ISSN: 1139-7489

AN ANNOTATION SYSTEM FOR SIGNED LANGUAGE INTERPRETING CORPORA

Un sistema de anotación para corpus de interpretación del lenguaje de signos

Ella WEHRMEYER
School of Languages, North-West University, Vanderbijlpark

ABSTRACT: Because of the complexity in transcribing and annotating a visual language, there exists a paucity of corpus-based/driven sign language interpreting research. In this paper, a transcription and annotation system for signed language corpora is outlined. The system was designed to analyse a corpus of interpretations into South African Sign Language, specifically to analyse corpora of signed interpretations using text-based concordance software, such as Wordsmith Tools or Antconc. Transcription of the manual signs is undertaken using lemmatized ID glosses, allowing for semantic comparison between the source message and its interpretation. The annotation system is designed to analyze five main groups of factors, namely linguistic features of the interpreter's sign language produced, characteristics pertaining to the quality of production of a face-to-face message, characteristics of the type of language used by the interpreter (i.e. *parole*) and features of the interpreting process, both in terms of describing differences between the source and target messages and in terms of identifying phonological, syntactic and discourse errors made by the interpreter.

© Fatiso

Key words: signed language interpreting, corpus annotation, corpus linguistics, signed language transcription, Interpreting Studies, interpreting strategies, interpreting errors, sign language studies, South African Sign Language.

RESUMEN: Debido a la complejidad de la descripción y anotación del lenguaje visual, existe muy poca investigación en la interpretación de la lengua de signos basada en u orientada a la compilación de corpus. En este artículo, se describe diseña un sistema de transcripción y anotación para la compilación construcción de corpus de lengua de signos. Dicho sistema surgió de la necesidad de analizar un corpus de interpretaciones hacia la lengua de signos de Sudáfrica y se diseñó de manera específica para analizar corpus de interpretaciones signadas que usan analizadores programas de concordancias como Wordsmith Tools o Antconc. La transcripción de los signos manuales se aborda mediante ID-glosas lematizadas, lo que permite la comparación semántica entre el mensaje original y su interpretación. El sistema de anotación está diseñado para analizar cinco grupos de factores principales, es decir, los rasgos lingüísticos de la lengua de signos producida por el intérprete, las características asociadas a la calidad de producción de los mensajes cara a cara, las características del tipo de lengua usada por el intérprete (es decir, el *habla*) y los rasgos del proceso de interpretación tanto en términos de la descripción de la diferencias entre los mensajes original y meta y en términos de la identificación de los errores fonológicos, sintácticos y discursivos cometidos por los intérpretes.

Palabras clave: interpretación de la lengua de signos, anotación de corpus, lingüística de corpus, transcripción de la lengua de signos, Estudios de Interpretación, estrategias de interpretación, errores de interpretación, estudios de la lengua de signos, lengua de signos de Sudáfrica.

1. INTRODUCTION

Since Mona Baker first introduced corpus linguistics into Translation Studies (TS), resulting in the now established discipline Corpus Translation Studies (CTS) (Baker 1993: 1995), corpus analysis is regarded as the most objective means to compare source and target texts (Setton 2002: 42). Manchester's TEC annotation scheme currently constitutes the benchmark for annotations in Translation Studies (cf. Cencini & Aston 2002: 47-62; Setton 2002: 29-34) and the TEC corpus has contributed significantly in corroborating T/I-universals (cf. Setton 2011; Zanettin 2012) such as *simplification* (Laviosa 2002; Øverås 1998), *explicitation* (Olohan 2004; cf. Olohan & Baker 2000),

normalization (cf. Kenny 2000; Øverås 1998; Tirkkonen-Condit 2004) and *ST interference* (Baroni & Bernadini 2006; Mauranen 2008). However, there is still a need for improvement in terms of more empirical studies, larger corpora and more rigorous research methodologies, concept definitions and statistical testing (cf. Setton 2011; Zanettin 2012: 27).

Corpus-based/driven research in Interpreting Studies is much less established than in Translation Studies, however, due to the discipline's youth and to difficulties in transcribing and annotating face-to-face (i.e. oral or signed) communication (cf. Alexieva 1997; Bendazzoli & Sandrelli 2005, 2008; Pöchhacker 1995: 17-32, 2007:129; Setton 2002: 29-34; Shlesinger 1998). Hence there are greater gaps in the field, and a greater need for rigorous empirical research and suitable transcription/annotation systems that can offer solutions to the challenges of representing paralinguistic and prosodic features (Alexieva 1997; Pöchhacker 2007: 134; Setton 2011: 51; Shlesinger 2000: 239, 1998: 487), accurately transcribing irregularities (Bendazzoli et al. 2011; Shlesinger 2008: 239); synchronising transcriptions with their video image (Pöchhacker 2007) and standardising conventions (Shlesinger 2008).

Although most are the result of doctoral studies and are not available to other researchers (cf. Bendazzoli & Sandrelli 2005; Setton 2011), the body of *spoken language* interpreting corpora has considerably expanded since Oléron and Nanpon's (1965) foundational study (cf. Chernov 1978, 1994; Lederer 1981; Pöchhacker 1994; Schjoldager 1995; Kalina 1998; Wadensjö 1998, 2002; Shlesinger 2000; Wallmach 2000; Diriker 2004; Setton 2002, 2011; Vuorikoski 2004, Bendazzoli & Sandrelli 2005, 2008, 2011). Research focii in Interpreting Studies include exploring strategies and norms (Pöchhacker 1994; Schjoldager 1995; Setton 2002, 2011; Shlesinger 2000; Wallmach 2000), interpreting quality (Bendazzoli & Sandrelli 2011; Chernov 1994), differences between simultaneous and consecutive modes (Russell 2002, 2005), differences between interpretations and translations (Dragsted & Hansen 2007; Shlesinger 2008), differences between interpretations and original speech (Russo et al. 2006), differences between signed and spoken simultaneous interpretations (Isham 1994, 1995) and categorization of universals (Setton 2011: 45) such as *ST interference* (Dam 1998; Shlesinger 2008), *simplification* (Shlesinger 2008) and *normalization* (Jakobsen et al. 2007; Shlesinger & Malkiel 2005).

In contrast, corpus analysis in *signed language interpreting* is an undeveloped field. Apart from the SASL corpus reported in this paper, the only other

signed language interpreting corpora are those of Isham (1994, 1995), who compared spoken and sign language interpreters' sentence recall, Russell (2002), who compared simultaneous and consecutive interpreting, and Savvalidou (2011), who investigated interpreter awareness of politeness strategies. None of these projects utilizes concordance software; therefore, the SASL study presents the first annotated corpus for signed language interpreting analysis using concordance software. Apart from the relative newness of the discipline, the main reason for the paucity of signed language corpora lies in the even greater difficulty of transcribing and annotating a visual language. This is succinctly expressed by Segouat and Braffort (2009: 65):

> Annotations can be made with glosses or complete translations, but these written data cannot describe in an efficient way typical SL [sign language] properties such as simultaneity, spatial organization, non-manual features, etc. In our opinion, it would thus be difficult to apply the computations used on written comparable corpora or on parallel corpora to comparable or parallel SL [sign language] corpora.

Although the field of corpus-based/driven signed language interpreting studies is limited, a number of annotation and transcription systems have been developed by researchers in the discipline of sign language linguistics (cf. Bungeroth et al. 2008; Crasborn & Hanke 2010; Hoiting & Slobin 2002; Johnston 2014; Koizumi et al. 2002; Leeson & Saeed 2012; McKee & McKee 2009; Neidle et al. 2001; Nonhebel et al. 2004; Özyürek et al. 2009; Paabo et al. 2009; Pichler et al. 2009; Prinetto et al. 2011; Segouat & Braffort 2009:65; Wallin et al. 2010; Wallin 2012). These corpora are based on original, not interpreted, discourse conducted by either native or proficient users of the signed languages involved and the corpora have been constructed for the purpose of providing a full linguistic description of the signed languages studied. Most of these corpora are currently completing the construction phase and are gradually being made available as online resources to researchers (cf. Johnston 2014; McKee & McKee 2009; Leeson & Saeed 2012). These linguistics-based corpora use a multi-tiered program called ELAN or EUDICO (European Distributed Corpus Linguistic Annotator) (cf. Johnston 2010: 110) to synchronize the video material directly with annotations, precluding the need for transcription (cf. Johnston 2014; Koizumi et al. 2002; Leeson & Saeed 2012; Paabo et al. 2009; Pichler et al. 2009).

However, in Interpreting Studies, the researcher is primarily interested in semantic comparisons between a *source text* (ST) (which can be written, oral or signed) with transcripts of its interpretation, i.e. the *target text* (TT) (which is either oral or signed), especially in terms of investigating semantic transfer across the two languages, linguistic features of the TT (e.g. collocations, grammatical, syntactic and discourse constructions, intonation, facial expressions, mouthings, etc.), issues related to the *production* of face-to-face communication (e.g. hesitations, slurs, filled/empty pauses, clarity of articulation, visual/oral noise, etc.) and features related to the *act of interpreting* (e.g. substitutions, omissions, additions, strategies, interpreting errors, etc.). In ELAN, the ST needs to be loaded as an annotation tier rather than as a separate file, which makes it difficult to compare multiple STs with their TTs. Moreover, because of the inclusion of the video material, the ELAN files are considerably bulkier than the simple text files used in word-based corpus packages and the program requires some training before a researcher is able to use it efficiently.

This paper presents a transcription and annotation system designed to overcome these challenges and allow researchers to investigate signed interpretations as monolingual corpora using readily-available word-based concordance packages such as Wordsmith Tools (Scott 2011) or AntConc (Anthony 2011). If the TT metalanguage is *not* the same as the ST, Paraconc (Barlow 2003) can also be used. Moreover, suggestions are also offered in order to adapt the system to annotate spoken-language interpretations, thereby offering a significant contribution generally to researchers in Interpreting Studies who use corpus-based/driven approaches.

2. CONTEXT AND LIMITATIONS OF THE RESEARCH

The transcription and annotation system presented in this article derives from the construction of a corpus of interpreted news broadcasts from English into South African Sign Language (SASL) (Wehrmeyer 2013). This corpus was constructed in order to investigate why members of the Deaf target audience did not understand the interpretations and thus had to consider linguistic features as well as production and interpreting features. However, because timing was not considered vital to the research question, only rudimentary annotations were developed (for lag time, pauses, punctuation and chunking) and the system therefore requires further development in terms of recording simultaneity etc. between ST and TT.

To provide a basis for this pioneer corpus, annotation systems used for the corpora constructed for linguistic research into signed languages reported above were investigated and adapted according to a set of principles. Firstly, conciseness was important, since for effective comparison with the STs, the SASL corpus should not be cumbersome due to long annotations. Thus simple alphanumeric codes were prioritized over descriptive terms. Secondly, since concordance software such as *Wordsmith Tools* and *Antconc* use text files, the annotations had to consist of symbols that could be recognized as plain text. This precluded the use of graphic symbols (e.g. arrows) in favour of alphanumeric codes. Thirdly, the annotations had to be unambiguous and mutually exclusive in order to facilitate search operations. Fourthly, in line with current practice in sign language linguistics, incorporation of prior theoretical analysis was minimized by preferring linguistic annotations based on observable, physical characteristics where these existed.

3. REPRESENTATION OF SIGNS

Signs can be represented by notational codes based on their phonological characteristics (cf. Sutton 2012; Hanke 2004), or by glosses based on their semantic meaning. While phonological representations are useful to linguists, researchers in Interpreting Studies want to compare what was said/written/signed in the ST and how it was interpreted (said/signed), i.e. semantic comparison is prioritized (cf. Pöchhacker 1994; Schjoldager 1995; Setton 2002, 2011; Shlesinger 2000; Wallmach 2000); hence a gloss system forms the basis of the transcription system.

3.1. THE GLOSS SYSTEM

Two types of glosses are used in the literature to represent signs, namely descriptive and ID glosses (Johnston 2011; Leeson & Saeed 2012; McKee & McKee 2009). Descriptive glosses represent signs by their contextual meanings and attempt to align sign language as naturally as possible to spoken language. In contrast, an ID gloss is defined as the primary denotative lemmatized meaning of a sign (Johnston 2010: 114). This allows signs to be consistently represented by the same gloss regardless of their contextual meanings. According to Johnston (2014), ID glosses are more objective than descriptive glosses since contextual meaning is sometimes a matter of subjective interpretation. Furthermore, since they are lemmatized, they allow greater consistency in compiling

corpus wordlists and carrying out search functions. ID glosses are used in the Australian, New Zealand, Irish, British, Dutch and Swedish linguistics corpora (Crasborn 2009; Johnston 2014; Leeson & Saeed 2012; McKee & Kennedy 2006); hence, in line with what is increasingly becoming the accepted practice in signed language corpus linguistics, ID glosses were also used in the SASL corpus. This represented a break from the convention of using descriptive glosses in signed language interpreting studies (cf. Lombard 2006; Savvalidou 2011).

The use of ID glosses implies that a reference lexicon of signs exists (Johnston 2010: 116). For SASL, the *Dictionary of South African Signs (*Penn et al. 1992) undertaken in conjunction with the Human Science Research Council (HSRC) was used since it constitutes the most authoritative reference work available. The first sign of each dictionary entry was taken as the primary (i.e. unmarked) variant, with other variants marked as dialect (discussed in Section 7 below).

Usually glosses are transcribed using capital letters (cf. Lombard 2006; Savvalidou 2011). However, in order to free capitals to be used for annotation symbols and fingerspelling, in the SASL corpus glosses are transcribed using small letters. An underscore separates the gloss from embedded annotations, e.g. in "rain_IM2" the gloss "rain" represents the sign and IM2 are annotations. By suffixing all glosses with underscores regardless whether they contained embedded annotations or not, a token count can be obtained for a transcription simply by counting all the underscores.

Corpora based on ID glosses are by definition lemmatized, so where it is important to distinguish grammatical forms (e.g. where the manual signs differ), the gloss is suffixed with an appropriate marker (e.g. "children" = "child-**pl_**"). Where it is important to annotate contextual meanings (e.g. in the case of polysemy or synonyms), these are transcribed as *contextual meaning=ID gloss*, e.g. "goal=aim_" indicates that the interpreter used the sign for "aim" as an equivalent for goal.

3.2. SINGLE AND COMPOUND SIGNS

Ideally, a single sign should be represented by a single gloss to distinguish them from compound signs (i.e. a unit of meaning consisting of at least two signs), which, by convention, are transcribed as the component glosses separated by hyphens (Johnston 2014: 13). The mapping of single signs onto single glosses also allows accurate corpus token counts. The assigning of single glosses to single signs belonging to the established lexicon (e.g. "rain_") is

unproblematic. However, signed languages also include a productive lexicon which encompasses signs that describe actions (termed descriptive verbs) or reflect the size, shape or spatial relationship of objects (cf. Johnston 2014). Single signs belonging to the productive lexicon can reflect the same meaning as a whole sentence in spoken language and therefore are more adequately translated by a descriptive phrase, e.g. "give money to". These were also assigned a single gloss condensed from the descriptive phrase, e.g. "givemoney", with separators only used if this gloss was ambiguous, e.g. "not-here" (cf. "nothere"). Following the convention used by Johnston (2014), these are prefixed by DV (descriptive verb), e.g. "DVgivemoney_".

Consecutive or simultaneous use of individual signs in a compound were marked by using hyphens for consecutive compounds (e.g. "longbeardsack_" = Santa Claus) and additions for simultaneous compounds (e.g. "DVwalkO+walk@_")[1]. This represents a finer distinction than in the Australian corpus (Johnston 2014) where a hyphen is used for both.

3.3. PRONOUNS AND DEIXIS

Transcription of pronouns is problematic in signed languages since their meanings are determined by context (cf. Johnston 2014, McKee & Kennedy 2006). They are conventionally represented as a coded grammatical category, e.g. PRO1SG = "I" (Johnston 2011), POS-2 = "your" (McKee & Kennedy 2006), or as INDEX (Leeson & Saeed 2012). Deixis is similarly represented by a range of codes, e.g. PT (Koizumi et al. 2002), PT: PRO-3G (Johnston 2014), IX-3 (McKee & Kennedy 2006), or as INDEX (Leeson & Saeed 2012). To achieve a consistent transcription policy, the SASL corpus designated pronouns and deixis according to handshapes with annotations for location and direction, similar to the Irish corpus (Leeson & Saeed 2012). Thus, as illustrated in Figure 1 below, the deixis sign (corresponding to various contextual meanings such as "there", "you", "s/he", "it", "this", "that", etc.) is glossed consistently as "index" (cf. Leeson & Saeed 2012), the polite form (fist turned outwards and corresponding to contextual meanings such as "you", "yours", "s/he", "his", "hers", etc.) as "you" and the first person (articulated with the fist turned inwards and corresponding to various contextual meanings such as "I", "my", "self") as "me" (to avoid using the single capital letter "I") , regardless of their contextual grammatical meanings.

[1] The example includes other annotations described below.

Figure 1: Transcription of pronouns

"index_" "you_" "me_"

3.4. QUANTITIES AND FINGERSPELLING

Quantities in corpora are generally transcribed fully (e.g. "twenty-two" instead of "22") (cf. Gollan et al. 2005). However, the sign language researcher may wish to exclude numbers from wordlists, which is very difficult to achieve if they are represented by glosses. Hence, provided the concordance program's token definition settings can be set to include number token classes as is the case with *Antconc*, quantities are transcribed as numerals (e.g. "217") in both ST and TT. This facilitates concordance operations since numerals are listed first in sorted lists and can therefore be easily eliminated from frequency counts. However, quantities expressed by a particular sign and not as a numeral are glossed as signs, e.g. "thousand_".

In signed languages, proper nouns are usually fingerspelled using a manual alphabet derived for that purpose. For example, SASL uses the one-handed American Sign Language alphabet. In the literature, fingerspelling is usually transcribed with an identifying code and sometimes a representation of the actual letters spelt, e.g. fs-OPOSSUM (McKee & Kennedy 2006), FS:WORD(WRD) (Johnston 2014: 37) Transcribing fingerspelling presents three main problems. Firstly, interpreters tend to abbreviate fingerspelled items, making it difficult to compare alternative spellings of a name, e.g. in the SASL corpus the name "Agliotti" is spelled as AGL, AGLIOTIE, AGLIOT, AGIOT and ALITI. Secondly, mouthing (see Section 5.2 below) can vary during fingerspelling. Thirdly, conventions outlined in the literature which isolate each letter with delimiters, e.g. "p.a.r.k.e.r." (cf. Leeson & Saeed 2012), prove diffi-

cult to manipulate with word-based concordance functions. To overcome these problems, a new convention was devised in which the letters actually spelled are rendered in capitals and omitted letters in lower case, e.g. in "PARKeR_" the interpreter fingerspelled the letters P, A, R, K, R of the surname "Parker". This allows spelling variants to be listed together in wordlists.

Should the researcher wish to study numerals or fingerspelling as a research *category*, an identifying annotation can be prefixed to facilitate search operations, e.g. for quantities: "Y217", "Ythousand". In the SASL corpus, fingerspelling is prefixed with category code (Z) together with axial categories for single letters (0), surnames (1), other names of people (2), names of organizations (3), names of places or directions (4), names of things or measurements (5) and fingerspelled conjunctions such as IF and SO (6), e.g. "Z1PARKeR", "Z3ANC", "Z4South-West", "Z6SO".

4. PUNCTUATION

In order to produce consistency between ST and TT, normal punctuation marks were used in the SASL corpus. Periods are used to mark instances of non-signing signaled by interpreters folding their hands (at or below waist level) and commas to mark pauses where the hands are still in signing space. Longer pauses are marked as "…". An alternative system in which each paused second is represented by "/" (cf. Allwood et al 2005; Leeson & Saeed 2012; Paabo et al. 2009) may be used if timing measurements constitute part of the research question; however, this notation does not specify whether or not the interpreter is engaged in signing space. Question marks are not used, since interrogatives are marked in signed languages by facial expression. Hence the identification of a piece of discourse as a question is a derived interpretation. Exclamation marks are used to mark emphasis, either after the gloss if the sign is exaggerated (e.g. "big!") or after the relevant annotation if the non-manual feature is exaggerated (e.g. "big_E22/5(pah)!").

Because they are used for fingerspelling and annotation codes, capital letters are not used to mark the start of new sentences or for proper nouns, unless the sign for the proper noun is articulated using a fingerspelling handshape, e.g. "Durban_" is signed using a "D" handshape and is thus transcribed with an initial capital letter, whereas "pretoria_" is not articulated with a "P" handshape and is thus transcribed only in small letters.

5. ANNOTATIONS FOR LINGUISTIC FEATURES

Phonological elements of a signed language include handshape, palm and finger orientations, non-manual features, location and movement (cf. Johnston & Schembri 2007; Koizumi et al. 2002). Non-manual features include facial expression (eyebrow and eye movements, eye gaze, blinking rates, mouth gestures, etc.), head/body movements and mouthing (cf. Neidle et al. 2000; Stone 2009). These parameters act as prosodic, discourse and syntactic markers, and even as morphemes. These elements are often intrinsic to a particular sign and thus of little interest to the interpreting researcher; thus only marked use may need to be annotated.

5.1. HANDSHAPES

Since a gloss-based corpus primarily focusses on the meaning of a sign and not on its phonological features, handshapes are usually only annotated when used as *classifiers*, i.e. when the hand is held in a specific shape as a discourse device. In linguistics-oriented corpora, classifiers are transcribed either as a description, e.g. "CL(pile of books)" (Leeson & Saeed 2012) or as a code derived from the handshape, e.g. "DSL(2-HORI)" (Johnston 2014), "pm'TL" (Hoiting & Slobin 2002). A prefixed annotation code distinguishes them from true signs and can also denote the class of classifier used, e.g. whether it describes an object's movement or shape, acts as a reference to a person or object or reminds the audience of the discourse topic (this type of classifier is termed a *buoy*) (cf. Johnston 2014). Buoys can be further subdivided into different types, e.g. the classification "LBUOY" = list buoy (Johnston 2014) includes discourse devices "firstly", "secondly", etc.

Because meaning and function are prioritized in the SASL interpreting corpus, the Irish convention (Leeson & Saeed 2012) was adapted to the convention established above for the productive lexicon of writing the description as a single word. Thus, for example, "CLcar_" signifies that the hand is held in the car classifier shape instead of the signer producing the established sign. Since buoys are held for over a period of discourse, they are annotated by inserting <buoy(CL-X)/> at the start and </buoy(CL-X)> at the end of the relevant segments, e.g. "<buoy(CLrein-

Figure 2: Buoy annotation

"<buoy(CLsecond)/> why_ […] </buoy(CLsecond)>

deer)/>…". Figure 2 illustrates a list buoy ("secondly") held over a period of discourse.

A second feature of handshape of interest to researchers studying signed languages is marked use of the hands in articulating signs. Some signs are performed by a single hand rather than by both hands, and certain discourse features involve both hands with different functions. The hand that performs the single signs is referred to as the dominant hand and usually corresponds to natural tendencies, e.g. a right-handed person would use his right hand as his dominant hand in signing (cf. Leeson & Saeed 2012). In ELAN-based corpora, the functions and handshapes of each hand are often represented on separate tiers (cf. Johnston 2014). In the SASL interpreting corpus, marked use of the hands is annotated by suffixing the gloss with "RH" (dominant) or "LH" (non-dominant), e.g. "personLH_" indicates that the sign for "person" was executed by the non-dominant hand. (It is evident that the annotation reflects the fact that most people are right-handed.) This notation is also used if part of a two-handed sign is executed, e.g. "carLH_" indicates that only the left-hand part of the two-handed sign for "car" was executed. Similarly, one-handed signs executed with both hands are suffixed with "2H", e.g. "fly2H_" (Figure 3a). Hand use can combine with classifier tags, e.g. "CLcarLH_" (Figure 3b) indicates that the non-dominant hand is held in the car classifier shape. Two-handed signs performed according to normal symmetry rules and one-handed signs performed by the dominant hand are by default unmarked. In simultaneous compound signs, the dominant hand sign is followed by the non-dominant hand sign, e.g. in "walkO+walk@_" (Figure 3c), the dominant hand performs a circular walking motion (O) while the non-dominant hand performs a straight walk (@).

Figure 3: Annotations for hand use

a) "fly2H_" = numerous flights

b) "CLcarLH_"(introducing a classifier)

c) "walkO+walk@_" = some walking down the road, others wandering aimlessly

5.2. MOUTHING

Mouthing is the phenomenon when a signer says words or parts of words audibly or silently while signing (Mohr 2011; cf. Leeson & Saeed 2012: 81; Stone 2010: 2). Mouthings are borrowings from spoken language. The practice has become integral to certain signs in some signed languages, and is also used by interpreters to facilitate understanding of signs (e.g. jargon) that the interpreter does not consider to be familiar to her target audience (cf. Crasborn 2009). Although discouraged in SASL (cf. Akach 1997), mouthing is prevalent in the SASL interpreting corpus. Mouthing should not be confused with *mouth gestures* (discussed below) that are established phonemes in sign languages and not borrowed from spoken language (cf. Leeson & Saeed 2012; Sutton-Spence 2007; Woll 2001).

According to Mohr (2011), mouthing may be identical, partial or different to the concept represented by the sign. In ELAN-based corpora, mouthing is annotated by transcribing the whole word represented on a separate tier with parenthesis to denote clipped portions, e.g. DELIB(ERATE) (Johnston 2014; Leeson & Saeed 2012), m(be(cause)) (Pichler et al. 2010).

In the SASL corpus, mouthing is annotated as "V*tql*" with category codes for timing (t), quality (q) and language use (l) after other embedded annotations. Timing may be simultaneous (0), subsequent (1) or prior to the sign (2). Following Mohr (2011), word quality is categorized as full (0), partial (1), different (2), unclear (3), differing in grammatical category (4) or related semantically (5). Language use is indicated as same (0) or different (1) to the ST. Thus, "win_V001(wen)" indicates that the interpreter simultaneously mouthed a full word but also code-switched into Afrikaans. It is evident that these categories can be further extrapolated if necessary.

For the sake of conciseness, full mouthing identical to the sign gloss (i.e. q = 0) is not transcribed, whereas partial and different mouthings (i.e. q = 1 to 5) are transcribed in parenthesis without spaces, e.g. "child_V020(youth)" vs "child_V000". Since the primary meaning is already carried by the sign gloss, only the part actually mouthed is transcribed in order to accurately reflect praxis, e.g. "parliament_v010(parl)". Words mouthed when the interpreter is not signing are annotated as <V9>, e.g. "<V9first>", thereby facilitating exclusion from token lists. Depending on the research question, it may also be useful to annotate for non-mouthing (V8), e.g. "child_V8" means that the sign is produced without mouthing.

Mouthing variations in fingerspelling and numerals are depicted using hyphens. For example, "SELEBI_v0" means that the interpreter fingerspells all letters and simultaneously mouths "selebi", whereas "S-E-L-EBI_v0" means that the interpreter mouths the first three letters S, E, L separately, then the remaining letters as a partial word, "ebi". Similarly, "217_V000" means that the interpreter mouths "two hundred and seventeen" (fully and simultaneously in the source language) while signing, whereas "2-1-7_V000" means that the interpreter mouths the individual numbers "two-one-seven".

5.3. FACIAL EXPRESSION

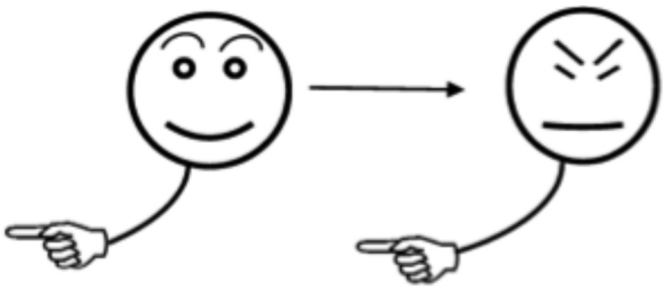
In signed languages, facial expressions encompass the eyebrows, eyes, cheeks and mouth and are important for constructing grammatical, syntactical and discourse meaning (cf. Pfau & Quer 2007; Sandler & Lillo-Martin 2006). Annotations found in the literature consist of descriptive words (e.g. "pout" (Johnston 2014)), initial-letter codes (e.g. /eb/ = eye blink, br = brow raised (Leeson & Saeed 2012, cf. Koizumi et al. 2002)) or physical detail (e.g. /BILABIAL/ (Nonhebel et al. 2004)), with each facet of facial expression described in a separate tier in ELAN-based corpora.

For the SASL corpus, the Irish (Leeson & Saeed 2012) and Japanese (Koizumi et al. 2002) annotations provided a useful framework for the types of facial expressions that required annotation. Applying the conciseness rule, their categories were assigned alphanumeric codes into the following annotation system: "E*bed/m*" for eyebrows (b), eyes (e), eye gaze (d) and mouth (m)[2]. A delimiter (/) is added in front of mouth codes to facilitate concordance searches of mouth gestures as a category. Eyebrows (b) are coded as relaxed (1), raised (2) or frowned (3). Eyes (e) can open normally (1), widen (2), squint (3), roll (4) or shut (5). Eye gaze direction (d) is an optional category and is annotated using the directional codes described below in Section 5.4. The mouth (/m) can be relaxed (0), smile (1), pout (2), pull down (3), grimace (4), mime spoken phonemes (e.g. "pah", "wh", "mm" etc.) (5), open wide (6), snarl baring the teeth (7), pull tight (8) and puff cheeks (9). Mouth gestures related to spoken language sounds (category 5) are transcribed in parenthesis, e.g. "big_E22/5(pah)". Hyphens depict changes in facial expression during the articulation of a sign, e.g. "index@4_E22/1-33/8" (Figure 4) indicates that the interpreter changes her initial expression (E22/1 = raised eyebrows, wide-open

[2] Cheeks are included under mouth actions, but a separate category can be constructed (cf. Crasborn 2009).

eyes, smile) to a frown with squinted eyes and pressed lips (E33/8) while performing the deixis sign.

Figure 4. Change of facial expression during a sign: index@4_E22/1-33/8



The codes thereby concisely record a large amount of detail without affecting the corpus token count.

5.4. LOCATION AND MOVEMENT

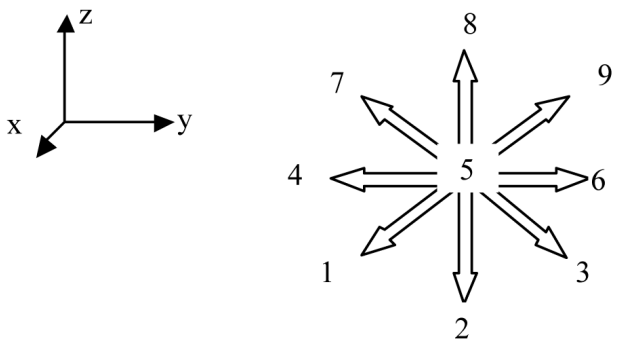
In SASL, as in other sign languages, location and movement are used to describe relative locations, manner and sequential ordering of events (Akach 1997: 18; Prinsloo 2003). However, as they are phonological parameters of any sign, only marked use needs to be annotated (cf. Johnston 2014; Leeson & Saeed 2012). In the literature, most linguistic corpora use alphanumeric symbols to express relative directions, e.g. 1GIVE2 (Leeson & Saeed 2012), S = sinuous downward movement (Paabo et al. 2009), or areas in the *signing space* (i.e. the three-dimensional sphere around the signer), e.g. numerical codes 1-10 (Paabo et al. 2009), c=centre, f=far centre (Leeson & Saeed 2012), l=left, 45º, u=up etc. (Nonhebel et al. 2004).

For the SASL corpus, a combination of spatial co-ordinates with the signer at centre and descriptive codes was derived. Locations are annotated by direction codes only, whereas movements are annotated by "M" together with a destination direction code. Horizontal (*x*) motion or direction is annotated as (M)50 (i.e. near centre or motion inwards towards the speaker) or (M)51 (i.e. far center or motion outwards away from the speaker)[3].

[3] If required, horizontal codes can be expanded to include signer plane (0), near space (1) and far space (2), e.g. to describe gazing into the distance.

The codes used for directions in the vertical (*yz*) plane are illustrated in Figure 5 below:

Figure 5. Direction codes for vertical (*yz*) plane



In signed languages, location and motion are described from the signer's perspective, e.g. 4 = the signers right. The codes, however, correspond to orientations on a computer keyboard number pad, enabling the transcriber to speedily select the correct direction without having to work out the signer's perspective. These directional codes are also used to annotate eye gaze and head/body movement, e.g. in "bad_E33**6**/4", the 6 digit indicates that the signer gazes to her left.

It is important to note that signers very seldom sign, look or face exactly to the left, right, up or down (i.e. in the exact plane of the signer). Especially in simultaneous interpreting where there are enormous time constraints, interpreters do not have the luxury of excessive body and hand movement. Hence the *yz* directions are used as approximate directions in a plane slightly in front of the signer, i.e. "to the left", "upwards", etc., rather than as strictly left, up etc. Strictly speaking, these directions can be coded as M61, M81 etc., to indicate a distance in front of the signer and allow for directions exactly in the signer's plane to be coded as M60, M80[4] etc.

[4] Note that because the *yz* plane is usually the most productive and therefore the most interesting in signing, the motion is given as M(*yz*)*x*, e.g. M71, or simply as M(*yz*), e.g. M7, with the x-co-ordinate unmarked and implied. However, mathematically-defined spatial or even polar co-ordinates can also be used if more accurate portrayal of location or movement is required, e.g. Mxyz= M(1;-1;1) in spatial co-ordinates is equivalent to M7 in the SASL code. Usually, however, this level of accuracy is not required in an interpreting corpus.

Complex forms of motion are assigned the following codes:

- ➪ Motion in an arc (MC), e.g. “all_MC”;
- ➪ Circular motion (MO), e.g. “trade_MO”;
- ➪ Wrist rotation (MG), e.g. “pretoria_MG”;
- ➪ Random motion (MM), e.g. “placepl_MM”;
- ➪ Repetition of a sign (MR), e.g. “man_MR”;
- ➪ Up-down alternating motion (MW), e.g. “maybe_MW”;
- ➪ Up-down together motion (MV), e.g. “weightlifting_MV”;
- ➪ Left-right alternating motion (MZ), e.g. “compete_MZ”;
- ➪ Left-right expansion motion from centre (MX), e.g. “expand_MX”;
- ➪ Alternating motion away from speaker (MK), e.g. “fly2H_MK”;
- ➪ No motion (M0), e.g. “trade_M0”.

Sign languages contain many direction verbs which require movement of the sign or classifier towards a particular location, e.g. “look-at-something”, “give-to-someone”. Since this directionality is usually linked to discourse referencing techniques (discussed below), it is expressed as @ plus a direction code, e.g. “look@6_” indicates that the signer moves the sign for “look-at” towards an invisible object set up in his signing space to his left. For conciseness, direction towards a location directly ahead of the speaker (M51) is used as unmarked default, i.e. “look@_” is equivalent to “look@51_”.

5.5. HEAD AND BODY MOVEMENTS

Head movements and to a lesser extent body movements are important prosodic, grammatical and syntactic markers (cf. Sutton-Spence & Woll 2006). In signed language linguistics-based corpora, they are annotated as coded physical gestures, e.g. htb = head tilt back (Leeson & Saeed 2012), grammatical functions, e.g. DIFFICUT-neg (head shake indicating negation) (McKee & Kennedy 2006), descriptions, e.g. “nod” (Johnston 2014; Koizumi 2002) or as a separate category with a description, e.g. NMS-nod (McKee & Kennedy 2006). There appears to be variation in the annotations within corpora, e.g. the New Zealand corpus (McKee & Kennedy 2006) uses both grammatical categories and coded descriptions to annotate head movement.

For the SASL corpus, a single system based on physical gesture was selected using the following annotation codes in angle brackets together with a direction code: h = head, b = body, c = cock (rotate) to one side but face still forward, n = nod, s = shake from side to side, sh = shrug, sw = sway. Although

derived independently, some of these codes are similar to those used in the SOI corpus (cf. Leeson & Saeed 2012). Centre position (i.e. facing directly forward) is default and unmarked. Directionality is represented by the number codes in the previous section and indicate the direction of the face (for head movements) or the chest (for body movements). As noted above, the *yz* locations are slightly in front of the signer (i.e. at a convenient neck angle) rather than in the same plane, e.g. <h8> represents the head tilted upwards but not awkwardly looking at the ceiling. Thus <h4> <b6> means that the face is turned to the signer's right whereas the body is turned to the signer's left. For <h9>, <h6> and <h3>, the face is turned to the signer's left, but the first represents a head tilt, the second a head turn, and the third, the signer looking downwards towards the left. However, <hc6> means that the head is cocked to the signer's left, whereas the eyes still look straight ahead. Likewise, <hn> indicates a head nod (as in affirmation), <hs> a head shake (as in negation), <bsh> a shrug of the body and <bsw> swaying movement of the body.

If the head or body gesture is maintained over a single sign, it is designated with a hyphen, e.g. "<h8-> british_". If it is maintained over a number of signs, it is designated as <*x*/> … </*x*>, e.g. "<h8/> british_ tour_ person_ </h8>" indicates that the head tilt is held over the whole phrase.

6. ANNOTATIONS FOR PRODUCTION QUALITY

In Interpreting Studies, quality of production of the target message is a key factor affecting comprehension of the message as well as user satisfaction of the interpreting services (cf. Kurz 1993; Pöchhacker 2007). Moreover, according to Gile's (1995) capacity model, it is also indicative of the cognitive capacity available to the interpreter for message output (cf. Shlesinger 2000). However, few annotations are reported in the literature for production quality. Unclear signs are annotated as [=?], e.g. SICK[=?] (Pichler et al. 2010) if the meaning can be inferred, or as YYY or XXX (Pichler et al. 2010) or UNDECIPHERABLE (Johnston 2014) if not. Sounds such as coughs and laughs are annotated as &=, e.g. &=laughs (Pichler et al. 2010).

In the SASL corpus, quality of articulation, signing speed, lag time and chunking were annotated as factors affecting the comprehension of the message. Quality of articulation was investigated by annotating for poor visibility against background (U0), careless or incomplete articulation (X0) and incorrect phonology, i.e. incorrect hand classifier, finger/palm orientation, movement,

etc. (X1), e.g. "index_U0". The latter two categories also constitute errors and are therefore included in Section 9 below. Although not done for the SASL corpus, other phonological errors (e.g. incorrect facial expression) may also be annotated. Signing speed is annotated as exceptionally fast (F1), very slow (F2) or held (F3), e.g. "index_F1". Since these features are not inherent to a particular sign, embedded annotations are used.

Annotations of lag time, also known as ear-voice span (cf. Gile 1995; Kalina 1998), are inserted using angle brackets, e.g. <lag=-1>. A positive value indicates that the interpreter began after the source speaker, whereas a negative value indicates that the interpreter began before the source speaker (indicating anticipation or prior access to the ST). Since lag time was not pertinent to the SASL research question, the values merely reflect an approximation to the nearest second. However, more accurate values (e.g. using milliseconds) can also be used, e.g. <lag=0.010> = 10 ms.

7. ANNOTATIONS FOR LANGUAGE USE

The researcher may wish to investigate the nature of the language used by the interpreter. This can be done by annotating categories of interest such as parts-of-speech, discourse topics, *iconicity* (i.e. a sign that displays physical resemblance to the object denoted), use of the productive lexicon, information density, dialectal variants, *referencing* (i.e. the placing of objects and people at specific locations in the signing space) and role-play (i.e. mime). Although these annotations facilitate investigation of these categories, it must be noted that they introduce prior theoretical analysis into a corpus.

Parts-of-speech annotations are common in the literature, not only in signed-language corpora (e.g. Johnston 2014; cf. McEnery & Wilson 2001; Setton 2002). In ELAN-based corpora, annotations for parts-of-speech and/or grammatical categories are usually done with abbreviated codes (e.g. "Prep" = preposition) in separate annotation tiers (cf. Johnston 2014; McKee & McKee 2011). The Australian corpus (Johnston 2014) also contains complex category codes for gesture (e.g. "G:PHOOEY"), role play (e.g. "CA:NARRATOR"), syntactic categories (e.g. "V1" = the first verb in a serial construction) and clausal analysis (e.g. "CLU TJ1aCLU#01"). Annotations for sociological variation are also used in some corpora, e.g. "V1", "V2", etc. (McKee & McKee 2011).

In the SASL corpus, categories that are integral to the sign (and therefore context-independent) are annotated using prefixes, e.g. V=verb, Q=interrogative question, U=conjunction, I=iconic sign, J=information-dense signs (e.g. "Vcompare_", "Qwhy_", "Uand_", "Irain_", "Jpretoria_"). Annotated items belonging to the productive lexicon include DV=descriptive verb (e.g. "DVhideface_") and G=gesture (e.g. "Gnotbother_") (cf. Johnston 2014; Leeson & Saeed 2012). If categories are context-dependent (e.g. if a sign functions both as noun and as verb), embedded annotations are preferred in order to retrieve words systematically from wordlists, i.e. "trade_V" instead of "Vtrade_".

However, dialectal variation is annotated as "D1" using suffixes so that glosses can be grouped alphabetically in wordlists for comparison, e.g. "woman_" vs "womanD1_". If more than one variant appears, these are annotated as "D2", "D3", etc. A second category code is added if the origin of the variant is known, e.g. "D1B" = British Sign Language, "D1$" = American Sign Language, etc. In the SASL corpus, variants belonging to the Afrikaans dialect ("D1A") are also annotated since this dialect was found by Vermeerberger et al. (2011) to be markedly different to other SASL forms.

Topics are marked by non-manual features in signed languages, but it may be of use to the researcher to identify discourse topics as a collective category, e.g. to study facial expressions used. Since topics are context-dependent and thus not an inherent feature of a sign, this is done using an embedded tag, "_T", e.g. "<h8> mandela_TE22". Topics that extend over a number of signs are tagged by bracketed annotations before and after the discourse segment, e.g. <T/> ….. </T>.

In signed languages, *referencing* (i.e. the assigning of people or objects (*referents*) to a location in the signing space so that they can be discussed by simply pointing to the relevant location) is an important discourse device (cf. Braffort et al. 2010: 453; Neidle et al. 2000: 36). In ELAN-based sign language corpora, referencing is annotated as a subcategory of deixis, e.g. PT:LOC/PRO = pointing (PT) at a referent (PRO) situated at a specific location (LOC) (Johnston 2014). In the SASL corpus, references were annotated with "@"[5] plus a location code, e.g. "index@6_" means that the interpreter points to a reference which he has set up to his left.

[5] Researchers need to ensure that annotation codes are not assigned to functions in the concordance software used to analyze the corpus. Because @ is a wildcard in *Antconc*, the relevant wildcard was changed to &.

Since signed languages are visual languages, spatial relationships between two objects can be depicted with great accuracy and are therefore also used to reference objects or persons. The dominant hand signs the object of interest (the *figure*), whereas the non-dominant hand signs the location object (the *ground*) (cf. Özyürek et al. 2009). Usually (but not always) the dominant hand articulates a sign and the non-dominant hand a classifier. In ELAN-based corpora, the function of the different hands can be described in separate tiers to indicate the spatial relationship. In the SASL corpus, the signs are transcribed as *figure@ground*, e.g. "DVsit@CLcar_" depicts a person sitting (depicted by the dominant hand) in a car (depicted by the non-dominant hand classifier). If the two hands perform the same action on each other, this is glossed as "@eo" (= each other), e.g. "DVshoot@eo_" (= two people/groups shooting at each other). Referents assigned to the non-dominant hand are similarly annotated, e.g. "index@CLcar_" (= this car) means that the dominant hand points to the non-dominant hand which assumes a car classifier handshape. If the hand functions are swapped, the markedness is annotated, e.g. "indexLH@CLcar_" means that the signer points with his non-dominant hand. The different uses of the "@" annotation are illustrated in Figure 6 below:

Figure 6. Reference annotation "@"

a) "index@6_" = that person/object | b) "index@CLcar_" = this car | c) "DVsit@CLcar_" = person sitting in a car | d) "DVtalk@eo_" = talk among each other

8. ANNOTATIONS FOR INTERPRETING FEATURES

An important reason for corpus-based/driven research in Interpreting Studies is the investigation of interpreter strategies and norms (cf. Schjoldager 1995; Setton 2002, 2011; Shlesinger 2000; Wallmach 2000). In the SASL corpus, interpreting choices were identified using Toury's (1980, 1995) categories of *shifts,* i.e. omissions, skewed substitutions and additions. Embedded annotations were used for shifts at word level, whereas angle brackets were used for

shifts above word level. To the best of the researcher's knowledge, the SASL corpus constitutes the first and only example of a corpus in Translation or Interpreting Studies that annotates interpreting choices.

8.1. OMISSIONS

By definition, omissions constitute ST material that is not represented in the TT (cf. Cokely 1992). Although initially perceived as miscues (cf. Barik 1994; Galli 1990), omissions are increasingly perceived as interpreting strategies (Garzone 2002; Jones 1998; Kurz 1993; Moser 1996; Moser-Mercer 1996; Napier & Barker 2004; Pym 2008; Shlesinger 2000; Viaggo 2002; Visson 2005).

Since there are no associated TT signs, omissions are transcribed as <omit-R*x*> where R is a category code and *x* the omitted material, e.g. in "<omit-Vwasasked>", R = category V (predicate verb) and x = "was asked". To facilitate analysis, omitted ST material is transcribed in lower case (to distinguish annotation codes) as single word forms separated by hyphens if necessary. In the case of sentence, clause or list omissions, only the gist is transcribed. The following categories are used in the SASL corpus: predicate verb i.e. verb phrase head (V); subject i.e. sentence noun phrase (S), predicate or indirect objects upon which meaning depends (O), list items (L), information-dense adjectival or adverbial modifiers (Q), topics that are not also subjects (T), conjunctions (U) and propositions or propositional clauses (P). Two further categories were added, namely <omit-blocked> when the interpreter was absent from the screen (e.g. sports listings) and <omit-namereporter> when interpreters omitted the names of reporters. Other omissions that were not regarded as important to the research question (e.g. previously stated topics, repetitions, fillers etc.) were not assigned category codes (e.g. <omit-tomurderinvestigation>). It is evident that further omission categories can be devised, depending on the research interests.

8.2. SUBSTITUTIONS

Skewed substitutions, i.e. where a target language element exists but is not equivalent to the source text element (cf. Cokely 1992), are also increasingly regarded to be the result of interpreting strategies rather than errors (cf. Bartłomiejczyk 2006; Gile 1995; Kalina 1998; Shlesinger 2000; Wallmach 2000). These were annotated as "_S" or "<S/> … </S>" at word and above word level respectively, together with the following sub-category codes:

- S0: interpretation using a synonym, a related word on the same semantic level or a representation from the productive lexicon (cf. Al-Salman & Al-Khanji 2002; Bartłomiejczyk 2006; Camayd-Freixas 2011);
- S1: paraphrase, i.e. reformulation (cf. Bartłomiejczyk 2006; Napier 1998; Leeson 2005);
- S2: simplification or interpretation by a more general meaning (corresponding to the interpreter strategy of *chunking up*) (cf. Gile 1995; Russo et al. 2006; Sandrelli & Bendazzoli 2005);
- S3: explicitation or interpretation with a more specific meaning (corresponding to the interpreter strategy of *chunking down*) (cf. Gile 1995; Katan 1999);
- S4: literal interpretation of ST elements, producing unusual collocations in the target language (cf. Bartłomiejczyk 2006);
- S5: the target language element has a very different meaning to that of the ST element (i.e. the interpreter possibly misunderstood the ST element);
- S6: the target language element is meaningless or incoherent;
- S7: the target language element introduces a different perspective, tense or modality;
- S8: the target language element corresponds to an earlier ST utterance (i.e. the interpreter is using compensation strategies (cf. Bartłomiejczyk 2006);
- S9: repeated attempts to interpret a particular ST element (cf. Camayd-Freixas 2011).

These categories were identified using exploratory techniques and it is evident that the list of codes may be adapted or expanded, depending on the research interests.

8.3. ADDITIONS

Additions, i.e. the insertion of TT elements that cannot be assigned to a corresponding (although not necessarily equivalent) ST element (cf. Cokely 1992), are similarly perceived to be the result of interpreting strategies (cf. Bartłomiejczyk 2006; Dose 2010; Kalina 1998; Klaudy 2009; Leeson 2005; Ortiz 2011; Stratiy 2005; Stone 2009; Wallmach 2000). Additions are annotated as "_A" at word level or as "<A/> … </A>" above word level, with accompanying sub-category codes. The following types of additions were identified and coded in the SASL corpus:

- A1: repetition, e.g. “index@_ yes_ index@_A1 house_ bad_”;
- A2: addition of discourse markers, e.g. “index@_A2”, “first_A2”.
- A3: explicitation or explanation, e.g. “<A3/> finance_ group_ </A3> Z3DELOITE_”;
- A4: affirmation or emphasis, e.g. “yes_A4”;
- A5: new information, e.g. “SouthAfrica_ football_ association_ <A5/> say_ angry_ </A5>…;
- A6: meaningless or incoherent TT content not linked to a ST element, e.g. “same_ story_ … index@_ <A6/> say_ me_ </A6> community_ z4DOnTSE_”;
- A7: addition to set up a reference, e.g. “car index@CLcar_A7”;
- A8: anticipation, e.g. “<A8/> future_ look_ weather_ </A8>” (cf. Van Besian 1999).
- These categories can be further adapted according to individual research needs.

9. ERROR ANNOTATIONS

Finally, interpreters do commit errors and it may be of interest to the researcher to investigate them. What is categorized as an error depends largely on the theoretical model and question on which the research is based. The SASL corpus was based on a Descriptive Translation Studies (DTS) model (cf. Toury 1980; 1995) that does not categories semantic differences between ST and TT as errors, and was concerned with investigating comprehensibility; hence, only speech acts that negatively affected comprehensibility were annotated as errors. Errors are annotated as “_X” at word level and “<X/> … </X>” above word level, either as a single system or as a hierarchical system. In the SASL corpus a single system of codes was used, namely:

- careless or incomplete articulation of a sign (X0);
- incorrect articulation of a phonological parameter (X1)[6];
- inadequate translation i.e. s2 substitutions above superordinate categories (X2);
- misinterpretation (X3);

[6] Where applicable, the gloss of the intended sign is followed by that of the actual sign articulated separated by “/”, e.g. “charge/bad” indicates that the signer intended to sign “charge” but due to an incorrect handshape signed “bad” instead.

- ➪ too close following of the source text causing unnatural collocations in the TT (X4);
- ➪ insensitivity to Deaf cultural norms (X5);
- ➪ incorrect discourse markers (e.g. topic, reference) (X7);
- ➪ false starts (X8);
- ➪ pidgin language (e.g. signing keywords only) (X9);
- ➪ incorrect word order (XW);
- ➪ illogical pausing or lack of a logical pause (XP);
- ➪ incoherence at sentence level (XF).

On the other hand, a hierarchical error annotation system in which errors are classified into main and sub-categories allows investigation of errors occurring at multiple levels. For example, the annotation X*lip* records errors in language use (l), accuracy of information transfer (i) and production fluency (p) with further refinement of each category as suggested in Table 1 below:

Table 1. Suggested hierarchical error categories

Language use (l)	**Accuracy of information transfer (i)**	**Fluency of production (p)**
1 = phonological error (hand-shape, orientation, location, movement)	1 = over-translation	1 = careless/incomplete articulation
2 = syntactic error	2 = under-translation	2 = hesitations
3 = incorrect or omitted non-manual features[7]	3 = mistranslation	3 = incorrect or omitted pauses
4 = unnatural collocations	4 = incorrect referencing	4 = false starts
5 = pragmatic error e.g. violation of Deaf cultural norms	5 = topic incorrectly identified	
	6 = changes in perspective, tense etc.	

With this system, a "0" axial code represents correctness or adequacy in a particular parameter. For example, "index@4_X041" indicates a carelessly articulated deixis pointing to an incorrect reference point in signing space.

[7] Where relevant, these may form independent sub-categories.

Annotating interpreters' self-corrections allows the researcher to study interpreters' metalinguistic awareness of their interpreted product (cf. Bendazzoli et al. 2011; Napier & Barker 2004; Shlesinger 2000; Van Besian & Meuelman 2004). Corrections are annotated as "C" with the corresponding error code, e.g. "fire_X1 fire_C1" (using the single code system) or "fire_X100 fire_C100 (using the hierarchical code system).

10. CODE COMBINATIONS

The following example of transcribed text from the SASL corpus (Figure 7) illustrates how the annotation codes function together. In order to facilitate concordance searches, it is useful (but not necessary) to adhere to a hierarchy of codes in embedded annotations. Thus for the SASL corpus, the following embedded code order was adhered to: interpreting shifts (S, A), errors (X) or corrections (C), production quality markers (U, F), movement (M), topic (T), facial expression (E) and mouthing (V). To facilitate readability, the corresponding ST segments have been inserted after each line of TT transcription.

Figure 7. Example of SASL transcription

<time=2.32>
#19. <lag=0> <h8/> Jbritish_A3E330V0 tour_E330V000 person_E330V010(per) past_E310 Vdie_E310V000 Vshoot_A5 index@6_E310V030(who) Qwhat_E330V020(who) Z4GuGuLETU_E310V000 JCapetown_E310V000 </h8>
[ST: It now appears that the vehicle in which a tourist died in Guguletu…]
#20. <h2> Vthink_E220 car_tE220V000 index@CLcar_a2tE220V000 <hn> <V9yes> <hn> <h4> important_S0MRV000 <hn> <omit-Ptomurderinvestigation>.
[ST: … could be key to the murder investigation.]
#21. <lag=-1> <h1> man_MRV000 person_A3V0 <h8-> 26_E220V000 other_X0 <hn> <omit-Qfromcapetown> <h8-> VchargeD1_E310V000 for_F1 <h2> DVhijack_E220V000 Ualso_E330V000 <h8-> DVslitthroat_S3V000(kill) index@6_A2 index@CLperson_E310V0(per) <h4-> Z1DEWaNI_E310V000.
[ST: A twenty-six year old Cape Town man has been charged with the hijack and murder of Annie Dewani.]
#22. <lag=1> car_F1E250 CLcarLh_A2E250V000 <hc4/> behind_V000, index@4_A2E330 </hc4> <S1/> lie_E33/8 Vdie_V030(dead) </S1>
[ST: her body was found on the back seat of the car…]
#23. <omit-Phoneymoonchauffeur>.
[ST: she and her British husband were being chauffeured in, while sightseeing on their honeymoon.]
#24. <lag=0> <h2> <S2/> index@CLperson_TE220 <h1-> manD1_MRV000 </S2> Vsay_A6U0 tomorrow_E220V000 <omit-Qmagistrates> court_E310V000 KhayalitshaD1_V000.
[ST: The accused will appear in the Khayalitsha Magistrate's Court tomorrow.]

As can be seen from the excerpt, annotations for recording time and interpreter segmentation of material (termed *chunking*, cf. Moser-Mercer 1997/2002) are inserted in order to align source and target texts. Recording time is annotated as <time=min.sec>, e.g. <time=2.32>, whereas chunking was annotated using line numbers (e.g. #22.). The decision to base alignment on TT prosodic segments derived from the research question which investigated comprehensibility of the target message.

11. CONCLUSION AND SUGGESTIONS

In this article, a system of transcription and annotation for corpus-based/driven investigation of signed language interpreting using text-based concordance software packages is described. The annotations enable the researcher to investigate linguistic characteristics of the interpreter's sign language as well as interpreting features such as production quality and interpreter choices. The latter annotations are thus also pertinent to researchers investigating spoken language interpreting. The system comprizes six basic components, namely ID gloss tokens, a punctuation code, annotations for linguistic features, annotations for production quality, annotations for language use and annotations for interpreting features.

Firstly, the manual sign is represented by an ID gloss that reflects the sign's context-independent meaning, which prevents the imposition of derived theoretical knowledge on the one hand, and spoken language grammatical categories on the other. These glosses represent the corpus tokens. The system allows distinction between simple and compound signs, and also proposes a new gloss system for fingerspelling and numbers. In principle, the ID gloss system can also be used to transcribe spoken-language interpretations.

Secondly, punctuation is expressed using normal (spoken-language) punctuation markers, with the alternative option of using the pause delimiter (/). However, the use of capitals is restricted to spelling, annotations and signs based on fingerspelling handshapes. (The last category is specific to signed languages.)

Thirdly, linguistic information on handshape, mouthing, facial expression, location/movement and head/body movements can be annotated where relevant using simple alphanumeric codes. These parameters are important in the analysis of signed languages. It is also suggested that the annotations for facial expressions and head/body movements described in this paper can also be used to describe non-verbal communication in spoken interpretations.

Fourthly, the system allows for production quality to be annotated in terms of sign visibility, clarity of articulation and signing speed. These annotations can also be adapted for transcriptions of spoken language interpretations: the annotation for poor visibility (U0) could be used for poor audibility; the annotation for carelessly articulated or partially formed signs (X0) for mumbled or partial words; the annotation for incorrect phonology (X1) for mispronounced words. Similarly, the signing speed annotations (F1= faster than average and F2 = slower than average) can also be used to describe the production speed of spoken words. Although the annotation indicating that the sign is held over a period of time (F3) is specific to a visual language, it can be adapted to describe moments of hesitation in spoken interpretation that are not accounted for by normal punctuation pauses.

Fifthly, the system allows the researcher to annotate for interesting features of language use. Annotations for parts of speech (e.g. V, Q, U etc.), sociolinguistic variation (D1 etc.) and information density (J) are not specific to signed languages and can therefore also be used in spoken language transcriptions, whereas those for iconicity (I) and productive lexicon (DV) are specifically for signed languages. However, it is evident that similar annotations can be developed for spoken language features (e.g. onomatopoeia, neologisms, grunts, coughs etc.). Oral interpretations also include gestures (G) and it is suggested that these could be annotated in spoken-language corpora in a fashion similar to mouthing in sign language corpora, e.g. "look at that_G(index)", where G(index) indicates that the interpreter points while saying "that", or "<G(index)/> look at that </G(index)>" if the gesture occurs over a number of words.

Finally, the system allows categorization of interpreting features and errors. Within a descriptive model (cf. Toury 1995), interpreting features are annotated in terms of *shifts*, i.e. omissions (<omit-Rx>), additions (A) and skewed substitutions (S) with respect to the source text. However, it is suggested that a similar set of alphanumeric annotations can be derived to describe interpreter *strategies* of interest. Similarly, while the error annotations described in this paper were derived specifically to identify comprehension problems in signed interpretation, it is suggested that these annotations may be adapted or expanded to describe other relevant problems in both spoken and signed interpretation.

In conclusion, therefore, the transcription and annotation system outlined in this paper offers the researcher a comprehensive means of describing

various aspects of an interpretation. Although designed primarily to categorize comprehension problems in interpretations into a signed language (SASL), the system can be extrapolated to include further features of interest to a researcher or adapted to describe spoken language interpretations. Although it is also possible to use the annotation codes in the ELAN software package, it is primarily designed for text-based concordance software packages commonly used to analyze written corpora. It therefore significantly contributes to eliminating many of the obstacles and limitations previously faced by Interpreting Studies researchers using corpus-based/-driven approaches.

ACKNOWLEDGEMENT.

The research was partially funded by NRF Thutuka bursary TTK2006061700002 Grant No. 70261 linked to Dr Kim Wallmach.

BIBLIOGRAPHICAL REFERENCES

Akach, Philemon. "The Grammar of Sign Language." *Language Matters* 28 (1997): 7-35.

Alexieva, Bistra. "A Typology of Interpreter-Mediated Events." *The Translator* 3.2 (1997): 153-174.

Allwood, Jens, et al. *Guidelines for Developing Spoken Language Corpora*. Pretoria: University of South Africa, 2005.

Al-Salman, Saleh, and Raja'i Al-Khanji. "The Native Language Factor in Simultaneous Interpretation in an Arabic/English Context." *Meta* 47.4 (2002): 607-625.

Anderman, G. and M. Rogers, eds. *Incorporating Corpora: the Linguist and the Translator.* Clevedon, Buffalo & Toronto: Multilingual Matters, 2008.

Anthony, Lawrence. *AntConc (software) Version 3.2.4w (Windows)*. Tokyo, Japan: Faculty of Science and Engineering - Waseda University, 2011.

Baker, Mona. "Corpus Linguistics and Translation Studies: Implications and Applications." *Text and Technology: in Honour of John Sinclair*. Eds. Mona Baker et al. Amsterdam & Philadelphia: John Benjamins, 1993. 233-250.

–-, "Corpora in Translation Studies: an Overview and Some Suggestions for Future Research." *Target* 7.2 (1995): 223-243.

Baker, Mona et al., eds. *Text and Technology: in Honour of John Sinclair*. Amsterdam & Philadelphia: John Benjamins, 1993.

Barik, Henri. "A Description of Various Types of Omissions, Additions and Errors of Translation Encountered in Simultaneous Interpretation." *Bridging the Gap: Empirical Research in Simultaneous Interpretation*. Eds. Sylvie Lambert and Barbara Moser-Mercer. Amsterdam & Philadelphia: John Benjamins, 1994. 121-37.

Barlow, Mike. *ParaConc: a Concordancer for Parallel Texts*. Houston: Athelstanm, 2003. Available at: http://www.athel.com/paraconc.pdf [Consulted: 31-Oct.-2012].

Baroni, Marco and Sylvia Bernadini. "A New Approach to the Study of Translationese: Machine-learning the Difference between Original and Translated Text." *Literary and Linguistic Computing* 21.3 (2006): 259-74.

Bartłomiejczyk, Magdalena. "Strategies of Simultaneous Interpreting and Directionality." *Interpreting* 8.2 (2006): 149-174.

Bendazzoli, Claudio et al. "Disfluencies in Simultaneous Interpreting: a Corpus-Based Approach." *Corpus-based Translation Studies: Research and Applications*. Eds. Alet Kruger et al. London & New York: Continuum, 2011. 282-306.

Bendazzoli, Claudio and Annalisa Sandrelli. 2005. "An Approach to Corpus-Based Interpreting Studies: Developing EPIC (European Parliament Interpreting Corpus)." *MuTra* (2005): 12 pp. Available at: http://www.euroconferences.info/ proceedings/2005_Proceedings/2005_Bendazzoli_ Sandrelli. pdf [Consulted 2-Jul-2011].

—. 2008. "Corpus-based Interpreting Studies: Early Work and Future Prospects." *Revista Tradumatica* 7 (2008): 20pp. Available at: http://www.fti.uab.cat/tradumatica/revista/num7/articles/08/08 central.htm [Consulted: 22-Jul-2011].

Berman, Sandra and M Wood, eds. *Nation, Language and the Ethics of Translation*. Princeton: Princeton University Press, 2005.

Braffort, Annelies et al. "Sign Language Corpora for Analysis, Processing and Evaluation." *Proceedings of the76th International Conference on Language Resources and Evaluation (LREC 2010), Malta*: European Language Resources Association (ELRA), 2010. Available at: http://www. lrec-conf.org/proceedings/lrec2010/pdf/555 Paper.pdf [Consulted: 26-May-2012].

Bungeroth, Jan, et al. 2008. "The ATIS Sign Language Corpus." *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008), Marrakech*: European Language Resources Association (ELRA), 2008. 2943-2946. Available at: http://www.lrec-conf.org/lrec2008/pdf/748_paper.pdf [Consulted: 22-May–2012].

Camayd-Freixas, Eric. "Cognitive Theory of Simultaneous Interpreting and Training." *Proceedings of the 52nd Conference of the American Translators Association,* New York: American Translators Association (ATA), 2011. Aviillable at: http://www.dll.fiu.edu/people/full-time-faculty/erik-camayd-freixas/cognitive-theory-2011.pdf [Consulted: 7-Dec-2012].

Cencini, Mark and Guy Aston. "Resurrecting the Corp(us/se): Towards an Encoding Standard for Interpreting Data." *Interpreting in the 21st Century: Challenges and Opportunities*. Eds. Guiliana Garzone and Maurizio Viezzi. Amsterdam & Philadelphia: John Benjamins, 2002. 47-62.

Chernov, Ghelly. *Teoriya i Praktika Sinkronnogo Perevoda*. Moscow: Mezhdunarodniye Otnosheniya, 1978.

—. "Message Redundancy and Message Anticipation in Simultaneous Interpreting." *Bridging the Gap. Empirical Research in Simultaneous Interpretation*. Eds.

Sylvie Lambert and Barbara Moser-Mercer. Amsterdam & Philadelphia: John Benjamins, 1994. 139-153.

Cokely, Denis. *Interpretation: a Sociolinguistic Model*. Burtonsville, MD: Linstock Press, 1992.

Crasborn, Onno. "Corpus Studies of Mouth Behaviour." *BSL Corpus Project*. 2009. London: Sign Language Corpora: Linguistic Issues Workshop, 2009. Available at: http://www.bslcorpusproject.org/sign-language-corpora-linguistic-issues/ pro gram/ [Consulted: 16-May-2012].

Crasborn, Onno and Thomas Hanke. "Metadata for Sign Language Corpora. Background Document for an ECHO Workshop, May 8-9, 2003" Niemegen: Radboud University, 2010. Available at: http://www.let.ru.nl/sign-lang/echo /events.html [Consulted: 20-May-2012].

Dam, Helene. "Lexical Similarity vs Lexical Dissimilarity in Consecutive Interpreting." *The Translator* 4.1 (1998):49-68.

Diriker, Ebru. *De-/Re-Contextualizing Conference Interpreting: Interpreters in the Ivory Tower?* Amsterdam & Philadelphia: John Benjamins, 2004.

Dose, Stefanie. *Patterns of Growing Standardisation and Interference in Interpreted German Discourse*. Unpublished MA thesis, Pretoria: University of South Africa, 2010.

Dragsted, Barbara and Inge Hansen. "Speaking your Translation: Exploiting Synergies between Translation and Interpreting." *Interpreting Studies and Beyond*. Eds. Franz Pöchhacker et al. Copenhagen Studies in Language Series. Copenhagen: Samfundslitteratur Press, 2007. 251-274.

Galli, Christina. "Simultaneous Interpretation in Medical Conferences: a Case Study." *Aspects of Applied and Experimental Research on Conference Interpretation*. Eds. L Gran and C Taylor. Udine: Campanotto Editore, 1990. 61-81.

Garzone, Guiliana. "Quality and Norms in Interpretation." *Interpreting in the 21st Century: Challenges and Opportunities*. Eds. Guiliana Garzone and Maurizio Viezzi. Amsterdam & Philadelphia: John Benjamins, 2002. 107-120.

Garzone, Guiliana and Maurizio Viezzi, eds. *Interpreting in the 21st Century: Challenges and Opportunities*. Amsterdam & Philadelphia: John Benjamins, 2002.

Gile, Daniel. *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam & Philadelphia: John Benjamins, 1995.

Gollan, Christian et al. "Cross Domain Automatic Transcription on the TC-STAR EPPS corpus." Aachen: Aachen University, 2005. Available at: http://www-i6.informatik.rwth-aachen.de/ PostScript/InterneArbeiten/Gollan_Cross_ Dom ain_ Automatic_Transcrip tion_on_the_TC-STAR_ EPPS_Corpus_ ICA SSP_2005.pdf [Consulted: 1-Aug-2014].

Hanke, Thomas. "HamNoSys – Representing Sign Language Data in Language Resources and Language Processing Contexts." *DGS Korpus*. 2004. Hamburg: University of Hamburg, 2004. Available at: http://www.sign-lang.uni-hamburg.de/dgs-korpus/tl_files/inhalt_pdf/HankeLRECSLP2004_05. pdf [Consulted: 29-May-2012].

Hoiting, Nini and Dan Slobin. "Transcription as a Tool for Understanding: the Berkeley Transcription System for Sign Language Research (BTS)." *Directions in Sign Language Acquisition*. Eds. G. Morgan and Bencie Woll. Amsterdam & Philadelphia: John Benjamins, 2002. 55-75.

Isham, William. "Memory for Sentence Form after Simultaneous Interpretation: Evidence both for and against Verbalization." *Bridging the Gap: Empirical Research in Simultaneous Interpretation*. Eds. Sylvie Lambert and Barbara Moser-Mercer. Amsterdam & Philadelphia: John Benjamins, 1994. 191-211.

—. "On the Relevance of Signed Languages to Research in Interpretation." *Target* 7.1 (1995): 135-149.

Jakobsen, Arnt et al. "Comparing Modalities: Idioms as a Case in Point." *Interpreting Studies and Beyond*. Eds. Franz Pöchhacker, et al. Copenhagen: Samfundslitteratur. Copenhagen Studies in Language series, 2007. 217-249.

Janzen, Terry, ed. *Topics in Signed Language Interpreting: Theory and Practice*. Amsterdam & Philadelphia: John Benjamins, 2005.

Johnston, Trevor. "From Archive to Corpus: Transcription and Annotation in the Creation of Signed Language Corpora." *International Journal of Corpus Linguistics* 15.1 (2010): 106-131.

—. "Auslan Corpus Annotation Guidelines." *Auslan Corpus*. Sydney: Macquarie University, 2014. Available at: http://www.media.Auslan.org.au / attachments /Johnston_Auslan CorpusAnnotationGuidelines_14June2014.pdf [Consulted: 16-Sep-2015].

Johnston, Trevor and Adam Schembri. *Australian Sign Language: an Introduction to Sign Language Linguistics*. Cambridge: Cambridge University Press, 2007.

Jones, Roderick. *Conference Interpreting Explained*. Manchester: St Jerome, 1998.

Kalina, Sylvia. *Strategische Prozesse beim Dolmetschen: Theoretische Grundlagen, empirische Fallstudien, didaktische Konsequenzen*. Tübingen: Gunter Narr, 1998.

Katan, David. *Translating Cultures. An Introduction for Translators, Interpreters and Mediators*. Manchester: St Jerome, 1999.

Kenny, Dorothy. *Lexis and Creativity in Translation: a Corpus-Based Study*. Manchester: St Jerome, 2000.

Klaudy, Kinga. "Explicitation." *Routledge Encyclopaedia of Translation Studies*. Second edition. Eds. Mona Baker and Gabriella Saldanha. London & New York: Routledge, 2009. 104-108.

Koizumi, Atsuko et al. "An Annotated Japanese Sign Language Corpus." Hamburg: University of Hamburg, 2002. Available at: https://nats-www.informatik.uni-hamburg.de/intern/proceedings/ 2002/LREC/ pdf [Consulted: 25-May-2012].

Kruger, Alet et al., eds. *Corpus-based Translation Studies: Research and Applications*. London & New York: Continuum, 2011.

Kurz, Ingrid. "Conference Interpretation: Expectations of Different User Groups." *The Interpreters' Newsletter* 5 (1993): 13-21.

Lambert, Sylvie and Barbara Moser-Mercer, eds. *Bridging the Gap: Empirical Research in Simultaneous Interpretation*. Amsterdam & Philadelphia: John Benjamins, 1994.

Laviosa, Sara. *Corpus-based Translation Studies: Theory, Findings, Applications*. Amsterdam & New York: Rodopi, 2002.

Lederer, Michael. *La Traduction Simultanée*. Paris: Minnard Letteres Modernes, 1981.

Leeson, Lorraine. "Making the Effort in Simultaneous Interpreting: Some Considerations for Signed Language Interpreters." *Topics in Signed Language Interpreting: Theory and Practice*. Eds Terry Janzen. Amsterdam & Philadelphia: John Benjamins, 2005. 51-68.

Leeson, Lorraine and John Saeed. *Irish Sign Language: a Cognitive Linguistic Account*. Edinburgh: Edinburgh University Press, 2012.

Leeson, Lorraine et al., eds. *Signed Language Interpreting: Preparation, Practice and Performance*. Manchester: St Jerome, 2011.

Lombard, Susan. *The Accessibility of a Written Bible versus a Signed Bible for the Deaf-Born Person with Sign Language as First Language*. Unpublished MA thesis. Bloemfontein: University of the Free State, 2006.

Mauranen, A. "Universal Tendencies in Translation." *Incorporating Corpora: the Linguist and the Translator.* Eds. G Anderman and M Rogers. Clevedon, Buffalo & Toronto: Multilingual Matters, 2008. 32-48.

McEnery, Tony and Andrew Wilson. *Corpus Linguistics*. Edinburgh: Edinburgh University Press, 2001.

McKee, Rachel and David McKee. "Corpus Informed Lexicography: a Decade of Exploration." *Sign Language Corpora: Linguistic Issues Workshop, July24-25, 2009*. London: University College London, 2009. Available at: http://www. blscorpusproject.org/wp-content/uploads/mckee-corpus-wkshop2.pdf [Consulted: 25-Jul-2012].

McKee, David and Graeme Kennedy. "The Distribution of Signs in New Zealand Sign Language." *Sign Language Studies* 6.4 (2006):372-390.

Mohr, Susanne. *Mouth Actions in Irish Sign Language – their System and Functions*. Unpublished PhD thesis. Cologne: University of Cologne., 2011.

Moser, Peter. "Expectations of Users of Conference Interpretation." *Interpreting* 1.2 (1996): 145-78.

Moser-Mercer, Barbara. "Process Models in Simultaneous Interpretation." *The Interpreting Studies Reader*. Eds. Franz Pöchhacker and Miriam Shlesinger. London & New York: Routledge, 2002. 148-161.

—-. "Quality in Interpreting: Some Methodological Issues." *The Interpreters' Newsletter* 7 (1996): 43-55.

Napier, Jemima. "Free your Mind – the Rest will Follow." *Deaf Worlds* 14.3 (1998): 15-22.

Napier, Jemima and Roz Barker. "Sign Language Interpreting: the Relationship between Metalinguistic Awareness and the Production of Interpreting Omissions." *Sign Language Studies* 4.4 (2004): 369-393.

Neidle, Carol et al. *The Syntax of American Sign Language: Functional Categories and Hierarchical Structure*. Massachusetts: MIT, 2000.

Nonhebel, Annika et al. "Sign Language Transcription Conventions for the ECHO Project." Version 9, 20 Jan. 2004. Nijmegen: Radboud University, 2004. Available at: http://www.let.kun.nl/sign-lang/echo/docs/transcr_conv.pdf [Consulted: 15-May-2012].

Oléron, Pierre and Hubert Nanpon. "Recherches sur la Traduction Simultanée." *Journal de Psychologie Normale et Pathologique* 62.1 (1965): 73-94.

Olohan, Mauve. *Introducing Corpora in Translation Studies*. London & New York: Routledge, 2004.

Olohan. Mauve and Mona Baker. "Reporting 'That' in Translated English: Evidence for Subconscious Processes of Explicitation?" *Across Languages and Cultures* 1.2 (2000): 141-158.

Ortiz, Isabel. "Types of Error in the Learning of Spanish Sign Language as a Second Language." *Signed Language Interpreting: Preparation, Practice and Performance*. Eds. Lorraine Leeson et al. Manchester: St Jerome, 2011. 50-60.

Øverås, Linn. "In Search of the Third Code: an Investigation of Norms in Literary Translation." *Meta* 43.4 (1998): 571-588.

Özyürek, Asli et al. "Annotation and Coding of Spatial Expressions across Sign Languages." *Sign Language Corpora: Linguistic Issues Workshop, July24-25, 2009*. London: University College, 2009. Available at: http://www.bslcorpus project.org/sign-language-corpora-linguistic-issues /progra m/ [Consulted: 16-May-2012].

Paabo, Regina et al. "Rules for Estonian Sign Language Transcription." *TRAMES* 13.4 (2009): 401-424.

Penn, Claire et al. *Dictionary of Southern African Signs for Communication with the Deaf*. Pretoria: HSRC, 1992.

Perniss, Pamela et al., eds. *Visible Variation: Comparative Studies on Sign Language Structure*. Berlin: Mouton de Gruyter, 2007.

Pfau, Roland and Joseph Quer. "Non-manuals: their Grammatical and Prosodic Roles." *Visible Variation: Comparative Studies on Sign Language Structure*. Eds. Pamela Perniss et al. Berlin: Mouton de Gruyter, 2007. 1-21.

Pichler, Deborah et al. "Conventions for sign and speech transcription of child bimodal bilingual corpora in ELAN." *Language Acquisition and Interaction* 1.1 (2010): 11-40.

Pöchhacker, Franz. *Simultandolmetschen als Complexes Handeln*. Tübingen: Gunter Narr, 1994.

—. "Simultaneous Interpreting: a Functionalist Perspective." *Hermes Journal of Linguistics* 14 (1995): 31-53.

—. "Coping with Culture in Media Interpreting." *Perspectives* 15.2 (2007): 123-142.

Pöchhacker, Franz and Miriam Shlesinger. *The Interpreting Studies Reader*. London & New York: Routledge, 2002.

Prinetto, Paolo et al. "The Italian Sign Language Sign Bank: Using WordNet for Sign Language Corpus Creation." *International Conference on Communications and Information Technology (ICCIT). Aqaba*: IEEE, 2011. 134-137. Available at: http://ieexplore.ieee.org/ stamp/stamp.jsp? arnumber =05762664 [Consulted: 29-May-2011].

Prinsloo, Beatrice. *An Introductory South African Sign Language Grammar for the Beginner Sign Language Student*. Unpublished MA thesis. Bloemfontein: University of the Free State, 2003.

Pym, Anthony. "On Omission in Simultaneous Interpreting: Risk Analysis of a Hidden Effort." *Efforts and Models in interpreting and Translation Research: a Tribute to Daniel Gile*. Eds. Gyde Hansen et al. Amsterdam & Philadelphia: John Benjamins, 2008. 83-105.

Russell, Debra. *Interpreting in Legal Contexts: Consecutive and Simultaneous Interpretation*. Burtonsville, MD: Linstok Press, 2002.

—. "Consecutive and Simultaneous Interpreting." *Topics in Signed Language Interpreting: Theory and Practice*. Ed. Terry Janzen. Amsterdam & Philadelphia: John Benjamins, 2005. 135-164.

Russo, Mariachiara et al. "Looking for Lexical Patterns in a Trilingual Corpus of Source and Interpreted Speeches: Extended Analysis of EPIC (European Parliament Interpreting Corpus)." *FORUM: International Journal of Interpretation and Translation* 4.1 (2006): 21-54.

Sandler, Wendy and Diane Lillo-Martin. *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press, 2006.

Sandrelli, Annalisa and Claudia Bendazzoli. "Lexical Patterns in SI: a Preliminary Investigation into EPIC (European Parliament Interpreting Corpus)." Birmingham: University of Birmingham, 2005. Available at: http://www.corpus.bham.ac.uk-pclc [Consulted: 7-Dec-2012].

Savvalidou, Flora. "Interpreting (Im)politeness Strategies in a Media Political Setting: a Case Study from the Greek Prime Ministerial TV Debate as Interpreted into Greek Sign Language." *Signed Language Interpreting: Preparation, Practice and Performance*. Eds. Lorraine Leeson, et al. Manchester: St Jerome, 2011. 87-109.

Schjoldager, Anne. "An Exploratory Study of Translation Norms in Simultaneous Interpreting: Methodological Reflections." *Hermes Journal of Linguistics* 14 (1995): 65-87.

Segouat, Jérémie and Annelies Braffort. "Towards Categorization of Sign Language Corpora." *Proceedings of the 2nd Workshop on Building and using Comparable*

Corpora, ACL-IJCNLP, Singapore. Singapore: ACL, 2009. 64-67. Available at: http://www.aclweb.org/anthology/W/W09/W09-3111.pdf [Consulted: 14-May-2012].

Setton, Robin. "A Methodology for the Analysis of Interpretation Corpora." *Interpreting in the 21st Century: Challenges and Opportunities*. Eds. Guiliana Garzone and Maurizio Viezzi. Amsterdam & Philadelphia: John Benjamins, 2002. 29-45.

—. "Corpus-based Interpreting Studies (CIS): overview and prospects." *Corpus-based Translation Studies: Research and Applications*. Eds. Alet Kruger et al. London & New York: Continuum, 2011. 33-75.

Scott, Mike. *Wordsmith Tools, Version 5*. Oxford: Oxford University Press, 2011.

Shlesinger, Miriam. "Corpus-based Interpreting as an Offshoot of Corpus-based Translation Studies." *Meta* 43.4 (1998): 486-493.

—. *Strategic Allocation of Working Memory and Other Attentional Resources in Simultaneous Interpreting*. Unpublished PhD thesis. Ramat-Gan: Bar-Ilan University, 2000.

—, Miriam. "Towards a Definition of *Interpretese*: an Intermodal, Corpus-Based Study." *Efforts and Models in Interpreting and Translation Research: a Tribute to Daniel Gile*. Eds. Gyde Hansen et al. Amsterdam & Philadelphia: John Benjamins, 2008. 237-253.

Shlesinger, Miriam and Brenda Malkiel. "Comparing Modalities: Cognates as a Case in Point." *Across Languages and Cultures* 6.2 (2005): 173-193.

Stone, Christopher. *Towards a Deaf Translation Norm*. Washington, DC: Gallaudet University Press, 2009.

—. "Sign Language Interpreting: is it that Special?" *Journal of Specialized Translation* 14 (2010): 1-9. Available at: http://www.jostrans. org/Issue14/art_stone.php [Consulted: 6-July-2011].

Stratiy, Angela. "Best Practices in Interpreting: a Deaf Community Perspective." *Topics in Signed Languages and Interpreting: Theory and Practice*. Ed. Terry Janzen. Amsterdam & Philadelphia: John Benjamins, 2005. 231-250.

Sutton, Valerie. "A Global Writing System for a Global Age." 2012. Available at: http://www.valeriesutton.org [Consulted: 20-Jun-2012].

Sutton-Spence, Rachel. "Mouthings and Simultaneity in British Sign Language." *Simultaneity in Signed Languages, Form and Function*. Eds. Miriam Vermeerbergen et al. Amsterdam: John Benjamins, 2007. 147-162.

Sutton-Spence, Rachel and Bencie Woll. *The Linguistics of British Sign Language*. Cambridge: Cambridge University Press, 2006.

Tirkkonen-Condit, Sonja. "Unique items – Over- or Under-represented in Translated Language?" *Translation Universals: Do They Exist?* Eds. Anna Mauranen and Pekka Kujamäki. Amsterdam & Philadelphia: John Benjamins, 2004. 177-184.

Toury, Gideon. *In Search of a Theory of Translation*. Tel Aviv: Porter Institute, 1980.

—. *Descriptive Translation Studies – and Beyond*. Amsterdam & Philadelphia: John Benjamins, 1995.

Van Besien, Fred. "Anticipation in Simultaneous Interpretation." *Meta* 44.2 (1999): 250-259.

Van Besien, Fred and Chris Meuleman. "Dealing with Speakers' Errors and Speakers' Repairs in Simultaneous Interpreting." *The Translator* 10.1 (2004): 59-81.

Vermeerbergen, Miriam et al. "Constituent Order in Flemish Sign Language (VGT) and South African Sign Language (SASL)." *Sign Language and Linguistics* 10.1 (2007): 25-54.

Viaggo, Sergio. "The Quest for Optimal Relevance: the Need to Equip Students with a Pragmatic Compass." *Interpreting in the 21st Century: Challenges and Opportunities*. Eds. Guiliana Garzone and Maurizio Viezzi. Amsterdam & Philadelphia: John Benjamins, 2002. 229-44.

Visson, Lynn. "Simultaneous Interpretation: Language and Cultural Difference." *Nation, Language and the Ethics of Translation*. Eds. S. Berman and M. Wood. Princeton: Princeton University Press, 2005. 51-64.

Vuorikoski, Anna-Riita. *A Voice of its Citizens or a Modern Tower of Babel? Interpreting Quality as a Function of Political Rhetoric in the European Parliament*. PhD thesis. Tampere: Tampere University Press, 2004. Available at: http://acta.uta.fi/pdf/951-44-5878-8.pdf [Consulted: 12-Dec-2012].

Wadensjö, Cecilia. *Interpreting as Interaction*. London & New York: Longman, 1998.

—. "The Double Role of a Dialogue Interpreter." *The Interpreting Studies Reader*. Eds. Franz Pöchhacker and Miriam Shlesinger. London & New York: Routledge, 2002. 355-370.

Wallin, Lars. "Swedish Sign Language Corpus Project." *Institutionen för lingvistik*. Stockholm: Stockholm University, 2012. Available at: http://www.ling.su.se/english/research/research-projects/sign-language/swedish-sign-language-corpus/ [Consulted: 23-Apr-2013].

Wallin, Lars, et al. "Transcription Guidelines for Swedish Sign Language Discourse." Nijmegen: Radboud University, 2010 Available at: http://www.sign .net.ru.nl/ groups/slcwikigroup/wiki/ 4d6fa/ [Consulted: 26-May-2012].

Wallmach, Kim. "Examining Simultaneous Interpreting Norms and Strategies in a South African Legislative Context: a Pilot Corpus Analysis." *Language Matters* 31.1 (2000): 198-221.

Wehrmeyer, Jennifer. *A Critical Investigation of Deaf Comprehension of Signed TV interpretation*. Unpublished D. Litt. et Phil. Thesis. Pretoria: University of South Africa, 2013.

Woll, Bencie. "The Sign that Dares to Speak its Name: Echo Phonology in British Sign Language (BSL)." *The Hands are the Head of the Mouth: the Mouth as Articulator in Sign Languages*. Eds. Penny Boyes-Braem and Rachel Sutton-Spence. Hamburg: Signum Verlag, 2001. 87-98.

Zanettin, Federico. *Translation-driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies*. Manchester & Kinderhook: St Jerome Publishing, 2012.

Artículo recibido: 3/10/2013

Artículo aprobado: 10/4/2014