transcription-logbook was created with directions on how to approach the texts for use in future studies. One person analysing the data also reviewed all transcriptions to ensure uniformity.

The following aspects were analysed:

- Signs: Number of signs in 60 seconds
- Mph: Number of morphemes in 60 seconds
- Loci: Number of unique loci marked in the signing space. These could be pointing (index) signs, or nominal establishment by verb or classifier signs at a specific location in the signing space.
- Loci\*: Number of times the signer uses anaphoric reference via loci, regardless how many times is referred to a unique location.

- Agr verbs/classifiers: Use of verbs that show agreement with the subject and/or the object of the sentence. Also included here is the use of classifier forms.
- NMGM: non-manual grammatical markers, expressed by facial expression, head-tilts, head-nods, head-shakes or body-movement.
- Coverage lexicon: the signs in the text also occur in the lexicon-lists of the relevant NGT module and therefore expected to be familiar.
- Corrected coverage lexicon: see above, but excluding fingerspelling since this was not taught in course NGT A or B.

## Analyses

Below we present all results for the analysed texts (Table 2).

**Table 2.** Analysed samples from the NGT signed texts

| Text      | CEFR      | Signs       | Mph          | Loci       | Loci*       | Agr verbs/<br>classifiers | Non-<br>manual<br>signals | Coverage<br>lexicon | Corrected<br>coverage** | M/S         |
|-----------|-----------|-------------|--------------|------------|-------------|---------------------------|---------------------------|---------------------|-------------------------|-------------|
| 1         | A1        | 61.0        | 67.0         | 1.0        | 2.0         | 0.0                       | 8.0                       | 100.0%              | 100.0%                  | 1.10        |
| 2         | A1        | 58.0        | 65.0         | 2.0        | 5.0         | 0.0                       | 5.0                       | 97.0%               | 97.0%                   | 1.12        |
| <b>Av</b> | <b>A1</b> | <b>59.5</b> | <b>66.0</b>  | <b>1.5</b> | <b>3.5</b>  | <b>0.0</b>                | <b>6.5</b>                | <b>98.5%</b>        | <b>98.5%</b>            | <b>1.11</b> |
| 3         | A2        | 79.0        | 104.0        | 7.0        | 16.0        | 3.0                       | 0.0                       | 94.0%               | 96.0%                   | 1.32        |
| 4         | A2        | 63.0        | 83.0         | 7.0        | 10.0        | 1.0                       | 0.0                       | 90.0%               | 92.0%                   | 1.32        |
| 5         | A2        | 61.0        | 79.0         | 6.0        | 12.0        | 5.0                       | 0.0                       | 90.0%               | 92.0%                   | 1.30        |
| 6         | A2        | 63.0        | 84.0         | 5.0        | 12.0        | 3.0                       | 2.0                       | 92.0%               | 95.0%                   | 1.33        |
| <b>Av</b> | <b>A2</b> | <b>66.5</b> | <b>88.0</b>  | <b>6.3</b> | <b>12.5</b> | <b>3.0</b>                | <b>0.5</b>                | <b>91.5%</b>        | <b>93.8%</b>            | <b>1.32</b> |
| 7         | B1        | 47.0        | 79.0         | 6.0        | 13.0        | 1.0                       | 14.0                      | 91.0%               |                         | 1.68        |
| 8         | B1        | 84.0        | 138.0        | 5.0        | 15.0        | 7.0                       | 5.0                       | 99.0%               |                         | 1.64        |
| 9         | B1        | 83.0        | 134.0        | 3.0        | 19.0        | 10.0                      | 16.0                      | 96.0%               |                         | 1.61        |
| 10        | B1        | 82.0        | 106.0        | 4.0        | 15.0        | 14.0                      | 4.0                       | 98.0%               |                         | 1.39        |
| <b>Av</b> | <b>B1</b> | <b>74.0</b> | <b>114.3</b> | <b>4.5</b> | <b>15.5</b> | <b>8.0</b>                | <b>9.8</b>                | <b>96.0%</b>        |                         | <b>1.58</b> |
| 11        | B2        | 72.0        | 116.0        | 9.0        | 27.0        | 17.0                      | 17.0                      | 93.0%               |                         | 1.61        |

|           |    |             |              |            |             |             |             |              |             |
|-----------|----|-------------|--------------|------------|-------------|-------------|-------------|--------------|-------------|
| 12        | B2 | 83.0        | 147.0        | 6.0        | 35.0        | 14.0        | 21.0        | 96.0%        | 1.77        |
| 13        | B2 | 83.0        | 134.0        | 8.0        | 20.0        | 17.0        | 16.0        | 93.0%        | 1.61        |
| <b>Av</b> |    | <b>85.3</b> | <b>134.0</b> | <b>7.7</b> | <b>27.3</b> | <b>16.0</b> | <b>18.0</b> | <b>94.0%</b> | <b>1.66</b> |

Mph = morphemes Loci = location Agr = agreement NMGM=non-manual grammatical marker M/S morpheme-sign rate

AV = average

\* Number of times anaphoric reference via loci is used, regardless how many times is referred to a unique location

\*\* Corrected coverage shows the coverage excluding fingerspelling, as this is not explicitly taught for level A1

### *Familiarity of vocabulary*

The familiarity of vocabulary is rated by the percentage of signs occurring in the texts that students have explicitly learned at a certain level. For level A1 and A2 we corrected this coverage by excluding fingerspelling, as this is not explicitly taught for level A1 and A2. Overall the data show that all texts have a coverage of familiar vocabulary of 91-100%.

### *M/S rate*

The average signing rate and the average morpheme rate increase over the CEFR-levels (see Table 3), although the range is overlapping for the different levels:

**Table 3.** Morphemes per sign (M/S) rate analysed

| CEFR-level | Signs Av | Signs range | Mph Av | Mph range | M/S Av | M/S range | StdDev M/S |
|------------|----------|-------------|--------|-----------|--------|-----------|------------|
| A1         | 59.5     | 58-61       | 66.0   | 65-67     | 1.11   | 1.10-1.12 | 0.02       |
| A2         | 66.5     | 61-79       | 88.0   | 79-104    | 1.32   | 1.30-1.33 | 0.02       |
| B1         | 74.0     | 47-84       | 114.3  | 79-138    | 1.58   | 1.39-1.68 | 0.13       |
| B1*        | 71.3     | 47-84       | 117.0  | 79-138    | 1.65   | 1.61-1.68 | 0.03       |
| B2         | 85.3     | 72-83       | 134.0  | 116-147   | 1.66   | 1.61-1.77 | 0.09       |

Mph = morphemes M/S = morpheme-sign rate StdDev = Standard Deviation

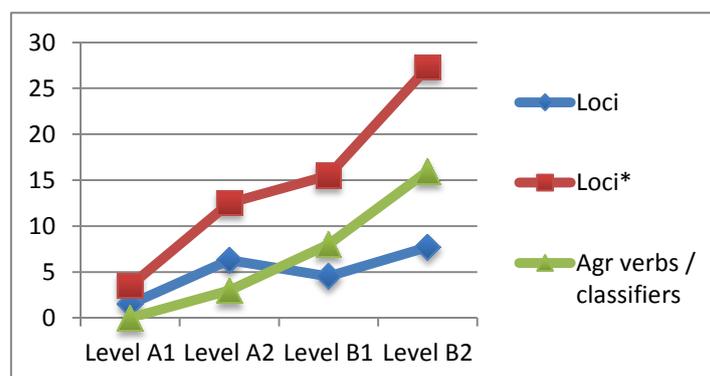
B1\* = level B1 measured over text #7-9

In Table 3 we see that the range of the number of signs per minute and the number of morphemes per minute is fairly broad and can overlap with a higher and/or lower level. However, the morpheme-sign rate for levels A1, A2 and B2

is very consistent with a small range and a small standard deviation (0.02; 0.02; 0.09 respectively). For level B1, where the M/S-rate standard deviation is 0.13, there is an outlier (text #10), which can be explained by taking a look at the signer and the circumstances (see Table 1). The text was translated by an inexperienced signer reading an autocue using Dutch text, which could have caused a reduced use of non-manual signals. If this text is left out, level B1\* is found, which shows a small morpheme-sign range and a small standard deviation (0.03) as well.

### *Use of space*

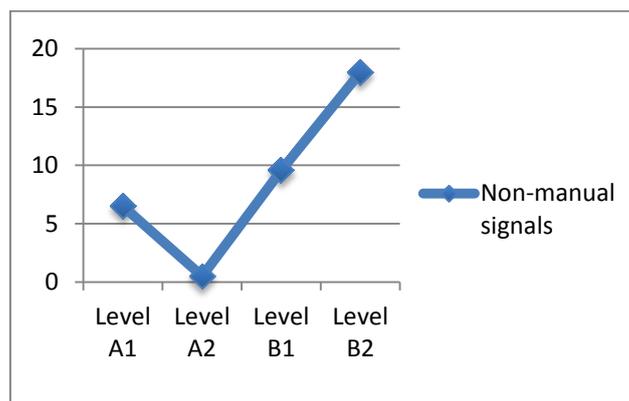
When looking at the use of unique loci (see Figure 4), there does not seem to be an increasing number of unique loci used over the levels. However, if we look at the number of times the signer uses anaphoric reference via loci regardless of how many times it refers to a unique location (loci\*), a clear increase is visible.



**Figure 4.** Use of space.

### *Non-manual signals*

Across the CEFR-levels an increased use of non-manual signals can be found (Figure 5). However, there is also an outlier here: at level A2 a non-manual signal occurred on average 0.5 times in 60 seconds text (see Figure 5). When looking at the meta-data (Table 1) this again can be explained by the use of the autocue by an inexperienced translator: also here written Dutch seems to cause a decrease in the use of non-manual signals. Furthermore the use of non-manual signals also seems to depend on the topic. For example, texts #8 and #9 were signed by the same signer on the same day: in the former text a non-manual signal is applied 5 times, and 16 times in the latter. Although both texts are children's stories, the stories themselves differ and therefore their topic could influence the number of times non-manual signals were used. This means that the feature 'non-manual signal' might not be a strong indicator of text difficulty by itself; however it does appear to influence the M/S-rate.



**Figure 5.** Average of non-manual signs in texts with estimated CEFR level.

## Conclusion

In summary, our institute is following a new sign language teaching approach that uses the descriptors of the CEFR (ATERK, 2013; Council of Europe, 2001; Leeson & Byrne-Dunne, 2009) as a guideline for NGT teaching and assessment. To align the assessment of our students' language proficiency to the CEFR, the Sign Language Proficiency Interview (Caccamise & Samar, 2009) was adapted for NGT. By implementing this NGT Functional Assessment (NFA) we could determine the NGT production/interaction level of our students during different stages of their studies. However for comprehension tests and self-study exercises, a set of criteria was needed to determine NGT text difficulty. We therefore started this initial study, to explore whether specific textual features can be used to predict the difficulty of NGT signed texts at each CEFR level.

For this purpose we analysed NGT signed texts that were known to match the comprehension level of students at certain stages in their studies. As we knew the production/interaction level of our students, we were able to align the texts with specific CEFR-levels. Then these texts were analysed, on the basis of the literature, to determine how specific features (*morpheme-sign rate*, *sign familiarity*, *use of space* and *use of non-manual signals*) characterise these texts.

The analysis showed a *familiar vocabulary* coverage of 91-100%, which seems to confirm our hypothesis that if students explicitly learned 90% of the vocabulary, they should be able to understand the text. In other words, the text would match their language level.

We also found that sign speed and morphemes per minute did not give a clear indication of text difficulty. However, if these are combined in the *morpheme-*

*sign rate* (the average number of morphemes per sign over a minute of text), this seems to be the strongest indicator for text difficulty with standard deviations of 0.02-0.09.

Furthermore, the use of unique loci did not give an indication of text difficulty. The number of times *anaphoric reference via loci* is used regardless how many times is referred to a unique location, and the number of *agreement verbs* or *classifiers* used to show agreement with the subject and/or the object of the sentence do seem to give an indication of text difficulty. This also applies to the number of times *non-manual signals* are being used. However, these factors are not decisive as the range was fairly broad.

There are a few limitations to our research design. The use of space and the use of non-manual signals seem to be strongly influenced by the topic of the text and the use of Dutch written text in the autocue. The use of an autocue with written key words or written texts that are translated into signed texts, strongly influenced the use of non-manual signals, unless the signer was an experienced translator. Further research should be conducted to define who is an 'experienced translator' and to develop alternative ways of auto-cueing in NGT.

Also the fact it is necessary to ascertain that a text matches a particular CEFR-level, as a prerequisite to analyse the features of a text for that language level, limits the choice of texts and complicates the analysis. However, once it is done, it does provide a good starting point.

This study, to our knowledge, is the first of its kind. In exchange with other institutions with similar programs, we have learned that our focus of research is interesting both for teachers and researchers, and fills a gap in our knowledge about the use of signed texts as teaching and testing materials. The impact of our research has yet to be determined, but we are confident that others will join us in this research, and that new knowledge will rapidly improve our teaching and testing approach.

As this is a work in progress, our team of researchers and teachers will continue to study the texts used for, and the CEFR-levels of, comprehension tests. In this we will continue to benchmark, standardise and validate our texts within the larger framework of the CEFR for signed languages.

### Acknowledgements

We are grateful to Annemiek Hammer for her useful comments, Karin Vinke for working with us by transcribing NGT texts, and to the teachers and students

who participated in our study. We would also like to express our thanks to the anonymous reviewers for their expert feedback.

## References

- Andringa, S. J., Hacquebord, H. (1994). De moeilijkheidsgraad van schoolboekteksten als grondslag voor het vaststellen van tekstbegripsvaardigheid. *Toegepaste Taalwetenschap in Artikelen*, 64, 83–94.
- ATERK (2013). *Europees Referentiekader voor Talen: Gebarentalen*. Hogeschool Utrecht. Available at [https://www.academia.edu/3748474/ATERK\\_NGT\\_2013\\_Europees\\_Referentiekader\\_voor\\_Talen\\_Gebarentalen\\_CEFR\\_CEFR\\_for\\_Sign\\_Languages\\_Dutch\\_version](https://www.academia.edu/3748474/ATERK_NGT_2013_Europees_Referentiekader_voor_Talen_Gebarentalen_CEFR_CEFR_for_Sign_Languages_Dutch_version)
- Bamford, J. (1984). Extensive reading by means of graded readers. *Reading in a Foreign Language*, 2(2), 218–260.
- Boers, E. & Van den Bogaerde, B. (2013, 28-31 August). *Second language learning in the visual modality: from phonology to narratives. Use of space in six L2 learners of a signed language*. Paper presented at the 23<sup>rd</sup> EUROSLA conference, Amsterdam, The Netherlands.
- Brentari, D. (1998). *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press.
- Brown, J. D. (1998). An EFL readability index. *JALT Journal*, 20, 7–36.
- Caccamise, F. C., & Samar, V. J. (2009). Sign Language Proficiency Interview (SLPI): Pre-negotiation interrater reliability and rater validity. *Contemporary Issues in Communication Science & Disorders*, 36, 36–47.
- Caccamise, F., & Newell, W. (1999). *Section 13: An overview of the Sign Communication Proficiency Interview (SCPI): History, development, methodology & use*. Manuscript, National Technical Institute for the Deaf, Rochester Institute of Technology.
- Caccamise, F., & Newell, W. (2010). *Section 3A: Program: Sign Language Proficiency Interview (SLPI): American Sign Language (SLPI-ASL). scheduling and interviewing procedures*. Manuscript, National Institute for the Deaf, Rochester Institute of Technology. Retrieved May 2014, from <http://www.rit.edu/ntid/slpi/materials>.
- Carrell, P. L. (1987). Readability in ESL. *Reading in a Foreign Language*, 4(1), 21–40.
- Chouise, C., Dugovicova, S., Fischer, J., & Virkkunen-Fullenwider, A. (2012). *Guidelines for task-based university language testing*. Graz, Austria: ECML.

- Colombo, L., Pasini, M., & Balota, D. A. (2006). Dissociating the influence of familiarity and meaningfulness from word frequency in naming and lexical decision performance. *Memory & Cognition*, 34(6), 1312-1324.
- Cormier, K., Schembri, A., Fenlon, J., Rentelis, R., & Reynolds, S. (2011). *Lexical frequency in British Sign Language conversation: A corpus-based approach*. LDLT3 main session submission.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Council of Europe. Retrieved from [http://www.coe.int/t/dg4/education/elp/elp-reg/Source/Key\\_reference/CEFR\\_EN.pdf](http://www.coe.int/t/dg4/education/elp/elp-reg/Source/Key_reference/CEFR_EN.pdf).
- Council of Europe (2011). *Common European Framework of Reference for languages: Learning, Teaching, Assessment (CEFR)*. Retrieved from [http://www.coe.int/t/dg4/education/elp/elp-reg/cefr\\_EN.asp](http://www.coe.int/t/dg4/education/elp/elp-reg/cefr_EN.asp).
- Crasborn, O. (2001). *Phonetic implementation of phonological categories in Sign Language of the Netherlands*. Ph.D. dissertation, University of Leiden.
- Crasborn, O. (2008). How to recognize a sentence when you see one. *Sign Language & Linguistics*, 38, 103–111.
- Crossley, S. A., Allen, D. B., & McNamara, D. S. (2011). Readability and intuitive simplification: A comparison of readability formulas. *Reading in a Foreign Language*, 23, 84–101.
- ECML (2011). *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). Highlights from the Manual*. Graz: Council of Europe Publishing.
- ECML (2014). *Signed Languages for professional purposes*. Retrieved 11 May, 2014, from <http://www.ecml.at/F5/tabid/867/Default.aspx>.
- Efsl. (2013a). *Learning outcomes for graduates of a three year sign language interpreting training programme*. efsl publication. Available at <http://efsl.org/publications/shop/>.
- Efsl. (2013b) *Assessment guidelines for sign language interpreting training programme*. efsl publication. Available at <http://efsl.org/publications/shop/>.
- Fenlon, J., Schembri, A., Rentelis, R., Vinson, D., & Cormier, K. (2014). Using conversational data to determine lexical frequency in British Sign Language: The influence of text type. *Lingua*, 143, 187–202.
- Gernsbacher, M. (1997). Coherence cues mapping during comprehension. In J. Costermans & M. Fayol (Eds.), *Processing interclausal relationships. Studies in the Production and Comprehension of Text* (pp. 3–22). Mahwah New Jersey: Lawrence Erlbaum Associates.
- Greenfield, J. (2004). Readability formulas for EFL. *JALT Journal*, 26, 5–24.

- Hacquebord, H., & Lenting-de Haan, K. (2012). *Kunnen we de moeilijkheid van teksten meten? Naar concrete maten voor de referentieniveaus*. *Levende Talen Magazine*, 13(2), 14–23.
- Johnson, R. E., & Liddell, S. K. (2010). Toward a phonetic representation of signs: Sequentiality and contrast. *Sign Language Studies*, 11, 241–274.
- Johnston, T., & Schembri, A. (2007). *Australian Sign Language. An introduction to sign language linguistics*. Cambridge: Cambridge University Press.
- Johnston, T. (2012). *Lexical frequency in signed languages*. *Journal of Deaf Studies and Deaf Education*, 17(2), 163–193.
- Leeson, L., & Byrne-Dunne, D. (2009). *Applying the Common European Reference Framework to the Teaching, Learning and Assessment of Signed Languages*. CDS: University of Bristol.
- Leroy, G., & Kauchak, D. (2014). The effect of word familiarity on actual and perceived text difficulty. *Journal of the American Medical Informatics Association*, 21(1), 169-172.
- Liddell, S. K. (2003). *Grammar, gesture and meaning in American Sign Language*. New York: Cambridge University Press.
- Marschark, M., Pelz, M., Jeff, B., Convertino, C., Sapere, P.; Arndt, M. E., & Seewagen, R. (2005). Classroom interpreting and visual information processing in mainstream education for Deaf students: Live or Memorex®? *American Educational Research Journal*, 42, 727–761.
- McKee, D., & Kennedy, G. (2006). *The distribution of signs in New Zealand Sign Language*. *Sign Language Studies*, 6(4), 373–390.
- McNamara, T. (2000). *Language Testing*. Oxford: Oxford University Press.
- McNamara, D. S., Kintsch, E., Butler-Songer, N., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14, 1–43.
- Morford, J., & MacFarlane, J. (2003). Frequency characteristics of American Sign Language. *Sign Language Studies*, 3(2), 213–225.
- Nation, I. S. P. (2008). Lexical awareness in second language learning. In J. Cenoz and N. H. Hornberger, (Eds.), *Encyclopaedia of language and education* (Vol. 6, 2<sup>nd</sup> ed.) (pp. 167–177). New York: Springer Science.
- National Technical Institute for the Deaf. (2014). *SLPI:ASL*. Retrieved from <http://www.rit.edu/ntid/slpi/>.
- Newell, W., Caccamise, F., Boardman, K., & Holcomb, B. R. (1983). Adaptation of the Language Proficiency Interview (LPI) for assessing sign communicative competence. *Sign Language Studies*, 41(1), 311–331.

- Nijen Twilhaar, J. (2009) *Lexicon van de gebarentaalwetenschap*. Deventer: Van Tricht uitgeverij.
- Nijen Twilhaar, J., & Bogaerde, B. van den (in press). *Concise dictionary of sign linguistics*. Amsterdam: John Benjamins.
- Nilsson, A.-L. (2010). *Studies in Swedish Sign Language. Reference, real Space blending, and interpretation*. Doctoral thesis, Stockholm University.
- Pearson Longman (n.d.). *Teacher guide to the CEF*. Retrieved May 2, 2014, from <http://www.euddansk.dk/wp-content/uploads/2013/04/cefguide.pdf>.
- Pfau, R., & Steinbach, M. (2006). Pluralization in sign and in speech: A cross-modal typological study. *Linguistic Typology*, 10, 135–182.
- Sandler, W. (1989). *Phonological representation of the sign: Linearity and nonlinearity in sign language phonology*. Dordrecht: Foris.
- Sandler, W., & Lillo-Martin, D. (2006). *Sign language and linguistic universals*. Cambridge, UK: Cambridge University Press.
- Staphorsius, G. (1994). Leesbaarheid en leesvaardigheid. De ontwikkeling van een domeingericht meetinstrument. *Cito*, Arnhem.
- Stokoe, W. C., Jr. (1960). Sign language structure: An outline of the visual communication system of the American Deaf Studies. *Linguistics Occasional Papers*, 8. University of Buffalo: Department of Anthropology and Linguistics.
- Stokoe, W. C., Jr., Casterline, D.C., & Croneberg, C. G. (1965). *A dictionary of American sign language on linguistic principles*. Washington, D.C.: Gallaudet College Press.
- The Language Archive (n.d.) *ELAN*. Retrieved from <http://tla.mpi.nl/tools/tla-tools/elan/>.
- Van den Bogaerde, B., Boers, E. & Terpstra, W. (2011) Sign language input to students: What and how to offer when? Paper presented at the 33<sup>rd</sup> Annual Conference of the German Linguistic Society (DGfs) *Texts: Structures and Processing*. Göttingen, Germany, 23-25<sup>th</sup> February 2011.
- Van Zeeland, H., & Schmitt, N. (2013). Lexical Coverage in L1 and L2 Listening Comprehension: The Same or Different from Reading Comprehension? *Applied Linguistics*, 34, 457–479.
- Vermeer, A. (2006, March). *Taalachterstanden in het primair onderwijs*. Lecture in Utrecht.
- Vinson, D. P., Cormier, K., Denmark, T., Schembri, A., & Vigliocco, G. (2008). The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behavior Research Methods*, (40)4, 1079-1087.

Winston, E., & Monikowski, C. (2000). Discourse mapping: Developing textual coherence skills in interpreters. In C. Roy (Ed.), *Innovative practices for teaching sign language interpreters* (pp. 15–66). Washington, DC: Gallaudet University Press.

## Appendix

| <b>Real-life Receptive Signed Language Skill</b>  | <b>Receptive Signed Language Exercises</b>  |
|---|---|
| <b>A1</b> I can understand familiar signs and everyday phrases concerning family, my immediate concrete surroundings and myself when people sign slowly and clearly.  | I can understand familiar names, signs and very simple sentences, for example in recorded videos or dvd.  |
| <b>A2</b> I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements.  | I can understand very short, simple recordings. I can find specific, predictable information in recordings, such as advertisements and videos. I can understand short simple personal messages.                             |
| <b>B1</b> I can understand the main points on familiar matters regularly encountered in work, school, leisure, etc. when clear standard language is used. I can understand the main point of many TV programmes on current affairs or topics of personal or professional interest when signed relatively slowly and clearly.  | I can understand recordings that consist mainly of high frequency every-day or job-related language. I can understand the description of events, the expression of feelings and wishes in personal messages.                |
| <b>B2</b> I can understand extended lectures and discourse and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programmes.   | I can understand recordings concerned with contemporary problems in which the authors adopt particular attitudes or viewpoints. I can understand a short theatrical or poetic performance.                                  |
| <b>C1</b> I can understand an extended discourse even when it is not clearly structured and when connections are only implied and not signalled explicitly. I can understand television programmes and works of fiction without too much effort.  | I can understand long and complex, factual and fictional discourse, appreciating distinctions of style. I can understand specialized discourse and longer technical instructions, even when they do not relate to my field. |
| <b>C2</b> I have no difficulty in understanding signed language whether live or broadcast, even when delivered at fast speed, provided I have some time to get familiar with the signer's particular "accent".  | I can understand with ease virtually all forms of recording, including abstract, structurally complex texts, such as lectures, specialized conferences and works of fiction.  |
| <b>Overall Signed Language Production</b>   |   |
| <b>A1</b> Can produce simple mainly isolated phrases about people and possessions.  |   |
| <b>A2</b> Can give a simple description or presentation of people, living or working conditions, daily routines, likes/dislikes, etc. as a short series of simple phrases and sentences linked into a list.   |   |
| <b>B1</b> Can reasonably fluently sustain a straightforward description of one of a variety of subjects within his/her field of interest, presenting it as a linear sequence of points.   |   |
| <b>B2</b> Can give clear, systematically developed descriptions and presentations, with appropriate highlighting of significant points, and relevant supporting detail.<br>Can give clear, detailed descriptions and presentations on a wide range of subjects related to his/her field of interest, expanding and supporting ideas with subsidiary points and relevant examples. |   |
| <b>C1</b> Can give clear, detailed descriptions and presentations on subjects that are complex or even unusual in this signed language, integrating sub-themes, developing particular points and rounding off with an appropriate conclusion.   |   |
| <b>C2</b> Can produce clear, smoothly flowing, well-structured discourse on a topic that may even be unusual in this signed language, with an effective logical structure which helps the recipient to notice and remember significant points.  |   |