Gjert Kristoffersen
University of Bergen

The tone bearing unit in Swedish and Norwegian tonology

In most Norwegian and Swedish dialects lexical tone is employed to distinguish word meaning. Examples from Norwegian are *[1]selen*, 'the seal' (zool.) vs. *[2]selen*, 'the suspender', and *[1]tanken*, 'the tank' vs. *[2]tanken*, 'the thought'. The two 'accents' are commonly referred to as accent 1 and 2. Even if the contrast can be described as lexical, it is deeply embedded in the intonational and metrical systems of the languages. Thus the accentual melodies also signal the location of primary stress, and a necessary, although not sufficient condition for a phrase being focused is that it is headed by an accented syllable.

In this way, the Norwegian and Swedish tonal system can be said to combine features found in 'pure' tone languages, such as most Bantu languages, where tone has a more indirect relationship with metrical and intonational (pragmatic) prominence, and intonational features found in other Germanic languages, where use of F0 is intimately bound up with marking of such prominence.[1]

The phonetic realization of the contrast varies with dialect. In East Norwegian, accent 1 is marked by a low tone (henceforth L) on the (primary) stressed syllable, while accent 2 is marked by a high tone(henceforth H) on the stressed syllable, followed by a low, so that the result is a fall through the stressed syllable. A similar pattern is found in Stockholm Swedish, as described in e.g. Bruce (1977). One way of analyzing this contrast would be to focus on the difference in timing of the L which is common to both accents: Early L in accent 1, late L in accent 2. In other dialects, such as Skåne Swedish and several West and North Norwegian dialects, we find a similar timing difference, but connected with high tones. In these dialects early H marks accent 1, while late H marks accent 2.

At least since the advent of autosegmental phonology in the 1970s (see e.g. Goldsmith 1976, 1989) tone has been seen as relatively independent of the segmental makeup of a phonological representation, in the sense that a tonal melody remains the same irrespective of the length of its domain, and it will often be preserved when the segmental string undergoes reductions. This means that the tonal melodies and the segments can be seen as two partly independent strings of units, tones and segments, that must be synchronized in order to render a complete, phonological structure. Autosegmental phonology provided a proposal of how this synchronization could be represented, and accordingly, analyzed. The assumption that tones and segments can be assigned to different tiers connected by association lines has proved enormously influential during the last 25 years.

This way of analyzing phonological tone (and in fact the autosegmental model itself) was first developed in analyses of African tone languages. It was soon adopted in analyses of e.g. East Asian languages, and also in Scandinavian. Indeed, the first autosegmental analysis of Norwegian appeared as early as 1977 (Endresen 1977), and Gösta Bruce' seminal dissertation from the same year (Bruce 1977) also draws on this model. Later works on Norwegian that have assumed the autosegmental model include Lorentz (1984, 1995), Hognestad (1997) and Kristoffersen (1993, 2000). In all these works the syllable or the mora was assumed as the tone bearing unit (henceforth tbu) with which the tones that make up the accentual melodies, associate. The underlying assumption is that *every* tone in the accentual melody ideally should be associated with its own tbu. But the converse principle, that every syllable should be linked to a tone, is not consistently assumed in these analyses, even though it is part of the classic autosegmental theory.

But already in 1983 the contours of an alternative analysis emerged, in Bruce (1983). As can be seen from the following quote, only the timing between the melody and the so called stress group boundary should be seen as critical.

> The hypothesis states that segments or syllables do not carry specific F0 movements. Instead, an F0 pattern characterizes the stress group which may vary in phonetic composition and length. F0 is only constrained to coincide with segments at the stress group boundaries, where the timing is critical. (p. 235)

---

[1] For a more thorough introduction to tone in Norwegian and Swedish, see e.g. Bruce (1977), Bruce & Gårding (1978), Gårding (1977) and Kristoffersen (2000).

Since the implication is that it is the timing of the word accents with respect to the stressed syllable that is critical, the stressed syllable can be seen as the only tone bearing unit in a given stress group, to which the melody as a whole is associated, and hence synchronized.

This idea is further developed in Bruce (1987), from which the following quote is taken:

> In this kind of analysis, the stress group (- foot) and not individual syllables is taken as the domain for accentuation (tone bearer). According to this view, apart from the highly critical synchronization demand of word accent H* at the stress group boundary, there is no timing demand for the following word accent L or focal accent H with reference to particular syllables oar segments within the actual stress group. (p. 43)

The evidence for this analysis is a timing constancy of the turning points of the accent 2 HLHL-contour found by Bruce in data from Stockholm, which suggests that their timing may be independent of the segmental and syllabic makeup of the string. Although Bruce concedes that "the data are not incompatible with the traditional analysis with its tone - syllable associations", he claims that the fact that the HLH-parts of the contours show the same absolute timing when the number of syllables in the stress group is systematically varied, suggests that a unique association between the first tone and the stressed syllable should be assumed. The timing constancy can then be interpreted as an effect of factors having to do with concatenation on the tone tier which are independent of the syllabic makeup of the stress group.

In Gussenhoven & Bruce (1999: 237-39) this analysis is further refined. The tone bearing unit is here defined as the stressed syllable of a word. In the case of the accent 2 HLHL contour, where the first HL represents the word accent, the second H sentence accent and the final L a boundary tone, only the first H is associated with the syllabic string. The following L is concatenated with the H by means of a '+', and not associated with any syllable. The sentence accent H is likewise said to be floating. In compounds, the L of the word accent will associate with the final, secondary stress, thus forcing the floating H to be realized immediately after this stress. The typological analysis of Scandinavian in Riad (1998) is likewise based on the view that only stressed syllables are actually linked with a tone.

We are therefore faced with two views with respect to how tones are associated with the segmental string. The first is the orthodox, autosegmental view that all tones are

autosegmentally associated with a tone bearing unit, which must be either a syllable or a mora, subject to dialect variation. As noted already, most recent analyses of Norwegian are based on this premise. The second is the 'revisionist' view first proposed by Bruce that only stressed syllables should be counted as tone bearing units in Swedish and Norwegian, a view that entails that other tones will be left either floating or concatenated with the linked tone. They are thereby left indirectly associated with the segmental string.

The goal of the present article is not to argue that one of the views are less adequate than the other. Instead I shall suggest that both capture important aspects of Norwegian and Swedish tonology, and are therefore needed in any exhaustive analysis. My point of departure will be that Swedish and Norwegian can be said to belong to two language 'types' at the same time. They are on the one hand (marginal) tone languages, in that tones are used to distinguish word meaning. This feature groups them with the languages whose tonal grammar inspired the development of the autosegmental analysis. But Swedish and Norwegian also share with the other Germanic languages an intonational system where tonal events associated with strong metrical positions and phrase boundaries can be used to contextualize the content of a given utterance.

This distinction gives rise to two different, hierarchically related research questions. The first addresses the details of the lexical contrast itself, extracting it from its embedding within the intonational system. The other defines the complete intonational system as its primary focus, and analyzes the lexical contrast as part of this more comprehensive system.

It should be clear by now that I regard the 'brucian' approach where only the stressed syllable is analyzed as a tone bearer as a representative of the latter, holistic approach. It has important features in common with the autosegmental-metrical approach to intonation represented by e.g. D. Robert Ladd and Janet Pierrehumbert,[2] where the intonational contour is divided into smaller tunes which is further divided into pitch accents, boundary tones and phrase tones. These tunes may comprise more than one tone, just as the melodies analyzed in Bruce (1987) and Gussenhoven & Bruce (1999). The two approaches also have in common that in the unmarked case there is only one association line between the central 'starred' tone of a given tune and a metrically strong syllable. Thirdly, they have in common that they seek to account for the whole intonational pattern. All local events, including the lexical distinction, are seen as part of the intonational pattern as a whole.

---

[2] See Ladd (1996) for a comprehensive introduction. Gösta Bruce is indeed credited as one of the principal inspirational sources of this approach.

This approach has the virtue that it reveals the basic likeness between Swedish and Norwegian, and other Germanic languages with respect to intonation. Even if the Scandinavian intonational systems are perhaps more limited due to the constraints imposed by the lexical contrast, it reveals how the lexical contrast, by the fact that it can only be realized in connection with metrically strong syllables, can be seen as parasitic upon the more general constraints that intonational tones seek out metrically strong syllables (and boundaries) as the basic anchoring points for the intonational melody. I shall henceforth refer to this account as the IA (for Intonational Account)

But the question remains whether this tells the whole story. The claim of Gussenhoven & Bruce that the position of the unlinked tones can be predicted as soon as the central tone has been linked to the stressed syllable, seems to imply a positive answer to the question. Below I shall try to show that a closer look at how the tones that constitute the lexical contrast align with respect to the stressed syllable, in some dialects at least cannot be accounted for without assuming a more fine grained association pattern than the one assumed in the IA.

Let us first take another look at East Norwegian. Above, I noted that the lexical contrast could be analyzed in the following way: In accent 2, the melody is HL.[3] The H associates with the stressed syllable, and in domains with three syllables or more, the L will often coincide with the following unstressed syllable. Whether it should be seen as linked to this syllable or is concatenated with the preceding H can remain open here. The accent 1 melody can be seen as L (again disregarding the final H, cf. footnote 3). Being the initial (and only) tone of the melody, it will also associate with the stressed syllable. So far, the analysis is compatible with the IA. But all analyses of the East Norwegian pattern that I am aware of, shows that the accent 1 trough and the accent 2 peak are not synchronized with the stressed syllable in the same way. While the accent 2 peak is reached early in the stressed syllable, the accent 1 trough comes late. In a stressed syllable with a long vowel, the accent 2 peak will be reached within the first half of the vowel, while the accent 1 trough occurs within the second half of the vowel.

An exhaustive analysis of East Norwegian tonology must be able to account for this difference. Even if the East Norwegian pattern is compatible with the IA, in that both melodies are synchronized with a stressed syllable, the finer timing difference between the two accents cannot be captured directly within this model. What seems to be needed in a full

---

[3] I disregard the phrase final H, that is a combined boundary tone and focus marker, associated with the left edge of the accent phrase. This tone can be seen as purely intonational.

account of the lexical contrast is the mora as a timing unit, and therefore as a tbu on the more abstract, phonological level. This must be entered into the analysis in addition to the observation that the stressed syllable is the intonational tone bearing unit.[4]

Even more problematic for the IA are data that I have recently analyzed from the East Norwegian Oppdal dialect.[5] Contrary to the more southern variety of East Norwegian discussed above, the melodies in this dialect appear to be HL with respect to both accent 1 and accent 2. The H of both melodies is aligned with the left edge of the stressed syllable, the sole difference between the accents being the difference in timing of the following L that we have already seen in the southern variety analyzed above. Also in this dialect there may well exist a fixed timing between the two tones that comprise the melodies, as assumed in the IA, but that timing must be different for the two melodies, and cannot be accounted for by a mere '+' linking the two tones.

It is of course conceivable that we are dealing with two kinds of fixed timing, close and distal, for example, but if one bases an analysis on this, contrastive timing is made into a phonological prime that would blur the distinction between phonetics and phonology. The alternative is to derive the phonetic timing difference, whether it is fixed or not, from established phonological primes, and the tbu here comes in as the natural candidate. The Oppdal pattern can be accounted for by assuming that the accent 1 L associates with the second mora of the stressed syllable, while the accent 2 L skips this mora and associates with the first available mora following the stressed syllable.

Both examples we have discussed till now, concern the tonal organization within the bounds of the stressed syllable. Both suggest that it is not enough to assume that the leftmost (or starred) tone of the accentual tune is associated with the stressed syllable. But even if the analysis has to be modified in order to account for these facts, it is still possible that the IA hold with respect to the timing of the tone following the associated tone. This is in fact the central claim of the IA.

---

[4] Assuming the mora as the primary tbu in East Norwegian, however, cannot explain why the L of the accent 2 melody associates with the next syllable, and not with the second mora of the stressed syllable. Space limitations do not allow for a discussion of this problem here, but the point remains that at some level in the analysis, the different timing of the accent 2 H and the accent 1 L must be accounted for.

[5] The results have not yet been published, but were discussed in a talk I gave at the conference *Lexical tone and intonation in Germanic Languages*, held in Lillesand, Norway in June 2002. Oppdal is located near the southern

A problem with the data analyzed by Bruce, apart from the fact that it is limited to a few examples from one dialect, is that the syllable structure of the data was not varied systematically. While the number of syllables can be expected to influence the synchronization between syllable heads and tones when there are more tones than syllables, adding additional syllables when this state is reached is not expected to give any difference in timing on neither hypotheses. A better way of testing the IA hypothesis would therefore be to use material where there is enough syllables to accommodate every tone of the melody, and then vary the distance between syllable heads by using words with a different number of consonants between the vowels, so that the duration from vowel to vowel will vary. If the distance between two tones is governed by timing, we would expect the turning point of the second tone to fall earlier with respect to the second vowel in tokens where there is more segmental material between the vowels. At the same time we would expect the durational distance between the two tones to be invariant. If on the other hand the synchronization of the second tone is governed by the second vowel, we would expect covariation between some fixed point in the vowel and the timing of the tone as manifested by an F0 turning point.

A possible problem with this approach is that the two measures, timing of the F0 turning point from a preceding turning point, and timing of the F0 turning point with respect to a given vowel, are not independent, since the vowel with which the turning point is correlated may also exhibit some constant timing relationship with, say, the beginning of the stressed syllable. But one conceivable way of establishing the relative power of the two approaches with respect to explaining the timing of the F0 turning point of the second tone, would be for a given data set to compare the variation in the absolute timing of this turning point with respect to the preceding turning point, with the variation in its timing with respect to the midpoint of the second vowel. Using the standard deviation as a measure, the test that renders the lower amount of variation can be argued to suggest the correct solution.

I shall very briefly illustrate this approach be examining a set of accent 2 words as spoken by three speakers from the West Norwegian town of Bergen, all born in 1982.[6] The

---

border of the Trøndelag region in central Norway. Whether the pattern reported here generalizes to other dialects in the region, remains to be seen.

[6] The data is a small subset of a systematic body of data covering a major part of West Norwegian varieties, which has been collected in connection with the project *Norsk tonelagstypology* ('Typology of Norwegian Tonal Accents'), funded by the Norwegian Research Council from 2000 through 2002. See also www.hf.uib.no/Nordisk/ntt/.

accent 2 melody usually assumed for this dialect is LH, with the L synchronized with the early part of the stressed vowel, and H synchronized with the following vowel when there are three syllables or more, cf. Lorentz (1995). The test words are those cited in the following table:

Word/phrase	*Transcription*	*Gloss*	*Mean distance between end of V1 and beginning of V2*[7]	*N*
Hannemor	[ˀhɑn.nə.muːʁ]	female name	125 ms.	24
lærerforbundet	[ˀlæː.ʁɐʁ.fɔʁ…]	'teachers' union'	137 ms.	6
delingsmodellene	[ˀdeː.li**i**ŋ.smu…]	'the division models'	146 ms.	5
levende	[ˀleː.və.nə]	'living'	151 ms.	8
hjernemassene	[ˀjæʁ.nə.mɑs…]	'brain substance'	164 ms.	5
selvmålet	[ˀsɛl.m**o**ː.lə]	'the own goal'	187 ms.	6
hundre av de	[ˀhʉn.dʁ**ɑ**.di]	'hundred of the'	192 ms.	6

If the timing of the H is governed by the syllable being the tbu, the peak should coincide with the second vowel, which in each example is printed in bold for ease of reference. The first test that was run, took the correlation between the timing of the tonal peak with respect to the midpoint of the vowel. This is in other words a measure of to what extent the timing of the tonal peak can be inferred from the timing of the midpoint of the vowel. The first measure in the column marked R^2 *Peak to Vowel* gives the correlation across all tokens, while the following lines gives the result for each speaker. The standard deviations of the differences between peak and vowel midpoint are given in the next column, marked *St.dev.1*.

	R^2 Peak to Vowel	St.dev.1	Mean T1toT2 (ms.)	St.dev.2
Across all tokens	0,82	27,2	196,2	48,8
Speaker A	0,73	41,7	217,8	58,7
Speaker B	0,88	13,5	169,3	31,7
Speaker C	0,95	19,4	205,8	41,6

A corresponding correlation cannot be calculated with respect to the absolute timing between the trough of the initial tone and the peak of the high tone. But if the timing of the H is a constant we would expect the difference to be stable as measured in ms. The mean differences are given in the third column, marked *MeanT1toT2*. The standard deviation calculated over the differences in timing will again give us a measure of the variation. These measures are given in the fourth column, marked *Std.dev.2*.

If constant timing is the most important factor governing the placement of the H peak, we would expect the St.dev.2 to be smaller than St.dev.1. If the vowel is the most important factor, we would expect St.dev.1 to be smaller than St.dev.2. We see that it is the latter relationship that holds. This result suggests that at least for this dialect, the timing of the H of the accent 2 LH melody is indeed governed by the vowel more than by a constant timing constraint with respect to the first tone of the melody. This supports a phonological analysis where the H is autosegmentally linked to the second vowel and not only concatenated with the first tone.

These results must of course be seen as preliminary, and as a demonstration of a possible way of testing the IA-hypothesis more than a basis for stating firm conclusions. When more data have been analyzed within the typology project referred to in footnote 6, I intend to run similar tests on data from more dialects.

Bibliography

Bruce, Gösta. 1977. *Swedish Word Accents in Sentence Perspective*. Lund: CWK Gleerup.
Bruce, Gösta. 1983. "Accentuation and timing in Swedish". *Folia Linguistica* XVII/1-2, p. 221-238.
Bruce, Gösta. 1987. "How floating is focal accent?". In Gregersen, Kirsten; Basbøll, Hans (eds.). 1987. *Nordic Prosody IV*, p. 41-49. Odense: Odense University Press.
Bruce, Gösta; Gårding Eva. 1978. "A prosodic typology for Swedish dialects". In Gårding, Eva; Bruce, Gösta, Bannert, Robert (eds.). 1978. *Nordic Prosody*, p. 219-228. Lund: Dept. of Linguistics, Lund University.
Endresen, Rolf T. 1977. "An Alternative Theory of Stress and Tonemes in Eastern Norwegian". *Norsk Tidsskrift for Sprogvidenskap*, 31: 21-46.
Gårding, Eva. 1977. *The Scandinavian word accents*. Lund: CWK Gleerup.
Goldsmith, John A. 1976. *Autosegmental Phonology*. Bloomington: Indiana University Linguistics Club.
Goldsmith, John A. 1989. *Autosegmental & Metrical Phonology*. Oxford: Blackwell.

---

[7] In words with long vowels, the final half of the vowel has been included in the measure, in order to make the measures comparable.

Gussenhoven, Carlos; Bruce, Gösta. 1999. “Word prosody and intonation”. In Hulst, Harry van der (ed.). *Word Prosodic Systems in the Languages of Europe*, p. 233-272. Berlin, New York: Mouton de Gruyter.
Hognestad, Jan K. 1997. *Tonemer i en høytonedialekt*. Oslo: Det Norske Samlaget.
Kristoffersen, Gjert. 1993. “An Autosegmental Analysis of East Norwegian Pitch Accent”. In Granström, Björn; Nord , Lennart (eds.). 1993. *Nordic Prosody VI*, p. 109-122. Stockholm: Almquist & Wiksell.
Kristoffersen, Gjert. 2000. *The phonology of Norwegian*. Oxford: Oxford University Press.
Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.
Lorentz, Ove. 1984. “Stress and tone in an accent language”. In Elert, Claes-Christian; Johansson, Iréne; Strangert, Eva (eds.). 1984. *Nordic Prosody III*, p. 165-178. Umeå, Stockholm: Almquist & Wiksell.
Lorentz, Ove. 1995. “Tonal Prominence and Alignment”. *Phonology at Santa Cruz* 4: 39-56.
Riad, Tomas. 1998. “Towards a Scandinavian accent typology”. In Kehrein, Wolfgang; Wiese, Richard (eds.). 1998. *Phonology and Morphology of the Germanic Languages*. Tübingen: Max Niemeyer Verlag.