//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 246 – [246–265] 20.6.2005 6:05PM

15 Prosody

In spoken language, the flow of speech is not a steady unbroken stream, nor is it uttered monotonally. Instead, it is broken up into rhythmic chunks; some of its elements are more prominent than others; and it is characterized by meaningful excursions of pitch, called intonational tunes. This prosodic pattern is such an integral and systematic part of language that it enables newborn babies to notice when a speaker changes from one language to another, even when the segmental information is filtered out of the signal, leaving only prosodic properties (Mehler, Jusczyk, Lamberz, Halsted, Bertoncini, Amiel-Tison 1988). We intend to show that central properties of the prosodic system are common to languages in both modalities, spoken and signed.

Prosody is often thought of as an area of phonology, and that is understandable, under the broad definition of the term *phonology* proposed in Chapter 8: *phonology is the level of linguistic structure that organizes the medium through which language is transmitted.* This broader definition implies that the realm of phonology includes material above the word as well, encompassing, for example, the phrase, the utterance, or even the discourse. However, many linguists maintain that prosody comprises a separate component of the grammar, independent of other levels of linguistic analysis, because it has units and rules for their distribution and combination that are specific to the prosodic component. This prosodic component systematically interacts with all other components – with phonology, syntax, semantics, discourse, and pragmatics.

Research has motivated a hierarchy of prosodic constituents (Selkirk 1984, Nespor and Vogel 1986). The hierarchy shown in (1), adapted from Nespor and Vogel,[1] ranks prosodic constituents, from smallest to largest. These constituents exist alongside morphological and syntactic constituents, but are often not isomorphic with them. For example, *Jane's* in the sentence *Jane's singing* is two morphosyntactic words, *Jane* and *is*, but a single prosodic constituent – one syllable. The non-isomorphism of

[1] We have omitted the clitic group, which has proved controversial in spoken language research, and subsume it with the prosodic word.

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 247 – [246–265] 20.6.2005 6:05PM

prosodic constituents with morphosyntactic constituents is one of the strongest arguments that prosody is an independent component of the grammar, and cannot be relegated exclusively to an interaction between syntax and phonology without mediation.

(1) Prosodic hierarchy
mora < syllable < foot < prosodic word < phonological phrase < intonational phrase < phonological utterance

In earlier chapters, we have dealt with the prosodic unit, syllable, and with prosodic templates employed by the morphology. The discussion of prosody in the present chapter will move up to higher levels of the hierarchy: to the prosodic word, the phonological phrase, and the intonational phrase.

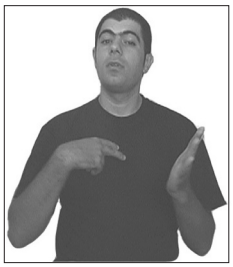
Our goal here is to describe the elements that go into the prosodic system in sign languages, and in this way to demonstrate that sign language grammar has a prosodic component. We'll also take a close look at the claim that sign language has intonation, expressed on the face. The discussion lays the groundwork for investigations of non-manual elements in connection with syntax, to be explored further in Unit 4. As intonation is part of prosody, and prosody is related to syntax, it is not surprising that some scholars have attributed to facial expression an explicitly syntactic role. We will suggest instead that the function and distribution of the relevant facial articulations correspond more closely to an intonational system.

15.1 The Prosodic Word

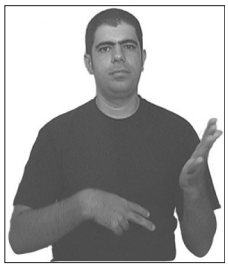
Just as morphemes and syllables are not isomorphic, morphosyntactic words and prosodic words are not always the same thing. It's very common, for example, for unstressed function words to group together with a nearby word prosodically, to form one prosodic word. Examples in English are contractions formed with auxiliaries like *Bill is* → *Bill's*. In many languages, like French, pronouns cliticize onto verbal hosts: *je aime* → *j'aime*. In these examples, the function word loses its syllable nucleus altogether, so that it would be ill-formed if pronounced as a full syllable; instead, it attaches prosodically to the host. In ISL, and apparently in ASL as well, there is also a distinction between morphosyntactic and prosodic words. In particular, pronouns can cliticize to lexical words, losing some of their phonological integrity, and forming one prosodic word together with the host. We will describe two such phenomena.

It has been noticed that in ASL, the handshape of pronouns can assimilate to that of a neighboring sign (e.g., Corina and Sandler 1993; Lillo-Martin 1986b). This phenomenon has also been reported for Quebec Sign Language (LSQ – Parisot 2000), as well as for Danish Sign

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 248 – [246–265] 20.6.2005 6:05PM



a. I (ISL)



b. I-READ (ISL)

Figure 15.1 Cliticized pronoun and host with handshape assimilation

Language (Engberg-Pedersen 1993). It occurs in ISL as well for personal, possessive, and deictic pronouns, and has been attributed to cliticization (Sandler 1999b, 1999c). In Figure 15.1, the pronoun 'I' takes on the handshape of the verb READ.

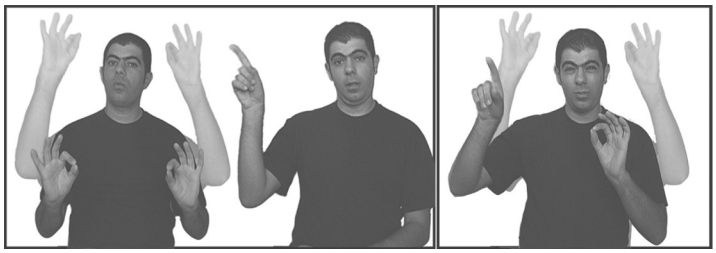
It is clear that cliticization of a function word to a host is involved, rather than a more general phonological assimilation rule, since it is always the case that the pronoun is the word that loses its underlying handshape, and never the neighboring full lexical word. Pronouns are typically unstressed in all languages, and are commonly cliticized. Given relative freedom of word order in ISL, assimilation can be either progressive or regressive; it is the lexical status of the words that determines the direction, and not the word-level phonology.

Another type of pronoun cliticization, coalescence, has also been observed in ISL (Nespor and Sandler 1999, Sandler 1999b, 1999c). In this type, in which the host is a two-handed h2-S sign, the host and pronoun reduce to a single syllable. Specifically, the non-dominant hand completes the full lexical sign, but the dominant hand only signs half of the host sign, and then signs the pronoun clitic while h2 completes the host. The sequence of two movements of the dominant hand is simultaneous with the single movement of the non-dominant hand, which creates a monosyllabic envelope for the newly formed prosodic word (see Chapter 14, example (1) for a definition of the sign language syllable). The coalescence process is illustrated in Figure 15.2. Figures 15.2a and b illustrate the signs SHOP and THERE uttered independently. In Figure 15c we see that the dominant hand switches from SHOP to THERE in "midstream," while the dominant hand simultaneously completes the sign SHOP.

The prosodic words formed by assimilation and coalescence are different from lexical words in some ways and similar to them in others. Let us examine the two cliticization processes more closely.

The handshape assimilation pictured in Figure 15.1 produces a single handshape specification for the host and clitic. The resulting form conforms to the Selected Finger Constraint (Chapter 14, Section 14.4),

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 249 – [246–265] 20.6.2005 6:05PM



a. SHOP b. IX 'there' c. SHOP-IX

Figure 15.2 Independent signs and cliticized host plus pronoun

rendering a surface form that bears a certain resemblance to a monomorphemic sign. However, this assimilatory process violates a constraint encoded in the hand configuration hierarchy presented in Chapter 10, Section 10.3, and results in an anomalous form. According to that hierarchy, if handshape assimilates, as in compounds, then orientation assimilates as well. This means that complex lexical words like compounds can have the same orientation with two different handshapes, through orientation assimilation. But it also means that assimilation can't result in the same handshape on the two member signs with two different orientations.[2] Yet in host plus clitic forms like I-READ, just such a form is created: handshape assimilates (→ one handshape), while orientation does not (→ two orientations).

The prosodic words formed by the other type of cliticization, coalescence, shown in Figure 15.2 (SHOP-IX), also serve to make the prosodic word so formed more like lexical words. Like most lexical words, the prosodic envelope provided by the non-dominant hand is monosyllabic (see Chapter 14, Section 14.5). But, like assimilation, this process is also non-structure preserving, in this case violating the Symmetry Condition on the behavior of the non-dominant hand (see Chapter 12, example (1)). In morphemes in which both hands move, the Symmetry Condition requires the two hands to have the same shape, path, and movement. It seems that this constraint holds not only for the morpheme, but for the syllable and the morphosyntactic word as well – we know of no counterexamples at any of these levels of structure. But as Figure 15.2 shows, in the prosodic word, the dominant hand changes its shape from  to  , while the non-dominant hand remains throughout, and the dominant hand changes its movement trajectory in mid-word, while the non-dominant hand completes the full movement of the host, SHOP.

[2] Monomorphemic, disyllabic lexical words, like ASL COOK, may have two different orientations of the hand with the same handshape, but these orientations are always mirror images of each other, e.g., supine and prone, and always signed with respect to the same place of articulation, here, h2. In cliticization, which is postlexical, these constraints are violated.

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 250 – [246–265] 20.6.2005 6:05PM

The explanation for these apparent violations is this: clitic formation is a postlexical process, occurring not in the lexicon – not through a process of word formation – but rather "later," at the point where words are strung together in sentences. Such a form is permitted because the assimilation process that creates it is postlexical and therefore may be non-structure preserving, as are many postlexical processes generally (Kiparsky 1982, 2002). In English, for example, a geminate [t:] may occur in *night time* although English does not have lexical length distinctions, and lexically prohibited consonant clusters occur freely in connected speech as well. In the same way, the prosodic words formed by handshape assimilation as in Figure 15.1 (I-READ) are non-structure preserving, but permitted postlexically.

The discussion leads us to three conclusions about these cliticized prosodic words in ISL: they are not isomorphic with morphosyntactic words; they are non-structure preserving; at the same time, they take on certain characteristics that make them more like lexical words.[3]

15.2 The Phonological Phrase

The constituent in the prosodic hierarchy that is above the Prosodic Word is the Phonological Phrase. According to Nespor and Vogel (1986), this unit is projected from syntactic phrases according to an algorithm that starts with a phrasal head belonging to a major lexical category: Nouns, Verbs, or Adjectives. Once constructed, phonological phrases can be restructured or merged, especially if they are short. Phonetically, this prosodic constituent is identifiable by minor rhythmic breaks. For example, the square brackets divide the following sentence into phonological phrases that would be likely to occur at a normal to slow rate of speech: *[The very tall] [construction worker] [carefully walked] [under the ladder]*. In English, the rhythmically prominent or strong position in the phonological phrase is the last stressed syllable in the phrase.[4]

To further support their claim that the prosodic hierarchy includes phonological phrases, Nespor and Vogel provide evidence that is independent of phonetic rhythmicity. The evidence consists of phonological rules in several languages that have the phonological phrase as their domain. For example, the Italian rule of *Raddopiamento Sintattico* (RS) applies only within phonological phrases. RS is an external sandhi rule

[3] A constraint competition analysis of these facts is suggested in Sandler (1999b), in which the Symmetry Condition, the Selected Finger Constraint, and Monosyllabicity are in competition, and the constraints have different rankings lexically and postlexically.

[4] According to Nespor and Vogel's theory, the direction in which phonological phrases are formed from the head, and the position of prominence within them, are predicted by the direction of syntactic recursivity – i.e., the word-order properties – of the language.

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 251 – [246–265] 20.6.2005 6:05PM

(i.e., a rule of assimilation across a word boundary). The rule geminates a consonant at the beginning of a word after a lexically stressed syllable, *p* and *l* after the stressed *é* of [*é* ***p***iú ***l***oquace] shown underlined and in boldface in (2a). But if a phonological phrase boundary separates the consonant from the stressed syllable, as in (2b), then RS does not apply and the consonant *p* of [*piú attentamente*] does not get lengthened, because the preceding stressed [*ó*] is in a different phonological phrase.[5]

(2) a. [Il tuo pappagallo]$_P$ [é **p**iú **l**oquace]$_P$ [del mio]$_P$
'[your parrot] [is more talkative] [than mine]'
b. [Guardó]$_p$ [piú attentamente]$_P$ [e vide]$_P$ [che era un pitone]$_P$
'[He looked] [more carefully] [and saw] [it was a python]'

A study of prosodic constituents in ISL (Nespor and Sandler 1999) has shown that there are phonological phrases in that language. This investigation coded and analyzed thirty elicited sentences, each signed by three signers, providing a corpus of ninety sentences. The rhythmic phonetic cues that mark the end of phonological phrases are: hold (freezing the signing hand or hands in their shape and position at the end of the sign), pause (relaxing the hands briefly between signs), or reiteration of the last sign. Nespor and Sandler's findings suggest that the end of the phrase is the prominent position in the phrase.

The study also discovered that the surface number of iterations of a sign is often determined by position in a phrase. The lexical representation of signs usually specifies a single iteration. But some signs have two, a distinction that may be contrastive (see Figure 13.3). In Chapter 13 on movement, the feature [restrained] was used for signs with two iterations, i.e., signs that are reduplicated once lexically.[6] If a sign that is underlyingly marked for the feature [restrained], i.e., reiterated once, occurs in a weak position in a phonological phrase (i.e., not phrase finally), it is often signed only once, losing the reiteration that occurs in citation form. However, if a sign that is underlyingly non-reduplicated occurs in the prosodically strong position at the end of the phrase, it often is reduplicated, even as many as three times (four iterations).[7] In an investigation of the phonology of the sign language of Quebec (LSQ), Miller (1996) finds that reduplication is influenced by prosodic context in that sign

[5] Phonological phrases with a small number of words in them can be restructured into a neighboring phrase (Nespor and Vogel 1986).

[6] Lexical reduplication and phrase-final reiteration are each distinct functionally and distributionally from morphological reduplication that occurs for example in temporal aspect inflection.

[7] Laura Downing points out that a citation form is in a phonological phrase, implying that all signs should potentially be reiterated in citation form. However, this is not the case: the lexical distinction between single and double movement is observed in citation form but may be neutralized by higher level prosodic cues. A possible explanation rests on Nespor and Sandler's (1999) suggestion that reiteration marks prominence. As prominence is a relative property, it requires the presence of more than one word in order to surface.

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 252 – [246–265] 20.6.2005 6:05PM

language as well, suggesting that this may be a general property of sign language prosody.[8]

Phonological phrases in spoken language are marked not only by phonetic cues but by phonological processes, like RS in Italian. In ISL as well, a rule of external sandhi provides further evidence for the domain in that language – it is a rule of Non-dominant Hand Spread (NHS). Specifically, if there is a two-handed sign within a phonological phrase, the non-dominant hand, can anticipate or perseverate the triggering sign by articulating the configuration and location of that sign. The spreading extends to the beginning and/or end of the phonological phrase, while the dominant hand articulates other signs. Crucially, the researchers found that NHS stopped at the boundary of the phonological phrase. Example (3) is divided into two intonational phrases (each labeled with an I subscript), the first containing three phonological phrases, and the second containing two phonological phrases. Each phonological phrase (labeled with φ) was marked by a characteristic phonetic cue – hold, pause, or reiteration of the last sign.

(3) [I-TELL HIM]$_{\varphi}$ [BAKE CAKE]$_{\varphi}$ [TASTY]$_{\varphi}$]$_{I}$ [[ONE FOR ME]$_{\varphi}$ [ONE FOR SISTER]$_{\varphi}$]$_{I}$
'I told him to bake a tasty cake, one for me and one for my sister.'

Unlike Italian RS, NHS does not involve sequential segments. Rather, the spread of the non-dominant hand from the triggering two-handed sign is simultaneous with the signing of other words by the dominant hand.

Figure 15.3 illustrates NHS in an excerpt from (3). The illustration shows the signs BAKE and CAKE with NHS. Also shown are the sign HIM in the phonological phrase that precedes BAKE CAKE, and the sign TASTY in the phonological phrase that follows it. In this example, the non-dominant hand from the sign BAKE spreads to the end of the phonological phrase by remaining in the same configuration as in the source sign, BAKE, throughout the next sign, CAKE, which is a one-handed sign. The end of the phonological phrase is marked by a hold – holding the hand in position at the end of the last sign. Precisely at the onset of the next phonological phrase, [TASTY]$_{\varphi}$, the sandhi stops, and the hand assumes a neutral shape. In the actual signing of this sequence, the change in the handshape and location between HIM and BAKE, and the rapid retraction of the fingers to a neutral position between CAKE and TASTY, are both perceptually salient (Sandler, in press).

NHS is an optional process, and does not always occur. Unlike hold, pause, and reiteration of the last sign, NHS is not a phonetic cue to a phonological phrase boundary. Instead, it is a rule of external sandhi

[8] If reiteration is influenced by prosodic position in ASL as it is in ISL and LSQ, this could explain why the underlying distinction between nouns and verbs in noun/verb pairs (Supalla and Newport 1978; Chapter 4, Section 4.1) was not discovered earlier.

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 253 – [246–265] 20.6.2005 6:05PM

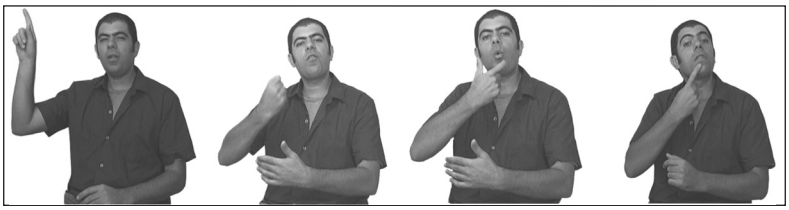


Figure 15.3 Non-dominant Hand Spread in the phonological phrase

which, by stopping at the phonological phrase boundary, is argued to provide further evidence for the existence of that constituent. What is important in the context of prosodic constituency is what the boundaries of NHS are when it does occur. In order to be convinced that the domain for the rule is the phonological phrase, we must rule out two other factors that might explain why the spread of the non-dominant hand stops where it does: the occurrence of another two-handed sign before or after the phonological phrase boundary, or the co-occurrence of a phonological phrase boundary with an intonational phrase boundary. When sentences with such co-occurrences were removed from consideration in two studies involving a total of seven signers and about eighty sentences, there were still no examples of h2 spreading beyond the phonological phrase boundary (Nespor and Sandler 1999, Sandler and Dachkovsky 2004). Therefore, the domain of the sandhi rule must be the phonological phrase boundary, and the rule provides evidence for the existence of this constituent.

15.3 The Intonational Phrase and intonation in sign language

At the next higher prosodic level, that of the Intonational Phrase, even more obvious prosodic breaks occur. Parentheticals, non-restrictive relative clauses, topicalizations and other extrapositions, vocatives, expletives, and tag questions form intonational phrases in many languages (Nespor and Vogel 1986). The salience of this break is due to clear rhythmic cues – they are typically separated by pauses and often by breaths – and also due to the distribution of intonational contours, to which we will return shortly. Sign languages have intonational phrases and intonational tunes as well, the latter expressed through facial expression. In ISL as in spoken languages, clear prosodic breaks were found for such syntactic constituents. For example, when the elicited sentences in (4) were signed, they were broken up into intonational phrases in the expected way.

(4) Intonational phrases in ISL

a. Parenthetical

$[\text{DOGS THOSE}]_I$ $[(\text{YOU})\ \text{KNOW}]_I$ $[\text{LIKE EAT COOKIES}]_I$

'Dogs, as you know, like cookies.'

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 254 – [246–265] 20.6.2005 6:05PM

b. Non-restrictive relative clause
[BOOKS HE WRITE PAST]$_I$ [I LIKE]$_I$ [DEPLETE]$_I$
'The books he wrote, which I like, are sold out.'

c. Right-dislocated element
[THEY TIRED]$_I$ [PLAYERS SOCCER]$_I$
'They're tired, the soccer players.'

d. Topic
[CAKE]$_I$ [I EAT-UP COMPLETELY]$_I$
'The cake, I ate up completely.'

Pronouncing the English translations of sentences (4a-d) above will give you a feel for the intonational phrase in spoken language, whose boundaries typically fall where the commas are. The breaks separating the intonational phrases in these ISL sentences of the Nespor and Sandler corpus had the following characteristics: they were marked by a change in head or body position and an across-the-board change in all aspects of facial expression. They were also optionally characterized by eyeblink. Eyeblinks often characterize phrase boundaries in American Sign Language as well (e.g., Baker and Padden 1978, Wilbur 1994b, Wilbur 1999a). The phrases described in the ASL studies appear to correspond to the Intonational Phrase (and not to the lower level Phonological Phrase).[9] This suggests that when eyeblinks occur, they are a reliable indicator of intonational phrase boundaries in sign languages generally, as breaths are in spoken language.

Comparison of the same sentences in different sign languages suggests that the change in head and body position together with facial expression at intonational phrase boundaries is common cross-linguistically (Sandler and Dachkovsky 2004). But one study indicates there may be some cross-linguistic variation in prosodic marking. Boyes Braem (1999) describes rhythmic side-to-side body sways for structuring certain kinds and levels of discourse, a cue not reported in other languages. Her work also shows that body sways of late learners differ from those of early learners on a range of measures, implying that the system is indeed linguistic.

In ASL and ISL, the ends of intonational phrases are prominent. In ISL, the last word in intonational phrases typically has more reiterations and larger signing than the last word in phonological phrases. In a study of prominence in American Sign Language, Wilbur and Zelaznik (1997) used an instrumental tracking device to determine prominence. They found that the final position in the intonational phrase was characterized by highest peak velocity, which they interpreted as prominence.

[9] For a comparison of methods and findings of Wilbur and those of Nespor and Sandler, see Sandler (1999a).

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 255 – [246–265] 20.6.2005 6:05PM

Intonational phrases in spoken languages are an important domain for intonational tunes (Pierrehumbert 1980). Intonational tunes impose a wide range of meanings on spoken utterances. In some languages, such as Hebrew, intonation may distinguish a declarative sentence from a yes/no question, as exemplified in (5) below. There is no syntactic difference between the two in these languages; only intonation distinguishes them.

(5) Hebrew intonational minimal pair

a. Yoni halax laxanut.
Yoni go-3rd-sg-m.-pst to-def-store
'Yoni went to the store.'

b. Yoni halax laxanut?
Yoni go-3rd-sg-m.-pst to-def-store
'Did Yoni go to the store?'

Within intonational phrases, the pitch accents fall on relatively prominent elements, and the boundary tones come at the edge, together forming the phrase's melody. The pitch accents and boundary tones themselves have meanings, and have been referred to as morphemes (e.g., Hayes and Lahiri 1991). This means that they have the dual function of delineating prosodic constituents and adding meaning to utterances. While dramatic pitch excursions tend to occur at intonational phrase boundaries, smaller changes at phonological phrase boundaries may also occur, effecting subtle nuances of meaning. Furthermore, intonational tunes in some languages have been analyzed as componential, building up complex meanings through sequences of meaningful tones and tone combinations.

An example from Bengali (Hayes and Lahiri 1991) given in (6) illustrates some of these properties. The L^* H_P L_I tune is the focus tune, consisting of an L^* pitch accent – that is, an accented low pitch – followed by an H (high) phonological phrase boundary tone and an L intonational phrase boundary tone. The focus tune has the effect of emphasizing the part of the sentence on which it falls. The succeeding H tone is a continuation rise, indicating that more information follows. As we can see here, phrasal tunes are componential.

(6)

[jodio ram [harlo,] P] I	(o k^hub b^halo k^helec^{h}ilo)
L* H_P L_I H_I	
'Although Ram lost,	(he very well played)'

We now return to sign language. The corpus examined in the prosody study of Nespor and Sandler, in addition to coding the behavior of the hands, coded each non-manual element of the face independently: eyebrows, eyes (upper and lower lids), cheeks, mouth, head position, similar to the system devised by Baker and Padden (1978). A (different-colored) line was drawn opposite the facial articulator label, and extended across the signs with which they co-occurred. An example is given in (7).

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 256 – [246–265] 20.6.2005 6:05PM

a. WRITE

b. INTERESTING

Figure 15.4 Complete change in non-manual markers in two adjacent intonational phrases in ISL

(7) Prosody coding (Nespor and Sandler 1999)

	[[book-there] P [he write] P]**I**	[[interesting] P]**I**
brows	up———	down———
eyes	squint———	droop———
cheeks		
mouth	‘O’———	down———
tongue		
head	tilt———	
mouthing	*‘book’*———	
‘interesting’–		
torso	lean———	
hold	=	
reduplication-1	×3	×4
pause		
speed		slow
size	big	big

The findings were consistent. The lines in that corpus were systematically discontinued at the intonational phrase boundary. In the example in (7), there are two phonological phrases in the first intonational phrase: BOOK THAT and HE WRITE, and one phonological phrase in the second intonational phrase, INTERESTING. The first IP is interpreted as the topic and the second as the comment (see also Rosenstein 2001). Although there is some small difference in non-manual articulation between the first two phonological phrases (in particular, there is a squint on the first phonological phrase only, and a non-neutral mouth shape only on the second), *all* facial configurations and the head and body positions change at the intonational phrase boundary in this example, and throughout the Nespor and Sandler corpus. This change is clearly indicated by the fact that all the lines on the coding sheet break between intonational phrases. Pictures of the two adjacent signs on either side of the intonational phrase boundary in this example, WRITE and INTERESTING, are given in Figure 15.4.

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 257 – [246–265] 20.6.2005 6:05PM

The fact that facial expressions coincide with intonational phrases is one of the motivations for the claim that they correspond to intonation, a claim we will support further in the next section.

15.4 Superarticulation: facial expression as intonation

It has long been known that facial expression and other non-manual markers play a significant linguistic role in sign languages.[10] Liddell (1978, 1980) presented the first detailed analysis of these markers in ASL. In that groundbreaking study, he demonstrated that yes/no questions, sentence topics, negated constituents, relative clauses, and other structures have characteristic facial expressions and head postures. He also examined the interaction between the scope of non-manuals and the scope of the constituent, finding that they were coextensive. His investigation provided evidence for word-order properties of ASL, to be presented in Chapter 18. Liddell also distinguished non-manual signals of emotional states such as surprise or anger from grammatical signals, pointing out that the emotional ones are more gradient, a claim confirmed in Baker-Shenk (1983). Analyses of the upper face/head and body positions in ASL showed that specific non-manual articulations mark questions, relative clauses, topics, conditionals, Wh-questions, and rhetorical questions (Baker and Padden 1978; Baker-Shenk 1983; Liddell 1978, 1980). Baker-Shenk (1983), Liddell (1986), Wilbur (1994b), Wilbur and Patschke (1999), and Bahan (1996) subsequently have identified functions for additional components. Together with all researchers since Liddell, we concur that non-manual signals are grammatically significant. But unlike Liddell and some recent researchers such as Neidle et al. (2000), we do not claim that facial expression is a direct reflection of syntax. Instead, we support the position taken by Reilly, McIntire, and Bellugi (1990b), and Wilbur (1991), that facial expression corresponds to intonation. We present evidence for this claim here, and take up the issue again in connection with WH-questions in Chapter 23.

In this discussion, we refer only to facial expression that corresponds to intonation, and not to other uses of facial expression such as lexical marking or adverbials, which were mentioned in Chapter 4.[11] To avoid the pitch-based label, "intonation," we call the intonational system of sign language *superarticulation*, and we use the term *superarticulatory arrays* for the combination of articulations corresponding to tunes (following Sandler 1999c).

[10] See, for example, Stokoe (1960), Baker and Padden (1978), Liddell (1980), Reilly, McIntire, and Bellugi (1990a), Nespor and Sandler (1999).

[11] We also exclude iconic mouth gestures (Sandler 2003) from our purview here, as they are argued to be the sign language equivalent to co-speech gesture, and therefore outside the formal linguistic system.

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 258 – [246–265] 20.6.2005 6:05PM

Studies of ISL have demonstrated that certain facial expressions consistently bear certain meanings in that language (Nespor and Sandler 1999, Sandler 1999c, Sandler and Dachkovsky 2004) as they do in ASL. As in spoken language intonation, superarticulatory meaning is broad, and gains more specific interpretation through its interaction with the meaning of the text with which it is associated. Furthermore, like the tones comprising intonational tunes in spoken languages, these superarticulaions may combine componentially with one another to give complex meanings. Finally, grammatical facial expressions in sign language can be distinguished from emotional facial expressions, just as linguistic and paralinguistic intonation in spoken language are distinguishable from each other. In the sections that follow, we will illustrate each of these characteristics.

15.4.1 Superaticulation is linguistic

Superarticulatory arrays similar to those that mark yes/no questions and WH-questions in ASL (Baker and Cokely 1980, Liddell 1980), in British Sign Language (Sutton-Spence and Woll 1999), Sign Language of the Netherlands (Coerts 1992), Danish Sign Language (Engberg-Pedersen, 1993), and many others, are found in ISL as well (Nespor and Sandler 1999, Sandler 1999c, Sandler and Dachkovsky 2004). Typical yes/no questions in ISL are marked by brow raise, wide eyes, and a forward head position – Action Units 1, 2, 5, and 57 in the Facial Action Coding System of Ekman and Friesen (1978). Typical WH-questions are marked by lowered brows (AU 4) and head forward (AU 57). A common systematic facial expression in ISL is a kind of squint used to mark shared information (lower lid contraction, AU 7). What appears to be the same superarticulation with the same interpretation is reported for Danish Sign Language (Engberg-Pedersen 1990). These superarticulatory arrays are illustrated in Figure 15.5

Finer grained superarticulatory arrays have also been found in ISL. For example, factual and counterfactual conditionals are distinguished by different superarticulatory arrays in that language, discussed in the next section.

The meanings attributed to these arrays are independent of the sentences they are articulated on, like the "morphemes" of spoken language intonation. This independence is exemplified by the fact that the superarticulations may combine with sentences whose syntax or lexical meaning do not match directly. For example, while prototypical WH-questions are accompanied by the articulatory array shown in Figure 15.5b, sentences that are WH-questions syntactically may be accompanied by different facial expressions if their pragmatic intent is not that of a WH-question. Conversely, the typical WH facial expression

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 259 – [246–265] 20.6.2005 6:05PM

a. Yes/no Q b. WH-Q c. shared information

Figure 15.5 Three grammatical superarticulatory arrays in ISL

may accompany strings that are not syntactically WH-questions, if the pragmatic intent is to ask a WH-question. We will provide examples of this dissociation in Chapter 23, where we present more specific arguments in support of our position that grammatical facial expressions in sign language are best understood as intonational "tunes." The next section demonstrates that meanings of superarticulatory arrays are built up componentially.

15.4.2 Superarticulation is componential

Coulter (1979) was among the first to identify component pieces of non-manual articulations, and to ascribe potential functions to them in ASL. Superarticulatory arrays can combine to form more complex arrays with more complex meanings in ISL as well (Nespor and Sandler 1999, Sandler 1999c). For example, in that language, a WH-question about information designated as shared is marked by a furrowed brow (Wh) plus squinted eyes (shared), shown in Figure 15.6.

Current work on ISL is showing that even arrays often interpreted holistically may be complex, i.e., that each individual action unit makes a contribution to meaning.[12] For example, brow raise in that language can co-occur with yes/no questions, factual conditionals, adverbial clauses, relative clauses and topics.[13] Dachkovsky (2004) proposes that brow raise conveys the general meaning of prediction. In these structures, it predicts that the first part of the utterance is going to be followed by some relevant information or consequence. In simple yes/no questions, the brow raise can be interpreted as predicting that a response will follow. Both yes/no and WH-questions are characterized by a forward head position (AU 57). Wilbur and Patschke (1998) suggest that the forward head position that

[12] A micro-component analysis of this sort might account for Reilly et al.'s (1990a) interesting finding that children master the non-manual components of the ASL conditional array bit by bit between the ages of 5 and 8, rather than all at once.

[13] Wilbur and Patschke (1999) and Wilbur (2000) isolate brow raise in ASL, a component in a large number of superarticulatory arrays, and provide an analysis predicting its occurrence on syntactic grounds. See Chapter 22 for a discussion.

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 260 – [246–265] 20.6.2005 6:05PM

Figure 15.6 WH-Q plus shared information: componential facial expression

occurs in these structures in ASL as well indicates inclusion of the addressee, a suggestion that is also compatible with the ISL data. Similarly, lower-lid squint (AU 7) in ISL occurs, often with various other superarticulations, on topics, relative clauses, parentheticals, and counterfactual conditionals, contributing to each array the same general meaning: designating the information so marked as shared between the interlocutors for the purpose of the utterance.

Dachkovsky (2004) analyzed counterfactual conditionals in ISL, sentences such as *If Ilan had more self-confidence, he would have passed his driver's test*. The first clause in such sentences is characterized by raised brows and lower-lid squint. Each superarticulation makes an independent contribution to the meaning, at once sharing with the addressee the knowledge that the event did not occur, and predicting the information in the next clause, i.e. what would have happened otherwise. This analysis demonstrates that the meanings associated with each action unit in the system are broad, gaining specificity and adding subtlety by combining with each other and with the meaning of the sentences they characterize. Intonation works like that.

15.4.3 The physical instantiation of intonation versus superarticulation

Superarticulation in sign language and intonation in spoken language, then, have three principal characteristics in common: (1) their functions, which are illocutionary, semantic, and pragmatic, (2) componentiality, and (3) the prosodic constituents that provide their domain: the phonological phrase and especially the intonational phrase. But the physical instantiation of tunes and arrays in each system is strikingly different, both in terms of the number of independent articulators that convey the tunes/arrays, and in terms of temporal distribution with respect to each other and to the co-occurring text.

In spoken language, the only intonational source is the vocal cords, which can vibrate at faster or slower frequencies resulting in higher or lower pitch. Intonational tunes are produced by changes in the frequency at which the vocal cords vibrate (the fundamental frequency or F_0) and are perceived as pitch excursions. Because only one articulator is involved, each tone is produced independently, and tunes consist of

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 261 – [246–265] 20.6.2005 6:05PM

sequences of these tones. Although a range of tones is implemented in this system phonetically, from a phonological point of view it is sufficient to distinguish only the two extremes, H (high) and L (low), and to account for the rest by rules of implementation (Pierrehumbert 1980). The sequences of individual H and L tones that comprise tunes typically are arranged at particular points of the text: on the stressed syllable of the head of focused constituents (see Selkirk 1984) and at phonological and intonational phrase boundaries (Beckman and Pierrehumbert 1986). While the tones occur simultaneously with particular syllables (hence the traditional term, "suprasegmental"), the syllables themselves are arranged in a sequence, and even tones falling on a single syllable also follow one another in a sequence.

In sign language, the physical system is quite different. There are several independent articulators – the brows, the upper and lower eyelids, the cheeks, the lips – and each articulator can perform more than one articulation. For example, the brows may rise or lower and the eyelids may contract or widen. The result is a system with a larger potential inventory of articulatory possibilities than spoken language intonation has. Whether or not this results in a richer intonational system is an empirical question. But one aspect of the physical instantiation is clearly different in the two modalities: the temporal instantiation of tones/superarticulations with respect to each other and in relation to the text. Instead of a linear sequence, the arrangement in sign language is simultaneous. Superarticulatory arrays typically co-occur with the entire prosodic constituent they characterize, and not only with the stressed syllable of a focused or boundary word as in spoken language, so that there is nothing in sign language that directly corresponds to pitch excursions, nor has the equivalent of a pitch accent been isolated. Furthermore, the superarticulations themselves within each array – i.e., whole "tunes" – co-occur simultaneously.

15.4.4 Grammaticization and language specificity of superarticulation

All humans use facial expression when they communicate, and many of these expressions are universal. How otherwise could we explain our ability to communicate attitudes and emotions to people with whom we have no common language? Or to empathize with people in news reports or characters in movies who speak different languages and come from cultures very different from our own? As communication through facial expression is universal, it should not be at all surprising that deaf people, for whom the visual medium is primary for communication, use facial expression as well. However, it should be clear by now that the use of facial expression in sign language is different. What is an idiosyncratic means of communication, one that may occur independently language or

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 262 – [246–265] 20.6.2005 6:05PM

as a supplement to it, has been grammaticized into a conventional system in sign language.

In an attempt to track the process of this grammaticization in ASL, Janzen argues that the yes/no question facial expression (raised brows and head tilted forward) evolved from a universal questioning expression, and that topic marking in that language evolved in turn from the yes/no non-manual configuration (Janzen 1998, 1999, Janzen and Shaffer 2002).

Evidence for the nonlinguistic source may be seen in a situation in which a person holding a drink makes eye contact with someone and then holds up the drink while raising his/her eyebrows. The interlocutor understands this to mean, 'Do you want a drink?' According to Janzen, this facial expression forms the basis of a conventional non-manual marker in ASL.

In Janzen's analysis, topics are seen as information from the interlocutors' shared world of experience, either new or old in the discourse. The meaning of the topic marker is associated with one of the meanings of yes/no questions: *Do you know x*? . . . . This overlap in meaning is claimed to underlie the further grammaticization of topics from yes/no questions. The phonetic difference between yes/no question and topic marking in ASL is the direction of head tilt: forward for yes/no questions and backward for topics. Janzen cites Wilbur and Patschke's (1998) explanation mentioned above: the forward head tilt on yes/no questions indicates inclusion of the addressee, while the backward head tilt found on the otherwise similar topic marking is interpreted as exclusion of the addressee.

An approach like Janzen's can explain why certain basic superarticulatory arrays, like yes/no questions, seem to be widespread across sign languages, and reveals another similarity between this system and spoken language intonation. In spoken language, for example, yes/no questions are nearly universally marked by a high tone (rising pitch) (Bolinger 1986, 1989). One explanation that has been offered for this universality is that certain pitch patterns evolved from purely biological factors and are now innate (Ohala 1984). Subsequently, such patterns are grammaticized (Gussenhoven 1999). A similar explanation for the grammaticization of universal facial expressions is suggested in Campbell, Woll, Benson, and Wallace (1999). In both modalities, the emotional or non-linguistic system exists alongside the grammatical intonation system. But if that is the case, how can we tell them apart?

15.4.5 Linguistic and nonlinguistic superarticulation

While differences between emotional or paralinguistic intonation and linguistic intonation in spoken language are not obvious, the two can

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 263 – [246–265] 20.6.2005 6:05PM

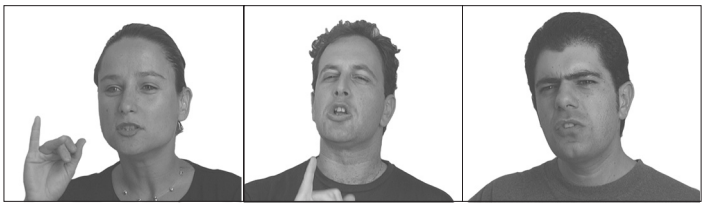


Figure 15.7 Grammatical "shared information squint" for three signers on the same phrase: 'Yossi's brother' in 'I just got a fax. Yossi's brother was killed in an accident.'

be distinguished. First, paralinguistic intonation reflects emotion and is therefore idiosyncratic, while linguistic intonational tunes have meaningful pragmatic functions and are conventionalized (Ladd 1996). Second, paralinguistic intonation is gradient. One can express more or less excitement, sadness, etc. by the degree of pitch excursion. Linguistic intonation, in contrast, is discrete and categorical (Gussenhoven 1999). The interpretation of a yes/no question melody does not depend on the mood of the asker, and small differences in the contour of a linguistic tune do not correspond to incremental changes in meaning or illocutionary force.

Signers use facial expression in both nonlinguistic and linguistic ways, another parallel with intonation of spoken language. Linguistic use of superarticulation is conventionalized, while paralinguistic non-manual articulations are idiosyncratic. Furthermore linguistic superarticulation is distributed across prosodic constituents discretely. Its onset is abrupt (Baker-Shenk 1983); and it does not begin before the constituent or continue after it (Baker-Shenk 1983, Sandler and Dachkovsky 2004). In all of these ways, linguistic superarticulation is different from nonlinguistic use of facial expression.

An utterance signed by three signers in a study of superarticulation in ISL illustrates the difference clearly (Sandler and Dachkovsky 2004). The elicited utterance is, *I just got a fax. Yossi's brother was killed in an accident*. In this utterance, the string *Yossi's brother* was characterized by the "shared information squint" (AU 6) superarticulation, shown in Figure 15.7. The onset and offset of this grammatical superarticulation was within three video frames of the onset and offset of the prosodic constituent established by manual signs in the string. (There are 25 frames per second of PAL videotape.)

But each signer produced a different array of facial articulations and head positions on the rest of the sentence, idiosyncratically reflecting different kinds, nuances, and intensities of emotion. Furthermore, these arrays, exemplified in Figure 15.8, were not synchronized with the signed text, instead beginning or ending up to 17 frames from the constituent boundary, and optionally crossing intonational phrase boundaries.

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 264 – [246–265] 20.6.2005 6:05PM

Figure 15.8 Idiosyncratic emotional facial expressions characterizing other parts of the same utterance

The distinction between linguistic and non-linguistic facial expression is supported by acquisition and aphasia studies as well. Anderson and Reilly (1998), Reilly and Bellugi (1996), and Reilly, McIntire, and Bellugi (1990a) provide evidence that affective (non-linguistic) and grammatical (linguistic) facial expression are acquired differently by children, and Corina, Bellugi, and Reilly (1999) report case studies on signers with brain lesions, which indicate that the linguistic and non-linguistic uses of facial expression are represented in different hemispheres of the brain.

15.5 Nonisomorphism

Although phonological and intonational phrases often correspond to syntactic constituents like the phrase or the clause, they are not strictly isomorphic with them. Rate of speech or signing, length of the syntactic constituent, and other factors have a clear influence on prosodic constituency. The two forces of syntax and rhythmicity in language don't always pull in the same direction, so that syntactic and prosodic constituents are not fully isomorphic, as shown in the excerpt from the children's story in (8).[14]

(8) Syntactic constituency vs. prosodic constituency
Syntactic: [This is [the cat that killed [the rat that ate [the malt]]]]
Prosodic: [This is the cat] [that killed the rat] [that ate the malt]

If the syllable were isomorphic with the morpheme, or the phonological phrase with the syntactic phrase, then there would be no need to posit a separate level of prosodic constituents in the grammar. Is such nonisomorphism found in sign language?

In Chapter 14, we presented evidence from sign language for nonisomorphism between the syllable, a prosodic unit, and the morpheme and the word, morphosyntactic units. In Section 15.1 above, we showed that the prosodic word and the morphosyntactic word are not one and the

[14] The example is taken from *The House that Jack Built*, a children's story compiled recursively into one long sentence.

//INTEGRAS/CUP/3-PAGINATION/SLL/2-FIRST_PROOF/3B2/0521482488C15.3D – 265 – [246–265] 20.6.2005 6:05PM

same. Evidence is presented in the Nespor and Sandler study for non-isomorphism at the phonological and intonational phrase levels of structure in Israeli Sign Language. Specifically, syntactic units which project their own prosodic constituents may be restructured and incorporated into nearby constituents if these syntactic units are short, or if the rate of signing is fast.

Especially compelling evidence of the dissociation between syntax and prosody can be found in the superarticulation system. This runs counter to a good deal of current research on ASL syntax, which relies on the assumption that facial expressions are explicit syntactic markers determined entirely by the syntax. We've provided a number of arguments here that the system is intonational instead, and will provide evidence of the dissociation between syntax on the one hand, and the rhythm and intonation of prosody on the other, in Chapter 23.

15.6 Summary, conclusion, and future research

Like all human behavior, communication among humans is subject to rhythmicity. It appears that language recruits this rhythmicity to interpret constituents that are not inherently rhythmic in nature, such as words, clauses, sentences, utterances, and higher discourse segments. This results in a prosodic system in which a hierarchy of prosodic constituents corresponds to morphosyntactic constituents to some extent, but not fully. Intonation superimposes itself upon this combination of syntactic structuring and rhythmic accentuation, systematically adding particular kinds of semantic information to the message. The existence of a prosodic system is a linguistic universal.

This chapter has served to substantiate the claim that prosody exists in sign language, and that it has certain key features in common with the spoken language system. But we have just begun to understand the structure of the sign language prosodic system in general, and of the intonational system in particular. How should intonational meanings best be characterized? How are they distributed and associated with the text? How are they interpreted? Is there an underlying system of intonational tunes that is altered by phonological and phonetic rule to produce the surface arrays, as has been demonstrated for spoken language (Pierrehumbert 1980; Gussenhoven 1984)? What are the prosodic differences across sign languages? Neither instrumental tracking and transcribing of this system in sign language, nor experimental work on its perception and interpretation, have yet been done. Our work is now cut out for us.