Prosodic Correlates of Sentences in Signed Languages

Ellen Ormel, Onno Crasborn
Centre for Language Studies, Radboud University Nijmegen

Accepted for publication in Sign Language Studies, 12:2 (2012)

Correspondence should be sent to: Ellen Ormel, Department of Linguistics, Centre for Language Studies, Radboud University Nijmegen, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands. E-mail: E.Ormel@let.ru.nl, Ph.: +31 (0)243616028, TT: +31 (0)243611516, Fax: +31 (0)243611070.

Abstract

This paper contains a literature review of prosodic evidence for large prosodic domains that can be equated with a syntactic unit (clause or sentence). In particular, different phonetic cues that may relate to clause or sentence boundaries are discussed in detail. On the basis of various ideas and views in the literature, we also describe two types of possible studies that would further our understanding of prosodic domains and their relation to sentence boundaries. The first type of study is a refinement to a recent series of perception studies in which the intuitions of signers are elicited on break points in stretches of discourse. The second type of study exploits new techniques from visual signal processing to detect salient events in a video recording of signing. The paper concludes with a discussion on how knowledge of the prosody of signed languages can be employed for language technology.

Acknowledgments

Production of this paper was made possible by EU FP7-ICT-2007-3 grant no. 231424 ‘SignSpeak’, ERC Starting Research grant no. 210373 ‘On the Other Hand’, and NWO VIDI grant no. 276-70-012 ‘On the Other Hand’.

1. Introduction

The identification of signed sentences in a larger stream of discourse is not a trivial task. On the basis of linguistic analysis, one can determine what counts as predicate, arguments and other syntactic material, and distinguish main clauses from embedded material. The semantic and syntactic analysis of signed languages is an ongoing matter of research and debate (e.g. Liddell 1980, 2003; Johnston 1991; Engberg-Pedersen 1993; Neidle et al. 2000; Sandler & Lillo-Martin 2006). How lexical and syntactic material actually appear to the viewer, i.e. the phonetic form of the language is mediated by a phonological level of organisation. Phonological forms from the syllable upwards have commonly been dubbed 'prosody', just as the rhythmic and melodic properties of spoken languages (Sandler 1999ab; Nespor & Sandler 1999). While many authors argue that the overall design of the grammar of signed languages shows large similarities to spoken language organisation (e.g., Meier 2002, Sandler & Lillo-Martin 2006, Sandler 2010), the phonetic substance of signed and spoken languages are very different (e.g., Crasborn 2001, Brentari & Crossley 2002; Sandler, to appear). The phonetic correlates of rhythm and intonation in signed languages consist of non-manual activities and modifications of manual phonological material. As at the syntactic level of organisation, it is not self-evident how clause-like or sentence-like units in signed languages can be identified by linguists on the basis of phonetic properties of the prosodic form.

The same would seem to be true of sign language users: perceivers cannot simply watch for one or two specific facial cues in order to segment the signing stream in sentences or large prosodic domains that are the phonological correlate of such syntactic units. At the same time, it is well-known from psycholinguistic studies on spoken language that prosody does help in the perception and recognition of spoken language (Gerken 1996; Cutler, Dahan, & van Donselaar 1997; Frazier, Carlson, & Clifton 2006); it is likely that this is also the case for signed language processing.

In a recent theme issue of the journal Sign Language & Linguistics on the unit 'sentence' in signed languages, both syntactic and prosodic perspectives were taken on the sentence in empirical studies of three signed languages (Crasborn 2007; Fenlon, Denmark, Campbell, & Woll 2007; Hansen & Heßmann 2007; Jantunen 2007). These all (implicitly or explicitly) subscribed to the conception of prosody as being related to, but not a direct expression of, syntactic structure (cf. Sandler, to appear; Crasborn 2010). Thus, phonetic events (whether in the face or on the hands) indirectly point to phonological structure, just as F0 (Fundamental frequency) and duration in spoken language are reflections of tones and prosodic groupings in speech (Pierrehumbert 1980, Selkirk 1984, Ladd 1996, Gussenhoven 2004). Phonological domains such as the utterance and the intonational phrase are related to syntactic structure, but the actual phrasing will depend in part on performance factors like speaking rate (Selkirk 1984, Nespor & Vogel 1986). Thus, the same syntactic string of words can be articulated in different ways, with different phrasing, showing more Intonational Phrases when it is articulated very slowly than when it is realised at high speed. The overall conception of the relation between syntax and phonetics in signed languages is thus highly similar to what Shattuck-Hufnagel and Turk (1996) sketch for speech. One of the possible models they conceive is presented in Figure 1. An alternative they suggest is a joint phonological component for prosody and segmental phenomena. While there is little consensus on the presence of a unit 'segment' in the phonology of signs, there surely is a sub-syllabic level of organisation that contains the phonological representation of lexical items (consisting of one vs. two-handedness, selected fingers and their configuration, orientation, a place of articulation, and movement properties) (Sandler 1989; Brentari 1998; Crasborn 2001; van der Kooij 2002; van der Kooij & Crasborn 2008).

Insert Figure 1 around here

In the present paper, we do not contribute new evidence to either prosodic or grammatical perspectives on the sentence, but we aim to present a literature review of prosodic evidence for large prosodic domains that can be equated with a syntactic unit 'clause' or 'sentence' (section 2). This paper was produced in the context of a European project on automatic sign language recognition and translation (SignSpeak; http://www.signspeak.eu), in which (after recognition of lexical items) the stream of signs in a video recording has to be segmented in some way by software in order to be able to translate sign language to the written form of a spoken language. This translation process is tied to sentences in the output language. For that purpose, it is an important question to ask whether sentence units are in some way visible in the video recordings of signed discourse. It is not crucial per se whether the units in question are actually sentences or sentence parts (clauses). For that reason, in this paper we will not address the topic of syntactic structure (and by consequence, neither the question how specific phonological domains are generated from the syntax). In addition, we will use the term 'sentence' and 'clause' fairly loosely, aiming to refer to a high syntactic domain that can receive a translation in a full written sentence. In reality, this will of course depend on the actual syntactic form of a given language, in our case Sign Language of the Netherlands.

On the basis of the various ideas and views in the literature, we describe two types of possible studies that can contribute to our understanding of prosodic cues to sentence boundaries (section 3). First of all, refinements are suggested to the perception studies by Nicodemus (2009), Hansen and Heβmann (2007), Hochgesang (2009), Fenlon et al. (2007), and Hermann (2009) in which the intuitions of signers and non-signers are elicited on break points in stretches of discourse. Secondly, we suggest that the use of new techniques from visual signal processing can be exploited to detect salient events in a video recording of signing. In this way, highly detailed information can be extracted on prosodic features. This can facilitate the analysis of larger data sets than would have been possible by phonetic transcription. Finally, in section 4 we end with a brief discussion on how knowledge of the prosody of signed languages can be employed for language technology, such as the automatic translation from signed to spoken language.

2. Literature review

In order to interpret a sentence adequately and to decide what pieces of information should be part of the same sentence and what belongs to a different sentence, several sources are used, such as syntactic, semantic, and discourse information. In addition, prosodic information provides additional support to interpret sentences in spoken as well as in signed languages. All utterances and components of those utterances must have a certain duration, amplitude, and fundamental frequency, and this is affected by the prosodic structure. When normal speech is recognized, prosodically determined variation is being processed (Cutler et al. 1997). The aim of many studies of prosody is to understand the process of recognition.

In contrast to prosody in spoken languages, prosody in signed languages has only been studied to a very limited extent, with empirical studies on only a very limited number of languages. First we will describe some studies on prosody in relation to sentences in spoken languages.

2.1 Prosody and sentences in spoken languages

In spoken languages, prosody is a widely discussed topic. Within the research on prosody, boundary features have been the most studied phenomena (Cutler et al. 1997). In a tutorial on prosody, Shattuck-Hufnagel and Turk (1996) explained the characteristics of spoken utterances that written equivalents do not have in terms of patterns of prosodic structure; intonation, timing, and variations in segmental implementation. The organization of a spoken utterance is not isomorphic to its morphosyntactic structure. Therefore prosody is of great importance for (auditory) sentence processing by human perceivers: we cannot hear all of the syntactic structure of a sentence directly, but are dependent on partly indirect cues such as those from prosody. The same is likely to be true for sign language processing, if one assumes the same general design of the grammar as in Figure 1. Definitions of prosody have been variable. Definitions can refer to the phonological organization of parts into higher-level constituents and to the hierarchies of relative prominence within these constituents (Selkirk 1984; Nespor & Vogel 1986). Constituents include intonational phrases, prosodic phrases, and prosodic words. Other definitions can refer to the acoustic parameters, presumably signaling constituent boundaries and prominence: F0, duration, amplitude and segment quality or reduction. A third group of definitions combine the phonological aspect of prosody at the higher level organization with the phonetic effects of the organization (e.g., F0, duration, segment quality/reduction). Shattuck-Hufnagel and Turk (1996) use the following working hypothesis: 'Prosody is both (1) the higher-level structures that best account for these patterns and (2) acoustic patterns of F0, duration, amplitude, spectral tilt, and segmental reduction, and their articulatory correlates, that can be best accounted for by reference to higher-level structures, (p.196). Evidence for prosodic constituents has been based on phonological observations as well as acoustic phonetic measurements. In an extensive literature review on prosody in the comprehension of spoken language, Cutler et al. (1997) similarly described that the term prosody is used in different ways: 'From at one extreme those who maintain an abstract definition not necessarily coupled to any statement about realization ("the structure that organizes sounds"), to those who use the term to refer to the realization itself, that is, effectively use it as a synonym for suprasegmental features ("e.g., pitch, tempo, loudness, pause") at the other extreme (p.142)'. The majority describe prosody between these two extremes, in using the term to refer to abstract structure coupled to a particular type of realization ("the linguistic structure which determines the supra segmental properties of utterances")' (Cutler et al. 1997, 142).

Evidence has been provided for the psychological reality of prosodic constituents, coming from studies of language behavior. For example, unit monitoring studies have provided evidence for the initial perceptual organization of spoken utterances. Other evidence on the psychological reality of prosody is based on the comparison between the perception of pauses or interruption points within constituents versus pauses at constituent boundaries.

One example of work concerning the psychological reality of prosodic constituents was carried out by Bögels, Schriefers, Vonk, Chwilla, and Kerkhofs (2009) who suggested that prosodic information could be sufficient for perceivers to determine the syntactic analysis of a spoken sentence. In other words, prosody not only provides some additional support for the interpretation of spoken sentences, but it directs the syntactic analyses. Cutler et al. (1997) explained that the presence of prosodic information cueing a boundary can influence syntactic analyses of listeners; however, they suggested that the evidence for this influence is not robust as yet. Prosodic information does not always cue syntactic information directly. In other words, prosody is informative towards syntactic structures, but whether it truly has a conclusive role is presently open for discussion.

Several studies have shown that the available prosodic cues in spoken languages are not always exploited by the listener (see Cutler et al. 1997). Furthermore, work on prosody in relation to sentence boundaries by Carlson, Clifton, & Frazier (2001) suggested that the interpretation of a prosodic sentence boundary is related to the existence and relative size of other prosodic boundaries in the sentence. This suggestion is in contrast to ideas that boundaries have a constant form independent of context. The different findings concerning the role of prosody to determine syntactic analyses as well as the findings concerning the exploitation and interpretation of prosodic cues indicate that conclusions about the specific role of prosody in spoken language processing are still refutable.

Shattuck-Hufnagel and Turk (1996) provided four helpful hints to support further research on auditory sentence processing studies that may form a starting point for looking at signed utterances: '1. Since prosody can't be predicted from text, specify the prosody of stimulus utterances as it was really produced; 2. Since prosody can't be read off the signal alone, inform acoustic measurements by perceptual transcription of the prosodic structure of target utterances; 3. Consider interpretation of results in terms of prosodic as well as morpho-syntactic structure; 4. Define those terms' (p.241).

Research on prosody in sign language has started even more recently than research on prosody in spoken languages. Many questions remain unanswered at this point. The study of sentence boundaries in relation to prosody in sign language is one of the areas that have been studied to a rather limited extent. The same four hints by Shattuck-Hufnagel and Turk (1996) on spoken sentence processing appear useful for looking at signed sentence processing. We will take up especially the first and the second point in our suggestion for future approaches concerning studies of prosody in signed languages, described in sections 3 and 4. The nature of prosody in sign language will be discussed next in section 2.2, followed by a review of a number of studies on prosody at sentence boundaries in sign languages in section 2.3. Subsequently, several recent developments are described on automatic detection of sentence boundaries in spoken and signed language in sections 2.4 and 2.5.

2.2 Prosody in signed languages

Similar to spoken languages, prosody can also be observed in sign languages, providing subtle and meaningful components of sentences (e.g., Dachkokovsky & Sandler 2009; Wilbur 2000). The finding that there is non-isomorphism between syntactic and prosodic constituents indicates that the requirement for rhythmic structure forms an independent property or phonological organization (see also Sandler & Lillo-Martin 2006). Nespor & Sandler (1999) defined the basic research theme of finding out whether prosodic patterns are exclusively present in spoken languages, or whether they also occur in sign languages. The latter would form an argument for prosody being a universal property of human language irrespective of modality.

Several studies on prosodic constituents have been conducted in sign languages. Nespor and Sandler (1999) presented evidence showing that Israeli Sign Language (ISL) sentences indeed do have separate prosodic constituents such as phonological phrases and, at a higher level in the prosodic hierarchy, Intonational Phrases (see also Sandler & Lillo-Martin 2006). Already in 1991, Allen, Wilbur and Schick studied the rhythmic structuring of sign language, in their case American Sign Language (ASL), by looking at three groups of signers: ASL fluent deaf signers, ASL fluent adult hearing children, and non-signing hearing adults. The participants were asked to tap a metal stick in time to the rhythm of signed narratives. The participants tapped in each of the three groups in a rhythmical way. In the paper, Allen et al. (1991) explained that non-native signers are often detected by the lack of rhythm is their signing. Some differences were found between the rhythmic identification patterns by the hearing non-signers and by the deaf signers, which showed that the rhythm of ASL is only observed fully when participants have a good grasp of sign language. Allen et al. (1991) mentioned that lexical stress distinctions as known for spoken languages, such as English, are not found in ASL, given that most ASL signs are monosyllabic. However, some signs involve reduplication of monosyllables, resulting in alternating accent patterns. For those repetitions, the recognition of primary and secondary stress was tested. One finding was that non-signers were not able to make the distinction between primary and secondary stress for those repeated syllables. Another difference was that adult hearing children of deaf adults (CODAs) showed much greater variability within the group than the other two groups did, suggesting that the rhythm of spoken English may have interfered for this group of CODAs.

Temporal aspects of signs which might function as prosodic elements were analyzed for Swiss-German Sign Language by Boyes Braem (1999), in an attempt to determine what makes the signing of native signers seemingly more rhythmic than the signing of late learners. For that reason, three early signers were compared to three late signers. Two kinds of rhythmic patterns (typically referring to beat or stress) were found in the study, which may have resulted in the observation of a sign as being rhythmic. One rhythmic pattern was the temporal balancing of syntactic phrases. Temporal balancing refers to phenomenon that the final phrase is produced with approximately the same duration as that of the preceding phrase. Here, no differences were found between early and late learners of sign language. The second kind of pattern was a regular side-to-side movement of the torso which appears to mark larger parts of certain types of discourse phonetically. Early learners showed this pattern more than late learners.

Prosodic cues can occur on the manual articulators (hands and arms) and nonmanual articulators (face, head, and body). Nonmanuals typically also add semantic information to the manual signs. Nonmanual markers include head position, body position, eye brow and forehead, eye gaze, nose position, and mouth, tongue, and cheek actions. As Wilbur (2000, 237) described: 'Nonmanual markers are integral components of the ASL intonation system, performing many of the same functions in the signed modality that pitch performs in the spoken modality'. In a number of studies, nonmanuals were investigated to gain insight in prosodic boundaries. In general, nonmanuals cue either the start or end of phrases (*boundary markers*) or their duration (*domain markers*). Wilbur also found that the large degree of layering in American Sign Language is a consequence of the modality. Layering serves prosodic purposes. The most apparent layering is that of the simultaneous production of nonmanual markings with manual signs.

One of the most frequently mentioned boundary markers are eye blinks. Already in 1978, Baker and Padden first introduced the importance of eye blinks in sign language research, in particular (what has later been referred to as) the *inhibited periodic blinks,* which often seemed to occur at the end of intonational phrases (see Wilbur 1994). There are also *voluntary blinks* that co-occur with lexical signs and indicate emphasis, assertion, or stress. The finding that voluntary blinks can be followed immediately by periodic blinks at the end of the constituents (slow and fast blinks) resulted in a study on the different functions of blinks (Wilbur 1994). Head nods likewise serve multiple functions, e.g., expressing focus, assertions, and/or existence. In other situations, head nods seem to function as edge markers, similar to blinks. Wilbur concluded that head thrusts do not seem to have a prosodic or syntactic function, but rather a semantic function, similar to some of the head nods; indicating the genuineness of or the speaker's commitment to the statement in a clause. Head thrusts are articulated with the lower jaw thrust forwards, and these occur on the last sign in some specific clauses (e.g. 'if' and 'when' clauses in ASL).

As we will emphasize several times in this paper, all these phonetic events that are related to lexical semantics, syntax, information structure, or pragmatics, also have a phonological form. The fact that they are clearly categorized and described as 'X' expressing 'Y' implies that the phonological analysis on the basis of the phonetic events has already taken place. However, this derivation from variable phonetic events to categorical phonological values ("brow raise") is typically not discussed in great detail in the literature.

The phonetics of speech prosody is extremely complex (cf. the work of Pierrrehumbert 1980, and related studies on phonetic implementation since), and we expect the same complex relation between phonology and phonetics in signed languages (Crasborn 2001).

As opposed to boundary markers, several nonmanual articulators can have a function as domain markers: eye brow position, eye gaze, head tilt, negative headshake, body leans, and body shift. These domain markers often appear to begin at the start of a constituent and end when the constituent ends. As they are articulated independently from the manual string of signs and each articulator can maintain a specific position, these positions can be kept constant for a certain time span. It is this phonetic affordance of the visual modality that makes the presence of domain markers specifically prominent and their roles quite diverse. Although tone in spoken language can also be raised or lowered for a certain domain, the many factors influencing F0 in speech (including a natural tendency to gradually lower one's voice over a sequence of syllables) make it quite hard to 'mark' domains for a perceiver in a constant way. For some nonmanual articulators in signed languages, it is slowly becoming apparent that there too, there are more influences on their phonetic appearance than was initially thought (see for example the study of De Vos, van der Kooij & Crasborn (2009) on eye brows in Sign Language of the Netherlands).

Three eye brow positions are often distinguished: raised, furrowed, and neutral (but see e.g. Baker-Shenk (1983) and De Vos et al. (2009) on the appearance of more detailed patterns). Raised brows occur on a number of apparently unrelated structures. Neutral brows occur on assertions, and furrowed brows occur with WH-questions and embedded WH-complements in some signed languages (Zeshan 2006). Prosodic modification of brow behavior can be observed as an effect of increased signing rate. The number of brow raises decreases as a result of fast signing rate without changing the syntactic structure (the number of constituents accurately marked by brow raise stay similar) (Wilbur 1994). Eye gaze as well as head tilts may express agreement information in some languages (Wilbur 1994). These can co-occur with each other and with different brow positions. Negative headshakes appear to have a grammatical function (Wilbur 1994). The interaction of negative headshakes with different head positions such as head tilt and head nod require further analysis, although it is to be expected that these are unable to coexist, given the semantic distinct reasons to use negative headshakes versus tilts and nods. Body leans can be side-to-side leans for prosodic purposes (as described by Boyes-Braem in 1994 on Swiss-German Sign Language), indicating other signers or different locative or temporal situations, but body leans can also move forward and backward (see also Wilbur & Patschke 1998, on ASL). The different leans can indicate prosodic emphasis on particular lexical items; semantic categories of inclusion 'even' and exclusion 'only'; and type of contrastive focus (Wilbur 2000, 235, on ASL). Van der Kooij, Crasborn and Emmerik (2006) found similar functions for Sign Language of the Netherlands. They found support for a multi-layered functioning of meanings: forward leaning was associated with one set of meanings, i.e. involvement, eagerness (lexical level), highlighting of information (subject focus), affirmation of sentence content, and belief; whereas backward leaning was associated with dislike, disgust, rejection (lexical level), highlighting of information (object focus), negation of sentence content, and disbelief. The authors emphasize that it was not always easy to distinguish linguistic and paralinguistic functions.

Body leans as well as body shift may be differentiated by their prosodic behavior, whereby body leans are restricted to signs and body shifts are restricted to left-edge markers within a certain (undetermined) domain (Wilbur 2000, on ASL). Finally, the nose has various contributions to sign language. Nose wrinkling can function as an affective marker but also as a contextual discourse marker. Wilbur (2000) surmises that there are probably other nose functions that are unknown as yet.

Part of the work on prosody in spoken as well as in signed languages has focused on segmentation and sentence boundaries. In the following part, we summarized a series of studies that have specific impact on the identification of clause and sentence boundaries in signed languages.

2.3 Prosody at sentence boundaries in signed languages

In spoken languages, pitch, length, spectral properties, and voice quality (and possibly others) together constitute the prosodic form of language (Lehiste 1972, Bolinger, 1986). Cues from these domains can be used by the listener to segment a string of sounds into words or larger units. According to Cutler et al. (1997) there has only been little discussion of the relative time-course of prosodic processing in sentence comprehension. Their statement refers to spoken languages, but of course this is even more so the case for signed languages. In the following overview and subsequent thoughts for future directions in section 3, we attempt to achieve more insight in prosodic processing in sign languages, in particular in relation to segmentation of signed discourse into clause or sentence units and the different prosodic boundary cues that relate to sentence boundaries. We start with a review of the literature on eye blinks, which is frequently discussed, and then move on to studies which aim to look at various cues in combination.

2.3.1 Studies of eye blinks

In a study by Wilbur (1994), signers of ASL were studied for eye blinking. Wilbur showed that the eye blinks of ASL signers were sensitive to syntactic structure. These findings provide insight into how intonational information shows in a signed language, information which is carried by pitch in spoken languages. Wilbur also looked into a range of other nonmanual markers, and distinguished between nonmanual markers carried on the lower part of the face and nonmanual markers on the upper face and head/body. The first were analysed as performing a lexical semantic function, whereas the latter ones carry grammatical and prosodic function. The nonmanuals that marked phrasal boundaries were blinks, head nods, change of head/body, and pauses. Wilbur distinguished three basic types of eye blinks: (1) reflexive blinks (not studied), (2) involuntary or periodic blinks, and (3) voluntary blinks.

The results showed that signers typically blink at the end of Intonational Phrases (right edge of ungoverned maximal projection); these are the involuntary blinks. Moreover, voluntary blinks are longer in duration and have greater amplitude than involuntary blinks. Voluntary blinks occur at lexical signs. Wilbur identified four different functions for blinks at boundaries: the marking of (a) syntactic phrases, (b) prosodic phrases (intonational phrases), (c) discourse units, or (d) narrative units.

Sze (2004) performed a study to examine the relationship between blinks, syntactic boundaries, and intonational phrasing in Hong Kong Sign Language (HKSL). Wilbur's findings on voluntary blinks (marking emphasis, assertion, or stress) and involuntary blinks (marking intonational bounderies) in ASL were evaluated empirically in HKSL. The classification made by Wilbur was insufficient to explain all data in the study on HKSL. Blinks in HKSL were also bearing a high correlation with head movement and gaze changes. Moreover, they could co-occur with syntactic boundaries of constituents smaller than a clause.

Wilbur made a distinction between lexical blinks and boundary blinks. Lexical blinks are voluntary and occur simultaneously with the lexical item, whereas boundary blinks are involuntary and occur at intonational phrase boundaries. The differences between the studies by Wilbur and by Sze may have resulted from differences in the exact measurement of the duration of a blink, differences in the exact measurement of the duration of a sign, and differences in determining whether a blink is produced voluntary or involuntary.

Whereas in Wilbur's study, 90% of the boundary blinks fall right on the intonational phrase boundaries, this was only true for 57% in HKSL. Secondly, many of the blinks (30%) in HKSL data co-occur with the last sign in a sentence, which seem to function as a boundary marker itself. However, according to Wilbur's description, these would have to be interpreted as a lexical blink. In addition, Sze described that some blinks are accompanied with another blink at the same sign. To account for these differences with Wilbur's findings, Sze proposes a new classification system of types of blinks: (1) Physiologically induced; (2) Boundary sensitive; (3) Related to head movement/gaze change, unrelated to syntactic boundaries; (4) Voluntary/lexically related; (5) Associated with hesitations or false starts. Type 1 and 5 do not have a linguistic function in the sense of being related to a specific lexical or grammatical form, although they can potentially have a function in perception in being related to performance phenomena related to false starts or hesitations. In addition to the new classification system, Sze suggested that changes in head movement may, in fact, be better than blinks to serve as clues for intonational phrase boundaries, given that blinks often co-occur with head movements. Moreover, it is rather uncertain whether the addressee makes use of blinks in perception, whereas head movements are larger and more easily observable, which makes it more likely to serve as an indication for intonational phrase boundaries in perception.

In a study of prosodic domains in ASL, Brentari and Crossley (2002) have used an extreme interpretation of Wilbur's findings, by adopting inhibited involuntary eye blinks in their methodology as the basic dividing marker between Intonational Phrases. On the basis of that division, further studies of manual and nonmanual markers of Phonological Phrases were sought. It appears that they did not take into account the 10% of IP boundaries in Wilbur's study on ASL (and 43% of IP boundaries in Sze's study on HKSL) that were not accompanied by an involuntary boundary blink. It seems to us that the conclusion that all voluntary blinks can be seen as reliable boundary markers is slightly premature.

Crasborn, van der Kooij, and Emmerik (2004) performed a small study similar to Wilbur (1994) and Sze (2004) concerning the prosodic role of eye blinks. Frequencies and locations of blinks were established in different types of data (monologues versus question-answer pairs), distinguishing between short and long blinks. The results indicated that a distinction seemed to be present between shorter and longer blinks. Longer blinks seemed to serve functions similar to other 'eye aperture' functions such as wide eyes and squinted eyes. Wide eyes express surprise, disgust and emphasis, whereas squinted eyes express fear and shared information. Closed eyes appear to express disgust and counter-assertion in NGT, and may be related to reference. With Sze (2004), the authors emphasize that frequency and duration of involuntary blinks and possibly also the frequency and duration of other types of blinks are likely to be influenced by the physical context, including factors such as the air humidity, amount of dust in the environment, and the temperature. This makes it hard to compare results across studies, and also findings from different recording sessions.

In sum, there does not appear to be a strict mapping of one articulator to one function in the eye opening parameter. One open question based on the literature is (1) what exactly is the distinction between a periodic and a voluntary blink. It is unlikely that this can be determined purely on durational criteria. The results also showed that blinks occurred at many locations in sentences, and it is an open question (2) whether they all relate to linguistic elements. Another open question is (3) whether sign language perceivers actually can and do perceive blinks, including the brief involuntary blinks, in normal interaction. There has been no targeted study for any sign language on the role of eye blinks in the perception of prosodic structure.

Finally, we would like to note that for studies of eye blinks (as in the other literature on nonmanuals), the difference between grammar and prosody is awkward if one adopts an overall model of language structure as in Figure 1. Nonmanual forms are often characterized in phrases like ‘Head shake is the grammatical marker of negation’. However, grammatical features (such as negation) have a prosodic phonological shape (such as head shake), and it is this phonological shape that can interact with the prosodic structure and with other phonological features. What we can actually see in a video recording is not a phonological shape but one specific phonetic instance of more general phonological categories. This phonetic implementation of a phonological form in sign is not likely to be much less complex than what we know from intonation research on languages like English or Dutch (see Crasborn 2001 on phonetic implementation of manual phonological categories in sign langauages; see Ladd 1996, and Gussenhoven 2004, for overviews of phonetic implementation in prosody). In other words, what can be observed in the phonetic appearance (a video recording, but also a kinematic recording of movement trajectories for example) is a cue that is somehow related to the phonological form that may have a semantic or grammatical function. It is likely that most instances of the phrase ‘nonmanual (grammatical) marker’ refer to the combination of a semantic value and its phonological shape. In relation to eye blinks, this implies that for the ‘lexical’, emphatic blinks, their duration and timing may vary, up to a point where their phonetic characteristics overlap with the prototypical involuntary blinks. The phonetic forms will have to be related to factors like signing speed in order to find any clear distinction between the two. For this, experimental phonetic studies will be indispensible.

In addition to phonological forms that are linked to a semantic or syntactic feature, there may appear to be prosodic phonological features that do not have any semantic or grammatical content, but only have the function to signal prosodic boundaries. If one would conceive of inhibited periodic eye blinks as similar to boundary tones in spoken languages, for example, then these still relate to the grammar in marking instances of the prosodic domain ‘Intonational Phrase’, for example, which is derived from the syntactic structure of the sentence. What makes them different is that their occurrence is induced by the physiology of the eye, only their timing appears to be related to prosodic structure. Thus, they would not form part of any phonological specification, but be generated purely by the phonetics. This makes it ever more likely that we will find large variation in the presence (vs. absence) of blinks, depending on various performance factors. The comparison with breathing comes in again here, as a comparable physiological condition related to speech events that may impact the appearance of some of the phonetic phenomena that we can observe and would otherwise be inclined to ascribe to linguistic structure.

2.3.2 Studies of multiple sentence boundary cues

Nespor and Sandler found that four markers almost always occur at Phonological Phrase boundaries in ISL: (1) Reduplication (reiteration of the sign); (2) Hold (freezing the signing hand or hands in their shape and position at the end of the last sign); (3) Pause (relaxing the hands briefly); and (4) Separate facial articulation. Hold, reduplication, and pause may all belong to the same phenomenon of lengthening given that these markers are almost never used simultaneously (see e.g. Byrd, Krivokapic, & Lee 2006, on lengthening in spoken languages). Moreover, the occurrence of an increased number of repetitions is dependent on the lexical sign having a specific distinctive feature [repeated movement]. Each of these specific prosodic cues may be perceptually prominent and suffice for the recognition of a domain. However, the modality favors a layering of information, for example for facial articulations such as eyebrows, eyelids, mouth which may be simultaneously present. The issue of simultaneous occurrence of prosodic cues requires further examination in future studies, as one would predict that the facial articulation would be able to co-occur with the manual cues.

According to Sandler and Lillo-Martin (2006), even more obvious prosodic cues occur at the boundaries of Intonational Phrases, as compared to Phonological Phrases. Intonational Phrases are at a higher prosodic level than Phonological Phrases. In spoken languages, Intonational Phrases are produced in one breath and begin with a new breath of air. There is a correspondence between breathing in spoken language and eye blinks in sign language in that both are imposed by the physiology of our body and will regularly occur irrespective of whether we are uttering linguistic units or not. It would therefore not come as a surprise if eye blinks would indeed indicate intonational boundaries in sign languages. In addition to blinks, at least two other characteristics noticeably indicate IP boundaries: changes in head position and major changes in facial expression. During an IP, head position seems to remain constant up to the point of boundary, where it clearly changes, providing a rhythmic cue for those phrases (Sandler & Lillo-Martin 2006).

Nespor and Sandler suggested that although rhythmic structures were clearly shown in a small data set, it would be interesting to use larger and more varied corpora, statistical analyses, and experimental studies to confirm to suggested patterns. In addition to providing evidence that prosodic cues form a highly valuable contribution for the detection of sentence boundaries in running sign language videos, one more specific interesting aspect of their findings so far is the apparent ability to define boundaries on the basis of the prosodic form, irrespective of the utterance content (assuming that more research will led to establishing strongly reliable boundary cues). This in turn is especially promising for the development of automatic sign language translation software, which will largely rely on sign form as opposed to content.

Fenlon, Denmark, Campbell, and Woll (2007) examined whether deaf native signers agreed on the locations of boundaries in the narratives and what cues are used when parsing the narratives. In addition, hearing non-signers were asked to mark sentence boundaries in order to compare the visual cues used in deaf native signers and hearing non-signers. Six native signers of BSL and six non-signers were asked to mark sentence boundaries in two narratives: in BSL and in SSL. Narratives were segmented in real-time, using the ELAN annotation tool (http://www.lat-mpi.eu/tools/elan/). Before assessing participants' responses, all Intonational Phrase (IP) boundaries in both BSL and SSL narrative were identified using a cue-based approach. Similar to the approach taken by Brentari and Crossley (2002), the occurrence of a blink between signs was used to indicate possible IP boundaries and these boundaries were further verified by the presence of other cues, such as pauses and head nods. Following identification, a strict 1.5 second window was applied to all IP boundaries in both signed narratives in which responses associated with that boundary could occur. Results indicated that the majority of responses from both groups in both narratives fall at IP boundary points (rather than only a Phonological Phrase boundary or a syntactic boundary).

The results showed that the number of cues present at each boundary varies from 2 to 8 (occurring simultaneously), showing prosodic layering in the sense of Wilbur's work (2000). Several cues were frequently present at boundaries, such as head rotation, head nods, blinks, and eye brows. Blinks formed one of the cues that highly frequently co-occurred with a boundary. However, many of those blinks were only detected by a limited number of signers (low boundary agreement), which can perhaps be related to the limited perceptual salience of eye blinks. Some other cues, such as pauses, drop hands and holds seemed to occur mainly at strong IP boundaries (i.e., boundaries agreed upon by at least four native signers or hearing non-signers). Because the cues could occur simultaneously and sequentially, it was difficult to detect the cues that were actually used by the participants to segment the narratives in the real-time online segmentation. No relationship was found between the number of cues present at an IP boundary and the identified boundaries. Further, participants seemed to be sensitive to the same cues in a signed narrative regardless of whether they know the language or not. However, segmentation was more consistent for the deaf signers than for the hearing signers.

The main conclusions were that phonetic cues of prosody form reliable indicators of sentence boundaries, while cues from grammar were argued not to be essential for segmentation tasks, as shown by the fast process in real-time, which does not seem to allow thorough processing of all cues from grammar for segmentation purposes. Some IP boundaries are perceptually stronger than others and can even be identified by those who do not know the language. For the cues at those boundaries (pauses, drop hands, holds), language experience seems to play a minor role only. It is therefore likely that the IP boundaries here coincide with even higher boundaries, such as of the prosodic domain Utterance, or an even larger discourse break. As far as the theory of prosodic phonology is concerned, all IP boundaries are equal, even though the phonetic cues of a specific IP boundary might be more prominent than those of another one, depending on context factors for example. Alternatively, it could be the case that some presumed IP boundary cues are actually cues for Utterance boundaries. Moreover, it was suggested that the many occurrences of head rotations and head movements may be unique to narratives. As a consequence, other kinds of materials in similar empirical studies might reveal (slightly) different findings. It would also be interesting to examine the differences and similarities in boundary identification between deaf and hearing participants if dialogues or monologues other than narratives would be used.

We find this type of study to be very valuable, as it aims to distinguish general phonetic cues from cues to language-specific structures. It is questionable whether at normal speed viewing, signers do not make use of their semantic and syntactic processing in determining boundaries, and really rely only on prosodic cues. Further studies are needed to verify the findings and confirm them for other languages and materials, and to analyze further any differences in boundary detection between those who are native users of a language and those who are not. We come back to this issue in a suggestion for further studies in section 3. Moreover, more research is needed to discover exactly how boundaries of identical and different layers differ (e.g., PP vs. IP vs. Utterance), whether in duration, intensity, or type of cue marking.

Hansen and Heβmann (2007) stated that none of the formal markers, such as blinks, change of gaze, lengthening and transitions, are conclusive in detecting sentence boundaries in German Sign Language (DGS), and that they cannot be determined independently from meaning. Nevertheless, three cues were argued to be useful indicators in determining sentence boundaries in DGS. As opposed to using a cue-based approach (blinks and some additional cues) to indicate Intonational Phrase boundaries, a short sample of DSG was segmented based on a functional analysis called TPAC (Topic, Predication, Adjunct, and Conjunct). The TPAC analysis supports the identification of boundaries of nuclear sentences. The results of this functional analysis were largely in line with, and to a degree refined, results based on intuitive judgements of sentences. As a next step, the occurrence of specific manual signs, interactively prominent gestures, head nods, eye blinks, gaze direction, pauses, and transitions were compared to the segments based on the TPAC analyses of propositional content to test if prosodic cues of sentence boundaries occur consistently or exclusively at relevant boundaries, as shown by the TPAC analysis.

The results showed that none of the prosodic cues consistently function as a boundary marker on their own. They stated: “As we will argue, signers recognize sentences by identifying propositional content in the course of a sense-making process that is informed but not determined by such form elements” (p. 146). Nevertheless, temporal adjuncts such as PAST and NOW seemed to point to an antecedent boundary of some sort. The palm-up gesture discussed in this study clearly requires attention in future studies on sentence boundaries (see also van der Kooij, Crasborn, & Ros 2006). Hansen and Heβmann showed that 60% of the palm-up gestures appeared at a sentence boundary. The remaining 40% referred to multiple further functions of the palm-up gesture, which altogether would appear to make it an inconsistent cue for sentence boundaries by itself. Similarly, head nods with a 'concluding force' (see also Bahan & Supalla 1995) may indicate sentence boundaries, and did occur to a minor extent in their data. This kind of head nods is interactively prominent, whereby the content of the previous part is affirmed before the signer continues. But as for the palm-up sign, Hansen and Heβmann suggest that it would be most peculiar to find head nods marking the boundaries of every sentence in ordinary discourse. Furthermore, eye blinks were found to co-occur rather often with sentence boundaries. For blinks, there seems to be some consistency in co-occurrence. At the moment, blinks do not seem to be conclusive for sentence identification, but suggestions for the occurrence of combinations of blinks with other cues have been provided in the past, for example by Wilbur in 1994, and also already in 1978 in an early study of nonmanuals by Baker and Padden. "Like breathing in spoken languages, a physiologically necessary brief closing of the eye may be expected to occur where it is least intrusive" (p.160). Hansen and Heβmann (2007) further indicated that eye blinks in ASL combine with eye gaze related features to indicate syntactic boundaries, but their data on DGS do not support the idea that this is conclusive.

A large-scale perception study of prosodic cues at sentences boundaries by native deaf participants on the production of signing by ASL sign language interpreters was performed by Nicodemus (2008). Fifty native deaf signers identified sentence boundaries in a video of an interpreted lecture at their own speed (as opposed to real-time judgments in Fenlon et al, 2007). The presence of twenty-one prosodic markers at the judged sentence boundaries were scored and grouped into one of the following four articulatory categories: *(1) Hands:* held hand shape, hand clasp, fingers wiggling, hands drop, signing space. *(2) head and neck:* head position (tilt (front and back), head position turn (left and right), head position: tilt (left and right), head movement: nod, head movement: shake, head movement (side to side), and neck tension. *(3) Eyes, nose, and mouth:* eyebrows, eye gaze, eye aperture, nose, and cheeks. *(4). Body:* breath, body lean, body movement, and shoulder actions. The most frequent markers present at the boundaries judged by the native signers were examined in each of these four categories. The results showed that in the category '*Hands'* the most frequent marker is hand clasp and the second most frequent is held hand shape. In the category '*Head and neck',* the most frequent marker is head tilt and the second most frequent is head turn. In the category '*Eyes, nose, and mouth',* the most frequent marker is increased eye aperture and the second most frequent is eyebrows. Finally, in the category '*Body',* the most frequent marker is body lean and the second most frequent is shoulders.

The cues involving larger articulators were the most frequent at boundary points (such as hand clasps and body leans). Markers of ongoing movements were used less frequently or in co-occurrence with a held marker. In addition to frequency, duration of the prosodic cues, the number of markers at each identified boundary, and the timing of the markers in relation to a target cue, which was the hand clasp. The longest duration was found for body lean and the shortest duration was found for eye aperture – which is what one would expect given the difference in mass between the articulators eye lid(s) and whole upper body. The maximum number of co-occurring cues of one sentence boundary was seven. Nevertheless, for 31% of the cues, a sequential timing pattern (occurring completely before or after a target cue) was found. The specific combinations of the occurring cues at the boundary points were not analyzed in detail, although the overall simultaneous use of (smaller) articulators was established for most of the cues (see also Nicodemus 2006).

The term 'sentence' was used to explain the task of segmenting in the written instruction. Which boundaries really refer to sentences in signed languages remains an open question. In contrast to studies the cue-based approach by Fenlon et al. (2007) and the TPAC analyses by Hansen and Heβmann (2007), Nicodemus only identified the boundaries based on native signers' boundary judgments, given that there is no easy morphosyntactic or semantic answer as to what sentences in sign languages should contain to form a coherent whole. As we indicated at the end of section 1, such an approach as taken by Nicodemus is not a problem if the perspective is that of automatic detection of prosodic boundaries; however, it does remain unclear what the 'sentences' look like. One of the further questions based on Nicodemus' extensive work is whether we can perceive the larger cues better simply because they are perceptually larger, or are they also most frequent if we would use video analysis to examine the occurrence of cues at boundary points? In other words, are the cues that lead to the identification of a sentence boundary driven by the perceptual needs of the viewer and are people therefore incapable to spot boundaries based on smaller cues during segmentation? Related is the question whether the most frequent cues are also the most reliable cues for segmentation (or do these cues at different locations in many occasions as well, which make them less successful for boundary identification). In addition, it might be the case that interpreters produce the various cues differently when compared to deaf signers. Further, it is unclear what exact consequences can be expected from the different time used in the tasks. Fenlon et al. (2007) presented the video in real-time whereas Nicodemus asked the participants to segment the video at their preferred speed. Finally, the specific type of instructions to the participants concerning the segmentation task may place a role in the task performance.

Hochgesang (2009) similarly studied sentence identification in ASL. Twenty-one deaf native and early sign language users from Gallaudet University looked at three clips of narratives. One of the topics examined by Hochgesang concerns the type of task instructions. The participants were divided into three groups (seven participants in each group). Each of the three groups received different instructions. Seven people were asked to identify sentences, seven were asked to identify where the periods should be, and seven were asked to identify where the narrative could be divided. The first time they saw the video they were instructed to look only without doing anything. On the second viewing, they segmented the data by reporting the time code of the video where they saw the end. The participants were instructed that they could change their answers if they wished. The results showed that the type of question asked to identify the boundaries of sentence-like units does not have much effect. Hochgesang also stated that the exact type of unit that was segmented is not quite clear. Transcription of sign language videos can be done at the level of intonation unit, utterance, idea unit, clause, sentence, Intonational Phrase, and possibly at yet other levels. Similar to Nicodemus, Hochgesang did not examine the content of the chunks that were identified by the deaf participants.

Herrmann (2009) also performed an extensive empirical study on prosody in German Sign Language (DGS), based on eight native DGS signers who were recorded for two hours each. Two hundred-forty short dialogs and contexts plus twenty-four picture stories were analyzed. Multiple prosodic cues were analyzed, referring to rhythm, prominence, or intonation. For rhythm, the following cues were analyzed: pauses, holds/frozen signs, lengthening, eye blinks, signing rate, head nods, reduplication, gestures. For prominence, the following cues were analyzed: head movement, eye brow movement, eye aperture, tense signing, lengthening and enlarging of sign. For intonation, eye brow movement, eye aperture, eye gaze, frown, facial expression, mouth gesture, and head movement were studied. Some cues are spread across multiple syllables, and function as domain markers. Domain markers that change at phrase boundaries include facial movement, head movement, and body movement (Herrmann 2009). Edge markers are observed at the prosodic phrase boundaries, for example, eye blinks, head nods, pauses, repetition of signs, holds, and final lengthening. Around a third of the blinks did not have a prosodic function according to Herrmann's analysis. At 78.8% of the Intonational Phrase boundaries, a blink was observed. In 94.7% of the Intonational Phrase boundaries, either blinks or other cues were observed. As in the previous studies discussed, there appears to be a complex interplay between prosodic markers as opposed to a one-to-one form and function relationship.

Equative sentences formed the subject of an investigation of Finnish Sign Language (FinSL) by Jantunen (2007). Equative sentences are nominal structures that are often applied for identification, such as introduction, defining and naming. In those equative sentences, Jantunen also studied the non-manual behaviors, including prosodic features such as eye blinks, eye gaze, eye brow movements, and movements of body and head position. Jantunen found that the non-manual behaviors in the different types of equative sentences showed substantial uniform occurrence of features. Similar to studies of blinks in American Sign Language (Wilbur 1994, 2000) and Hong Kong Sign Language (Sze 2004) Jantunen showed that in FinSL too, blinks were often present at sentence boundaries. However, blinks did not always occur at the boundaries, and moreover, blinks did also occur at sentence-internal phrase boundaries and in places other than sentence or phrase boundaries, for example within longer fingerspelled sequences. Sentence initial noun phrases showed combinations of widened/squinted eyes and raised/frowned eye brows. At the end of sentences, a head tilt was often observed. In general, alterations of head postures and also of body postures seemed to mark phrase or sentence boundaries, cf. the findings on ASL, HKSL, and ISL reported above.

Brentari (2007, in press) found that native signers and non-signers do sometimes differ in their segmenting strategies. In a segmentation study, signers and non-signers were asked to mark the edges of Intonational Phrases in passages of ASL which contained pairs of identical sequences of two different signs, either with an IP break between the signs or not. The sign pairs were produced as part of longer sequences of signs by instructors of ASL in Infant Directed Signing (IDS), and their interlocutor was a hearing toddler of 16 months of age who was familiar with baby sign. The use of IDS might have resulted in exaggerating their prosody. The cues in the stimuli in the two types of sign pairs (either with or without IP breaks between the signs) were eye blinks (70% between clause vs. 0% within clause), duration (mean 1100 msec between clause vs. 730 msec within clause), hold (mean 400 msec between clause vs. 66 msec within clause), pause (mean 780 msec between clause vs. 90 msec within clause), and drop hands (mean 70% between clause vs. mean 0% within clause). Brentari showed that native signers were more accurate at detecting the IP boundaries than non-signers.

In their work on identifying clauses in Auslan (Australian Sign Language), Johnston and Schembri (2006) found that signals such as pauses, blinks, changes in eye gaze, changes in brow position, changes in head position, and so forth, do not always systematically occur at boundaries. This suggests that any potential boundary cues are not completely grammaticalized, and most of these cues have a pragmatic function instead. As a result, seeing sentences in sign would present a challenge for linguistic analysis (Johnston & Schembri 2006).

Finally, in a paper by Kingston (1999) concerning the prediction of phonological analyses through experimental investigations of phonetic behavior, three issues in the analysis of prosody of signed languages were described from the perspective of laboratory phonology (Pierrehumbert, Beckman & Ladd 2000): the internal structure of the signed syllable, realization of lexical and phrasal prominence, and the marking of edges. Kingston discussed two specific topics related to boundaries: final lengthening and external sandhi. Similar to spoken languages, final lengthening marks the end of phrases, according to Kingston. Wilbur and Zelaznik (1997) showed that signs were longer in final position than in other positions, but that the velocity or displacements were not larger in final positions. However, this was different for phrase prominent signs, which were not longer in final positions than in other positions; yet phrase prominent signs do seem to have higher peak velocity and displacement. With respect to sandhi, it has been known that the pronunciation of the end of spoken words may influence the pronunciation of the edge of the next word, either as assimilation in the case of a categorical influence of phonological features, or by more a gradient gestural overlap. The absence of assimilation processes can also be informative about the presence of prosodic boundaries. This suggestion about the relation between assimilation and prosodic boundaries has not explicitly been taken up in subsequent research on sign prosody as far as we know. In a sense, spreading of the non-dominant hand (e.g. Sandler 1999a; Brentari & Crossley 2002) can be considered as a form of assimilation that stops at a prosodic boundary (typically a that between two Phonological Phrases) and not go beyond it, even if the subsequent signs would allow it in being one-handed.

2.3.3 Conclusion: combinations of phonetic cues to prosodic boundaries

The presentation of the rather diverse set of studies above has made clear, there is no evidence for a dominant role of one cues or a specific combination of cues in the signaling of prosodic boundaries. Multiple cues of both a durational and a punctual nature appear to be present in various sign languages including ASL, HKSL, DGS, FinSL, NGT, and BSL. Some authors point to the complex relation between syntax and phonetic form, with many authors proposing that there is a phonological level of organization as well. Although most authors would agree that this phonological level includes a hierarchical set of prosodic domains mediating between the two, cf. the speech model of Shattuck-Hufnagel and Turk (1996) that was presented in Figure 1 and cf. the seminal work of Nespor and Sandler (1999), very few if any authors explicitly discuss the fact that this should also hold for the actual phonetic events that are related to syntactic or semantic features, such as eye brow positions or head movements. As we have indicated above, the comparison with the overall linguistic organization of spoken languages makes this rather likely, yet there has been little research in this area. There have been very few observational or experimental studies of phonetic variation in the form of intonational features in signed languages.

In the next two sub-sections, we will briefly discuss how machine processing of speech and sign is trying to automatically recognize prosodic boundaries.

2.4 Automatic detection of sentence boundaries in spoken languages

Sentence boundary locations can be extracted from audio with reasonable reliability. Gotoh and Renals (2000) described an approach whereby sentence boundary information was extracted statistically from text and audio resources in broadcast speech transcripts. Pause duration based on speech recognizer outputs was used to establish boundaries, in addition to the conventional language model component that can identify sentence markers to some extent. The combination of the pause duration model and the language model appears to provide an accurate identification of boundaries.

As for text, it is important for the understanding of spoken language to find the location of sentence boundaries. In text, punctuation is structurally provided. This is not explicitly indicated in spoken language. Similar to Gotoh and Renals (2000), Stolcke et al. (1998) found that combining models (in their case a combination between prosodic and language model sources, modeled by decision trees and N-grams) led to better results than use of the individual models (see also, e.g., Shriberg, Stolcke, Hakkani-Tur, & Tur 2000). For their study, Stolcke et al. (1998) examined altogether three aspects of prosody: duration (of pauses, final vowels and final rhymes, normalized both for segment durations and speaker statistics), pitch (F0 patterns preceding the boundary, across the boundary, and pitch range relative to the speaker's baseline), and energy (signal-to-noise-ratio). These machine processing strategies indicated that for spoken languages, no single cue will ever be reliable enough to segment the stream of language production.

2.5 Automatic detection of sentence boundaries in signed languages

Sentence boundary detection in spoken languages is relatively new compared to the processing of textual data. Sentence boundary detection in signed languages is even more recent. Only a very limited amount of work has been done thus far. Nevertheless, it is as important as for spoken language to be able to detect boundaries in signed languages. In ongoing work by Koskela, Laaksonen, Jantunen, Takkinen, Raino, and Raike (2008), computer vision techniques for the recognition and analysis of gesture and facial expressions from video are applied to the area of sign language processing of FinSL. Existing video feature extraction techniques provide the basis for the analysis of sign language videos. Koskela et al. (2008) further apply an existing face detection algorithm to detect the eyes, mouth, and nose. To track the motion part of the manual signs, they apply a standard algorithm based on detecting distinctive pixel neighborhoods. The relation between motion and sign language sentence boundaries is currently under investigation. Results have not yet been published.

Related to the work by Koskela et al. (2008), Jantunen, Koskela, Laaksonen, and Raino (2010) described a technical method to visualize prosodic data in sign languages. Data on prosody in sign language were similarly represented graphically and analyzed semi-automatically from the digital video materials. Similar techniques are extensively used in spoken language research in software such as 'Praat' (http://www.praat.org), in which speech recordings are analyzed and presented graphically. In the past, several attempts were done to perform linguistic analysis of motion and other parameters when the signed videos were produced in pre-determined laboratory settings using complex motion tracking equipment. Jantunen et al described that one of the major advantages of the new techniques used is that videos no longer need to be produced in laboratory settings, but complex analysis of highly variable digital videos will become possible.

Three steps are taken in the analysis by Jantunen et al. (2010), as in other studies (e.g. Infantino, Rizzo & Gaglio 2006; Piater, Hoyoux & Du 2010): 1. Skin regions of the participants are detected; 2. Motion of these skin area is tracked; and 3. This motion is represented by using statistical descriptors. The technique is not ideal as yet: skin color detection cannot distinguish between skin on the hands and the face, and its success is dependent on lighting conditions and the color of clothing and background. Moreover, the distance between the signer and the camera cannot be measured effectively, so that movement in the front-back dimension has to be reconstructed, which can be a very complex task. This might get solved in the future, for example by using multiple video cameras with overlapping views.

As recently described by Piater, Hoyoux, and Du (2010), the automatic recognition of natural signing is demanding due to the presence of multiple articulators that each have to be recognized and processed (fingers, lips, facial expressions, body position, etc.), as well as to technical limitations such as restricted spatial and temporal resolution and unreliable depth cues. Similar challenges were mentioned by Jantunen et al. (2010) and by Crasborn, Sloetjes, Auer, and Wittenburg (2006). Two body areas were analyzed in the videos by Piater et al. (2010). The first part of the video analyses concerned detailed face tracking, extracting facial expressions such as mouth and eye aperture and eyebrow raise. The second part of the video analyses concerned hand tracking. The videos contained conversations and monologues in sign language. Further steps are currently being implemented to increase the overall tracking accuracy.

Any type of numeric data can be displayed along with video recordings in recent versions of ELAN, the multimodal annotation tool. Crasborn et al. (2006) described the development of this facility in ELAN. They presented the multimodal annotation tool in the context of the collection of kinematic recordings of hand and finger movements. Some of the major advantages of using such data rather than the output of video processing is the high spatial and temporal resolution that can be obtained by many motion tracking systems (the time resolution often ranging between 100 and 500 Hz compared to 25 Hz for PAL video). Among the possibilities for analyses based on raw position data retrieved are the calculation of velocity, acceleration and jerk, parameters that have been argued to be informative about stress (Wilbur 1990, 1999; Wilbur & Zelaznik 1997) that may also turn out to be informative about other aspects of prosody. One of the disadvantages of using kinematic equipment is the unnatural signing environment and the impossibility to analyze the growing number of video corpora on sign languages. However, as opposed to the method described by Jantunen et al. (2010), skin color does not need to be detected for this technique, and the two-dimensional data inherent in video recordings is no longer an issue, given that movement in all three spatial dimensions is recorded with equal accuracy.

Cutler et al. (1997) stated the following: "Of great value to future work for spoken languages are greater phonetic precision, consideration of cross-language variation, and a theoretical framework allowing explicit prediction (of prosody) towards processing effects" (p. 171). We suggest that the same is true for future work related to signed languages. It will be clear that the technical possibilities that have been discussed in this paragraph have not yet led to knowledge of segmentation of signed discourse into signs and sentences, but that these techniques hold great promise for the future. Knowledge of phonetic cues is slowly growing, including knowledge of prosodic cues of sentences. Nevertheless, none of the investigated prosodic cues seem to provide a fully reliable predictor function thus far for the presence of sentence boundaries. As Sandler and Lillo-Martin stated in 2006: "Neither instrumental tracking and transcribing of the prosodic system in sign language, nor experimental work on its perception and interpretation, have yet been done" (p. 265). In the next section, two possible research directions are described that can contribute to our understanding of prosodic cues at sentence boundaries, one involving experimental work on human perception and one involving instrumental tracking of phonetic cues.

3. Suggestions for two types of empirical studies

First of all, we suggest new tests of human segmentation of signed sentences. Secondly, we sketch how the use of new tracking techniques from visual signal processing can be engaged to detect salient events in a video recording of sign language.

3.1 Study 1: New tests of human segmentation of signed sentences

We would like to suggest several new tests of human segmentation of signed sentences, which include video manipulation of various kinds. The goal remains to gain further insight in the human detection of prosodic cues. The overall idea behind the tests is to prevent the semantic processing of the signing in the video by human subjects, signers or non-signers, as semantic processing can always interfere with the pure phonetic parsing. Similar methods have been used in studies of spoken languages whereby the spoken stream is manipulated in such a way that it becomes harder to understand the speech, while people could process prosodic features (van Bezooijen & Boves 1986; Mettouchi, Lacheret-Dujour, Silber-Varod, & Izre'el 2007). Typically, low-pass filters are applied which mask the segmental content while maintaining duration and melodic properties.

A parallel signal manipulation for sign language videos would involve a strong decrease in visual quality. This could be created for example by blurring the visual scene or by lowering the spatial resolution (dots per inch) in such a way that it becomes harder to understand the signing. In particular, the manual lexical content should become as incomprehensible as possible, as this is where most of the information load is located. Furthermore, if the aim would be to examine the specific contribution of prosodic features of the head and face in particular, separately from the contribution of the body and the manual prosodic features, only the face of the signers could be shown in an additional video manipulation without showing the remainder of the body. Conversely, if the aim would be to examine the specific contribution of prosodic features of the body and the hands, only the body of the signer could be shown without showing the head. Those suggestions are expected to strongly disrupt the processing of semantic information in the discourse when asked to make segmentation judgments. Both are also problematic given the many signs that have the head or face as the location for manual articulation. A technically simple way to circumvent this might be to have signers wear clothing and gloves with the same colour, or perhaps skin-coloured clothing, so as to obscure hand finger movements a bit. Alternatively, prosodic cues can be made to stand out in a (manipulated) video in order to make the prosodic cues more noticeable. For example, the eye contours can be highlighted with cosmetics to make short blinks more easily detectable, or body and head contours can be highlighted (e.g., changing colour when the head or body move in a certain way).

Quite a different method of hindering the semantic processing of the (sign) stream would be to process an unfamiliar (sign) language (e.g., Fenlon et al. 2007, for sign languages; Mettouchi et al. 2007, for spoken languages). In contrast to the processing of an unfamiliar spoken language, the processing of an unfamiliar sign language may be affected by the presence of some similarities between the familiar and unfamiliar sign languages, especially for those signs that are semantically motivated (showing some level of transparency between the meaning and the form of the sign). Fenlon et al. reported that the signers of British Sign Language in their study were able to partly understand stories in (unfamiliar) Swedish Sign Language. Signers of the familiar and unfamiliar sign language reported similar prosodic boundaries, possibly related to their partial understanding of the stories. This finding of partial understanding of signed stories emphasizes the additional benefit of using video manipulation to hinder the understanding of the content. In addition to processing of an unfamiliar sign language by signers, Fenlon et al. also involved hearing non-signers, who may have relied on cues they knew from gestures in face-to-face communication. Another problem in having users of a sign language process the prosody of an unfamiliar sign language is that the prosodic systems and the phonetic cues that are used may differ between the two languages. While the literature review in section 2 has mainly shown overlaps between languages in the type of prosodic cues that are used, it may still be the case that the precise timing and quality of the nonmanual articulation varies between languages in a more subtle way than has been studied so far. In fact, given what we know about spoken language prosody, it is highly likely that linguistic variation between sign languages is also located in the phonetic implementation of phonological features (e.g., Gussenhoven 2004). For that reason, we suggest that preventing access to semantic processing by video manipulation may be preferred over the use of subjects that do not master the sign language in question – whether foreign sign language users or non-signers.

More generally, independent of the precise method used for the elicitation of human segmentation judgments, post-hoc analyses of prosodic cues at high-likelihood segmentation regions would be useful to examine the co-occurrence of various combinations of cues (the end and start of domain markers in combination with boundary markers), as past empirical studies have shown that none of the individual prosodic cues provide sufficient predicative power for the occurrence of a sentence boundary. Several people have emphasised the presence of co-occurring cues at boundaries, however, specific combination of prosodic cues have not been specified thus far.

In summary, the video manipulations and the general suggestion for post-hoc analyses of co-occurring cues may provide new insights into the use of prosodic cues at sentence boundaries.

3.2 Study 2: Experimental tests of human vs. machine segmentation

In sign language research concerning prosody, video analysis and the use of movement capture techniques could also be applied to extract useful information on prosodic cues. Video analysis techniques, such as described by Piater et al. (2010) and Jantunen et al. (2010) are very promising with respect to the analysis of prosodic cues in non-restricted natural continuous sign language. The simultaneous access to numeric data and video recordings that is possible in tools like ELAN makes such analysis relatively accessible to a wide audience (Crasborn et al. 2006) Moreover, kinematic recordings of finger movements by using kinematic measurements are very promising, given the high temporal and spatial resolution, even though the recording setting will be somewhat less natural. One of the main advantages of video analysis is that large corpora can be processed this way. This in turn means that statistical patterns among the co-occurrence of different types of cues can then be calculated. These techniques would allow refinements of the recent human perception studies in which the intuitions of signers and non-signers were analyzed concerning boundaries in sign language discourse. It may prove useful to use data deriving from video analysis and data glove techniques to examine the actual co-occurrence of specific combinations of prosodic cues that appear to be most predictive for sentence boundaries cues. We should emphasize, however, that the discovery of statistical patterns in large data sets does not in itself constitute a linguistic analysis of the structures in question. They should be considered as tools for linguistic analyses, just as quantitative analyses are used in spoken language phonetic research. It is the linguistic model that should generate the hypotheses to test. For automatic recognition and translation of signed languages, however, there are more direct advantages of computer processing of phonetic cues.

To illustrate the possibilities of facial and head feature detection, Figure 2 shows an example of feature extraction of the face, developed by Piater et al. (2010, see also Dreuw et al. 2010) using video analysis. For each of the four video images, three drawings are presented of the fitted model. From top to bottom, the following models are presented: a full model instance, a meshed shape, and a plotted shape. In addition, three vertical lines are present in the left of the image, reflecting the measurement of several nonmanual cues. From left to right, the following features are presented: left eye aperture, mouth aperture, and right eye aperture. The three axes on the face represent information on the orientation of the face: with the origin at the tip of the nose, the red line is the X axis, the green line the Y axis, and the blue line is the Z-axis.

Insert Figure 2 around here

Shattuck-Hufnagel and Turk (1996) emphasized that in studies of prosody on spoken sentences, acoustic measurements should ideally be complemented by perceptual measurements of the prosodic structure, since it is difficult to detect prosody based on the signal alone. This is true in the case of sign languages as well, in particular because retrieving prosodic information of the sign signal has only just started. At the same time, as we noted in section 2.1, Cutler et al. (2007) found that not all prosodic cues are used by perceivers of spoken languages.

Currently, some prosodic signals can be detected more easily than others by video analyses. Some of the restricting factors are exceptionally large or small movements, the visual similarity between the skin of the hands and of the face when manual signs occur in front of or nearby the face, and the general quality of video materials. Theoretically, these specific restrictions do not apply to kinematic measures when using data gloves. However, although hands can be tracked relatively well with data gloves or small markers attached to the body, some features in the face, such as eye blinks, would require different techniques.

Regardless of these challenges in video analyses and kinematic recordings, video analyses, kinematic recordings, and intuitive judgments can be mutually highly informative for the increase of our knowledge on sentence boundaries. Video analyses and kinematic recordings can provide exceedingly detailed data on each of the possible cues. However, intuitive judgments can provide information of the actual presence of a prosodic boundary in sign language. Without native informants to confirm the feature extraction data during the initial phase of analysis of (combination of) features deriving from video analyses, it would be impossible to judge whether the data in fact pointed towards boundaries or not. Beyond the current technical restrictions in video analysis, the equivalent of the ‘acoustic’ measurements based on automatic measures should therefore be combined with the perceptual transcription of the prosodic boundaries. In an ideal situation, the perceptual measures and transcriptions should be elicited from (native or near-native) signers who can provide intuitive judgments on their own language.

As was already noted above, additional phonological analyses would subsequently be necessary to provide more information on the actual domains that are identified. Equally, the identification of boundaries by intuitive signers can be analyzed into much more depth if not only manual annotations of (co-)occurring cues were provided for the identified boundaries, but also information based on the feature extraction data that cannot be determined by the naked eye, such as temporal properties (detailed duration, velocity, and acceleration).

4. Conclusion

How can knowledge of the prosody of signed languages be employed for language technology, such as the automatic translation from signed to spoken language? As long we have not captured reliable predictors (specific combinations of prosodic cues) of sentence boundaries and a reliable method to capture these prosodic cues, at least three challenges will be present. First of all, a time-consuming manual procedure will remain to be required to annotate each sentence boundary, based on native signer’s intuitive judgments. Secondly, automatic translation will remain relatively difficult, as the only predictors of sentence boundaries (or put differently, the cohesion of a series of signs in a larger syntactic unit) are statistical patterns in the semantic cohesion between the lexical items that have been recognised by the system. Finally, as long as sentence boundary annotations need to be added manually, interpersonal variation in the annotation process is inevitable and some prosodic cues cannot be observed in sufficient detail. This implies that increased knowledge of prosody of sign languages and prosody related to sentence boundaries in particular can lead to better predictions of the presence of boundaries. This in turn will help to improve automatic translation from sign to speech.

Once these predictive cues can be captured automatically and reliably by means of video analysis (possibly combined with the detection of lexical items that often occur at sentence boundaries), sentence units can be automatically generated. For linguistic analyses, this can facilitate the analysis of much larger data sets than would have been possible by phonetic transcription. Moreover, the captured chunks can in theory be translated automatically. Development of sentence translation software is of course further dependent on a sufficient number of signs in the sentence that are correctly recognized by machine recognition algorithms. Many challenges are still to be overcome in this area, given our restricted understanding of sign languages and the relatively short research tradition in this area (Dreuw et al. 2010; Ormel et al. 2010).

References

Allen, G. D., Wilbur, R. B., and Schick, B. B. 1991. Aspects of rhythm in American Sign Langauge. *Sign Language Studies* 72: 297–320.

Bahan, B. N., and Supa;;a, S.J. 1995. Line segmentation and narrative structure: A study in eyegace behavior in American Sign Language. In *Language, gesture and space*, ed. K. Emmorey and J. Reilly, 171–191. Hillsdale, NJ: Lawrence Erlbaum.

Baker, C., and Padden, C. 1978. Focusing on the nonmanual component of ASL. In *Understanding language through sign language research*, ed. P. Siple, 25–27. New York: Acedemic Press.

Baker-Shenk, C. 1983. *A microanalysis of the nonmanual components of questions in American Sign Language*. Unpublished PhD Thesis, University of California, Berkeley, Berkeley.

Bezooijen, R. van, & Boves, L. 1986. The effects of low-pass filtering and random splicing on the perception of speech. *Journal of Psycholinguistic Research* 15(5): 403–417.

Bolinger, D. 1986. *Intonation and its parts. Melody in spoken English*. Stanford, CA: Stanford University Press.

Boyes Braem, P. 1999. Rhythmic temporal patterns in the signing of deaf early and late learners of Swiss German Sign Language. *Language and Speech* 42(2/3): 177–208.

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D. J., Kerkhofs, R. 2009. The interplay between prosody and syntax in sentence processing: The case of subject- and object-control verbs. *Journal of Cognitive Neuroscience* 22(5): 1036–1053.

Brentari, D. 1998. *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press.

Brentari D. 2007. *Perception of sign language prosody by signers and nonsigners*. Paper presented at Workshop on Visual Prosody in Language Communication. Max Planck Institute for Psycholinguistics. Nijmegen, May 10–11, 2007.

Brentari, D., González, C. , Seidl, A., and Wilbur, R. in press. Sensitivity to visual prosodic cues in signers and nonsigners. *Language and Speech.*

Brentari, D., and Crossley 2002. Prosody of the hands and face: Evidence from American Sign Language. *Sign Language and Linguistics* 5(2): 105–130.

Byrd, D., Krivokapic, J., and Lee, S. 2006. How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America* 120(3): 1589–1599.

Carlson, K., Clifton, C., and Frazier, L. 2001. Prosodic boundaries in adjunct attachment. *Journal of Memory and Language* 45: 58–81.

Crasborn, O. 2001. Phonetic implementation of phonological categories in Sign Language of the Netherlands. Utrecht: LOT.

Crasborn, O. 2007. How to recognize a sentence when you see one. *Sign Language & Linguistics* 10: 103–111.

Crasborn, O. 2010. *Prosody and the non-dominant hand in Sign Language of the Netherlands*. Paper presented at the 20th Colloquium on Generative Grammar, Barcelona, March 18, 2010.

Crasborn, O., Sloetjes, H., Auer, E., and Wittenburg, P. 2006. Combining video and numeric data in the analysis of sign languages within the ELAN annotation software. In *Proceedings of the 2nd workshop on the representation and processing of sign languages: lexicographic matters and didactic scenarios*, ed. C. Vettori, 82–87. Paris: ELRA.

Crasborn, O., Kooij, E., van der, and Emmerik, W. 2004. *Eye blinks and prosodic structure in NGT.* Talk presented at a European Science Foundation (ESF) Exploratory Workshop, Modality Effects on the Theory of Grammar: A Crosslinguistic View from Signed Languages of Europe. Barcelona, Spain, 14–18 November 2004.

Cutler, A., Dahan, D. and Donselaar, W. van 1997. Prosody in the comprehension of spoken language: a literature review. *Language and Speech* 40(2): 141–202.

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

Dreuw, P. Forster, J., Gweth, Y., Stein, D., Ney, H., Martinez, G., Verges Llahi, J., Crasborn, O., Ormel, E., Du, W., Hoyoux, T., Piater, J., Moya Lazaro, J.M. and Wheatley, M. 2010. SignSpeak – Scientific Understanding and Vision-Based Technological Development for Continuous Sign Language Recognition and Translation. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies,* ed. P. Dreuw, E. Efthimiou, T. Hanke, T. Johnston, G. Martinez Ruiz and A. Schembri, 65–72. Paris: ELRA.

Engberg-Pedersen, E. (1993). *Space in Danish Sign Language: the semantics and morphosyntax of the use of space in a visual language*. Hamburg: Signum.

Fenlon, J. Denmark, T, Campbell, R., and Woll, B. 2007. Seeing sentence boundaries. *Sign Language and Linguistics* 10(2): 177–200.

Frazier, L., Carlson, K., and Clifton, J. C. 2006. *Prosodic phrasing is central to language comprehension*. *Trends in Cognitive Sciences* 10(6): 244–249.

Gerken, L. 1996. Prosody's role in language acquisition and adult parsing. *Journal of Psycholinguistic Research* 25(2): 345–356.

Gotoh, I., and Renals, S. 2000. *Sentence boundary detection in broadcast speech transcripts*. Proceedings of the International Speech Communication Association (ISCA) Workshop: Automatic speech recognition: Challenges for the new millennium (ASR-2000), Paris, September.

Gussenhoven, C. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Hansen, M., and Heβmann, J. 2007. Matching propositional content and formal makers. Sentence bounderies in DSG text. *Sign Language and Linguistics* 10(2).

Herrmann, A. 2009. Prosody in German Sign Language. *Workshop on Prosody and Meaning,* 17–18 September 2009, Frankfurt am Main.
http://prosodia.upf.edu/activitats/prosodyandmeaning/home/presentations.php

Hochgesang, J. A. 2009. Is there a 'sentence' in ASL? Insight on segmenting signed language data. Talk presented at *Sign Language Corpora: Linguistic Issues Workshop*, July 24th, London.
http://www.bslcorpusproject.org/sign-language-corpora-linguistic-issues/programme/

Infantino, I., R. Rizzo and S. Gaglio 2006. A Software System for Automatic Signed Italian Recognition. In *Proceedings of the second workshop on the Representation and Processing of Sign Languages*, LREC 2006, ed. C. Vettori, 25–30. Paris: ELRA.

Jantunen, T. 2007. The equative sentence in Finnish Sign Language. *Sign Language and Linguistic* 10(2): 113–143.

Jantunen, T. J., Koskela, M., Laaksonen, J. T., Raino, P. I. 2010. Towards the automatic visualization and analysis if signed language prosody: Method and linguistic issues. *Paper to appear in the proceedings of the 5th International Conference on Speech Prosody*. May 11–14, Chicago, Illinois (USA).

Johnston, T. 1991. Spatial syntax and spatial semantics in the inflections of signs for the marking of person and location in AUSLAN. *International Journal of Sign Linguistics* 2(1): 29–62.

Johnston, T. and Schembri, A. 2006. *Identifying clauses in signed languages: applying a functional approach.* Paper presented at the DGfS [Deutsche Gesellschaft für Sprachwissenschaft] 2006 workshop How to recognize a sentence when you see one: methodological and linguistic issues in the creation of sign language corpora, 23–24 February, Bielefeld, Germany.

Kingston, J. 1999. Extrapolating from spoken to signed prosody via laboratory phonology. *Language and Speech* 42(2–3): 251–281.

Kooij, E. van der. 2002. *Phonological categories in Sign Language of the Netherlands. The role of phonetic implementation and iconicity.* Unpublished PhD Thesis, LOT, Utrecht.

Kooij, E. van der, and Crasborn, O. 2008. Syllables and the word prosodic system in Sign Language of the Netherlands. *Lingua* 118: 1307–1327.

Kooij, E. van der, Crasborn, O., and Emmerik, W. 2006. Explaining prosodic body leans in Sign Language of the Netherlands: Pragmatics required. *Journal of Pragmatics* 38(10): 1598–1614.

Kooij, E. van der, Crasborn, O., and Ros, J. 2006. *Manual prosodic cues: palm-up and pointing signs.* Poster presented at the 9th conference on Theoretical Issues in Sign Language Research, Florianópolis, Brazil.

Koskela, M., Laaksonen, J., Jantunen, T., Takkinen, R., Raino, P., and Raike, A. 2008. Content-based video analysis and access for Finnish Sign Language - a multidisciplinary research project. *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora,* ed. O. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood & E. Thoutenhoofd, ELDA, organised as a part of the 6th Language Resources and Evaluation Conference (LREC). 1 June, Marrakech, Morocco.

Ladd, D. R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

Lehiste, I. 1972. *Suprasegmentals*. Cambridge, MA: MIT Press.

Liddell, S. 1980. *American Sign Language syntax*. The Hague: Mouton.

Liddell, S. 2003. *Grammar, gesture and meaning in American Sign Language*. Cambridge: Cambridge University Press.

Meier, R. 2002. Why different, why the same? Explaining effects and nn-effects of modality upon linguistic structure in sign and speech. In *Modality and structure in signed and spoken languages.* ed. R. Meier, K. Cormier and D. G. Quinto-Pozos, 1–26. Cambridge: Cambridge University Press.

Mettouchi, A., Lacheret-Dujour, A., Silber-Varod, V., & Izre'el, S. 2007. Only Prosody? Perception of Speech Segmentation in Kabyle and Hebrew. *Nouveaux cahiers de linguistique française* 28: 207–218.

Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B., and Lee, R. G. 2000. *The syntax of American Sign Language: functional categories and hierarchical structure*. Cambridge, MA: MIT Press.

Nespor, M., and Vogel, I. 1986. *Prosodic phonology*. Dordrecht: Foris.

Nespor, M., and Sandler, W. 1999. Prosody in Israeli Sign Language. *Language and Speech* 42(2–3): 143–176.

Nicodemus, B. 2006. *Prosody and utterance boundaries in ASL interpretation*. Paper presented at the DGfS [Deutsche Gesellschaft für Sprachwissenschaft] 2006 workshop How to recognize a sentence when you see one: methodological and linguistic issues in the creation of sign language corpora, 23–24 February, Bielefeld, Germany.

Nicodemus, B. 2009. *Prosodic markers and utterance boundaries in American Sign Language interpretation*. Dissertation. Gallaudet University Press, Washington D.C. 20002.

Ormel, E., Crasborn, O., van der Kooij, E., van Dijken, L., Nauta, E. Y., Forster, J., and Stein, D. 2010. Glossing a multi-purpose sign language corpus. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies,* ed. P. Dreuw, E. Efthimiou, T. Hanke, T. Johnston, G. Martinez Ruiz, and A. Schembri, 186–191. Paris: ELRA.

Piater, J., Hoyoux, T., and Du, W. 2010. Video Analysis for Continuous Sign Language Recognition. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, ed. P. Dreuw, E. Efthimiou, T. Hanke, T. Johnston, G. Martinez Ruiz, and A. Schembri, 192–195. Paris: ELRA.

Pierrehumbert, J. 1980. *The phonetics and phonology of English intonation*. Unpublished PhD Thesis, MIT, Cambridge, MA.

Pierrehumbert, J. B., Beckman, M. E. and Ladd, D. R. 2000. Conceptual foundations of phonology as a laboratory science. In *Phonological knowledge. Conceptual and empirical issues*, ed. N. Burton-Roberts, P. Carr and G. Docherty, 273–303. Oxford: Oxford University Press.

Sandler, W. 1989. *Phonological representation of the sign: linearity and nonlinearity in American Sign Language*. Dordrecht: Foris.

Sandler, W. 1999a. Prosody in two natural language modalities. *Language and Speech* 42(2–3): 127–142.

Sandler, W. 1999b. The medium and the message: the prosodic interpretation of linguistic content in Israeli Sign Language. *Sign Language & Linguistics* 2(2): 187–215.

Sandler, W. 1999c. Cliticization and prosodic words in a sign language. In *Studies on the Phonological Word,* ed. T. A. Hall and U. Kleinhenz, 223–254. Amsterdam: Benjamins. (Current Studies in Linguistic Theory).

Sandler, W. in press. The uniformity and diversity of language: Evidence from sign language. *Lingua.*

Sandler, W. to appear. Prosody and syntax in sign languages. *Proceedings of the Philological Society*.

Sandler, W., and Lillo-Martin, D. 2006. *Sign language and linguistic universals*. Cambridge: Cambridge University Press.

Selkirk, E. 1984. *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.

Shattuck-Hufnagel, S., and Turk, A. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25(2): 193–247.

Shriberg, .E, Stolcke, A., Hakkani-Tur, D., and Tur G. 2000. Prosody-Based Automatic Segmentation of Speech into Sentences and Topics, *Speech Communication 32*(1–2), 127–154 (Special Issue on Accessing Information in Spoken Audio).

Stolcke, A., Shriberg, E., Bates, R., Ostendorf, M., Hakkani, D., Plauche, M., et al. 1998. Automatic Detection of Sentence Boundaries and Disfluencies Based on Recognized Words. *Proceedings of the International Conference on Spoken Language Processing*, vol. 5, 2247–2250. Sydney, Australia.

Sze, F. 2008. *Blinks and intonational phrasing in Hong Kong Sign Language*. In Quer, J. (ed.) Signs of the time. Selected papers from TISLR 2004 (pp. 83–107). Hamburg: Signum.

Vos, C. de, Kooij, E. van der, and Crasborn, O. 2009. Mixed signals: combining linguistic and affective functions of eyebrows in questions in Sign Language of the Netherlands. *Language and Speech,* 52–2/3: 315–339.

Wilbur, Ronnie B. 1990. An experimental investigation of stressed sign production. *International Journal of Sign Linguistics* 1 (1):41–59.

Wilbur, R. 1994. Eye blinks and ASL phrase structure. *Sign Language Studies*, 84, 221–40.

Wilbur, R. 1999. Stress in ASL: empirical evidence and linguistic issues. *Language and Speech* 42(2–3): 229–250.
Wilbur, R. 2000. Phonological and prosodic layering of nonmanuals in American Sign Language. In *The signs of language revisited: Festschrift for Ursula Bellugi and Edward Klima*, ed. H. Lane, and K. Emmorey, 213–241. Hillsdale, NJ: Lawrence Erlbaum.
Wilbur, R. B., and Patschke, C. 1998. Body leans and marking contrast in ASL. *Journal of Pragmatics* 30: 275–303.
Wilbur, R. B., and Zelaznik, H. N. 1997. *Kinematic correlates of stress and position in ASL*. Paper presented at the Annual Meeting of the Linguistic Society of America, Chicago IL.
Zeshan, U. 2006. *Interrogative and negative constructions in sign language.* Nijmegen: Ishara Press.