Pauses as a tool to ensure rhythmic wellformedness

Augustin Speyer

Department of Linguistics
University of Pennsylvania, Philadelphia, U.S.A.
speyer@babel.ling.upenn.edu

Abstract

Rhythmic wellformedness on the level of syllables and words and the mechanisms which are employed to ensure it are well known and researched on. The level of sentential accent is less often in the focus of such studies. In this paper I argue that rhythmic wellformedness plays an equally important role on the sentential level and that the preferred strategy to ensure it is the insertion of a pause (or lengthening of an already present pause), rather than the classical strategies of stress shift and destressing. This has to do with the role that focal accent plays for the semantic interpretation of an utterance. A reading experiment showed clearly that in clash cases the distance between the words in clash is increased by a pause; as we get the same effect in words in which a number of unstressed syllables intervenes between the clashing focal accents, it is clear that it is not simply stress clash resolution but resolution of a clash on a higher level of representation.

Index Terms: Prosody, pauses, rhythmic wellformedness, focal accent

10.21437/Interspeech.2006-256

1. Introduction

It has often been noticed that speakers make pauses between two accented phrases (e.g. Schmerling [1]). This has been interpreted as silent positions in the Metrical Grid (e.g. Selkirk [2]: 297ff.), and has also been linked to the presence of syntactic boundaries and thereby boundaries of units of the Prosodic Hierarchy (e.g. Byrd & Saltzmann [3]; to the Prosodic Hierarchy cf. Nespor & Vogel [4]). This goes so far that pauses are sometimes used as diagnostic for prosodic units such as phonological phrases or intonational phrases. This however cannot be right. First, the argument is at serious danger of becoming circular, if on the one hand one says 'a prosodic boundary is marked by pauses' and on the other hand 'wherever we find a pause must be a prosodic boundary'. Second, if we take the claim of Selkirk [2] and others seriously that there is some mapping between syntax and phonology such that any prosodic unit is ultimately derived from some syntactic structure, there should be no contrast between, say, (1) and (2), as both have pretty much the same underlying structure (sentential accent indicated by CAPITALS).

RICK they LIKE (1)

RICK _ THEY like (2)

In the case of (2) a pause is perceived between *Rick* and *they*, which is not necessarily the case in (1). If we believed that prosodic units such as Intonational Phrases are defined by pauses we would have to conclude that the structure of (1) would be one Intonation Phrase, whereas the structure of (2) would be two. Whereas one could argue that topicalized noun phrases, like Rick in both (1) and (2), form Intonational Phrases of their own, no matter whether the phrase boundary is marked by a pronounced break or not, the argument certainly would not hold of cases in which there is no topicalization but normal word order, and two focal accents happen to be on the subject and the verb. Compare (3) to (4).

RICK dashes at STUFF (3)

RICK _ DASHES at stuff (4)

There are two ways to cope with this contrast. One is to state that there is an obligatory Intonational Phrase boundary after any subject. The other is to assume that (3) and (4) have two different prosodic phrase structures. To the first possibility: To claim in such a case that there is an obligatory Intonational Phrase boundary between the subject and the rest of the sentence misses an important generalization, namely that it is usually not the case that the subject is of equal prominence than the nuclear accent of the utterance, but this is what we should expect if it formed an Intonational Phrase on its own obligatorily. To the secomd possibility: While it is possible to find a sufficiently strong syntactic boundary between the subject and the rest of the sentence (namely the boundary between SpecIP and I', in the framework of Chomsky [5]), the second claim would reckon that sometimes this boundary is translated into a prosodic boundary and sometimes not. This leaves a highly undesirable ambiguity, as there are no syntactic ways of predicting when a boundary has to be created. This leaves the burden to decide completely on the phonological or even phonetic side of the speech generation system, and this is tantamount to saying that phonology and/or phonetics creates a boundary whenever it is convenient. The idea that prosodic units mirror in some way syntactic structures is abandoned by that; what however would be the purpose of prosodic units if they do not give clues to the listener and/or language learner for the syntactic structure of the utterance?

The claim this paper makes is that the insertion of a pause is heavily dependent on rhythmic factors, and that the marking of prosodic units can be overridden by that. The evidence comes from experiments in which critical cases were contrasted.

2. The study

Six subjects (3 male, 3 female) were given the task to read out loud sentences that contained clash of focal accents; likewise minimal pairs of such sentences that did not involve a clash of focal accents. The participants read 'from scratch', i.e. no similar sentence was read loud to them beforehand in order not to bias them in their performance. For each sentence, an appropriate context was given. The factors that were varied in the experiment were the following:

- Topicalization or not: *RICK THEY like* vs. *RICK DASHES at things*
- Length of the first accented element: *RICK DASHES at things* vs. *ABERNATHY DASHES at things*.
- Length of the second accented element (only in topicalization cases: *RICK THEY like* vs. *RICK DENNIS respects*.

Especially the topicalization sentences involving full noun phrase subjects were judged as close to ungrammatical by some of the participants. Explanation and reading took the participants 10 minutes on average. They were undergraduate and graduate students with only limited linguistic background. They were rewarded for their time. The short target words were chosen such that the short first target word (= 1s), *Rick* or *Bob*, ended in a stop not homorganic to the coronal stop of the following target word. By the fact that two stops collide across the word boundaries we will get a short pause anyway; the prediction is that the pause should be longer if there is an accent clash. The long first target words (=1l), *Abernathy* and *Higginbotham* were chosen such that a complete foot follows the accented foot in the word; I could not find names of that phonological setup which ended in a stop. The second target word varied between pronoun (*they*; = 2s) and two-syllable name (*Dennis; Decker*; = 2m). In the case of non-topicalization, the second target word had to be a verb, and I chose verbs that begin at least approximately similar to the second target words in topicalization (namely *dashes* and *dawdles*; 2m). The following combinations were tested:

Top., 1s, 2s	Rick they like; Bob they like
No T., 1s, 2m	Rick dashes at things; Bob dawdles
Top., 1l, 2s	Abernathy they like; Higginb. th. like.
No T., 1l, 2m	Abern. dashes at things, Higg. dawdles
Top., 1s, 2m	Rick Dennis respects
Top., 1l, 2m	Abernathy Decker respects

The experiment showed these differences in pausing:

Type	Example	Pause mean	Difference
T1s2s	RICK THEY	0.2350	0.0750
	RICK they	0.1600	
	BOB THEY	0.2037	0.1137
	BOB they	0.0900	
N1s2m	RICK DASHES	0.2396	0.0798
	RICK dashes	0.1598	
	BOB DAWDLES	0.1437	0.0436
	BOB dashes	0.1001	
T1l2s	ABERN. THEY	0.0570	0.0226
	ABERN. they	0.0344	
	HIGG. THEY	0.0689	0.0471
	HIGG. they	0.0218	
N1l2m	ABERN. DASHES	0.0917	0.0318
	ABERN. dashes	0.0599	
	HIGG. DAWDLES	0.0459	0.0355
	HIGG. dashes	0.0104	
T1s2m	RICK DENNIS	0.2251	0.0702
	RICK Dennis	0.1549	
T1l2m	ABERN. DECKER	0.0964	0.0367
	ABERN. Decker	0.0597	

We see clearly from the data that there are quite large differences between the pause in the examples with accent clash as opposed to their counterparts without accent clash. The differences are not an artifact of calculating the respective means, in each single example there was a difference between the pause. The variation in the mean difference might have to do with the position of the target words in the utterance – all 'Bob' and 'Higginbotham'-examples were forming the second part of the contrastive utterance – and the fact that with the second contrastive pairs in such a setup it is not possible to derive real minimal pairs, so that there factors such as speaking rate etc. may play a role. Another factor which tends to shorten the difference is probably the length of the first target word; this could be explained by two things, speaking rate and compensation. About speaking rate: Although the two utterances *Rick they like* vs. *Abernathy they like* has the same number of words, the second utterance has twice as much syllables as the first one, as it is probable that the speaking rate depends not only on the number of words but also of syllables, we would expect a higher speaking rate there. About compensation: The clash in examples like *ABERNATHY THEY like* may be softened by the presence of the intervening extra unstressed foot. It is an accent clash, but at least it is not a stress clash (The distinction between (word, phrase) stress and (sentential) accent goes eventually back to Bolinger [6]). In the case of *RICK THEY like* it is both, an accent clash and a stress clash at the same time.

3. Discussion

We know from the literature about stress clash (e.g. Lieberman & Prince [7], Grabe & Warren [8]; cf. also Hayes [9]) that clash resolution usually is done by other means, viz. by stress shift or destressing. Stress shift is the well-known phenomenon that pushes the stress leftwards in a word that has underlying final stress if the following word begins with a stressed syllable (e.g. *thirTEEN*, but *THIRteen MEN*). Destressing is a process by which in clash cases one of the stresses in clash (usually the first one) looses some of its prominence so that it is not two equally strong stresses that stand in clash (*KalamaZOO NIGHTS* → *Kalamazoo NIGHTS*). Why do speakers fail to resort to this strategy in order to avoid accent clash?

An obvious objection might be that the phrases used in this study are so short that it there is no escape word etc. to which the accent could be shifted. But former experiments in which the phrases were more complex suggest that this cannot be the

reason: the participants failed to shift the stress even then, although escape words were available. The same goes for deaccenting, which would have been possible even in the examples used in the experiment, but the speakers did not make use of it.

The obvious answer to the question why speakers resort to pauses and not other strategies is that focal accent – as opposed to word stress and also probably to default sentence accent – is important for the semantic interpretation of the sentence. Focal accents – that is: any sentential accents that do not observe the default rules and thus are not predictable by using rules for stress assignment – is correlated with the semantic notion of focus which again can be represented in a very precise manner (cf. e.g. Rooth [10]). Since both the presence and the position of the focal accent are crucial for the interpretation of the sentence any strategy to resolve accent clash that messes with these two aspects is bound to fail, as it distorts the meaning of the sentence. In the classical cases of stress clash, which are within a phonological phrase and where the stresses are assigned by context-free word stress rules, shift and destressing are possible as the meaning of the words subject to shift / destressing are not affected by the position / presence of the stress.

A factor that both word stress and sentential accent are subject to is rhythmic wellformedness. It does not matter whether it is defined as 'absence of clashes' as in Lieberman & Prince [7] and subsequent literature, or as 'eurhythmy', following Hayes [11]. The crucial generalization, common to both approaches, is that two units of equal prominence must be divided by at least one unit of lesser prominence. This generalization, which I referred to as 'Trochaic Requirement' in Speyer [12], holds on all levels of metrical representation, between syllables, words and phrases.

It is planned to corroborate the results by perception studies – the recorded sentences with pause and manipulated sentences out of which the pause has been cut out or shortened are going to be presented to participants and they should judge which one is better; in another experiment participants are played such sentences both with normal pause and shortened pause and asked whether they perceive the pause – but this study could not be finished yet.

4. Conclusions

Focal accents, that is, prominences on utterance level which are determined semantically, are subject to rhythmic wellformedness, just as syllables within words and words within phrases: Two neighboring focal accents are not allowed. As other stress clash resolution strategies such as stress shift and deaccenting are not available, since they would distort the semantic representation of the sentence, another strategy, namely insertion of a pause, has to be used to restore rhythmic wellformedness in cases of clash.

5. Acknowledgements

This research has been done at the University of Pennsylvania as part of my PhD thesis. I would like to thank Jiahong Yuan for discussion of this work, as well as Eugene Buckley, Anthony Kroch, Marjorie Pak, and Jonathan Wright for their criticism and support. All remaining mistakes are my own.

6. References

[1] Schmerling, S., Aspects of English Sentence Stress, University of Texas Press, Austin, 1976.

[2] Selkirk, E.O., Phonology and Syntax, MIT Press, Cambridge, Ma., 1984.

[3] Byrd, D. and Saltzman, E., "Intragestural dynamics of multiple prosodic boundaries", J. Phon. 26, 1998, p. 173-199.

[4] Nespor, M. and Vogel, I., Prosodic Constituency, Reidel, Dordrecht, 1986.

[5] Chomsky, N., Barriers, MIT Press, Cambridge, Ma., 1986.

[6] Bolinger, D.L., "Accent is Predictable (if you're a Mind-reader)", Language 48, 1972, 633-644.

[7] Lieberman, M.Y. and Prince, A., "On Stress and Linguistic Rhythm", Linguistic Inquiry 8, 1977, 249-336.

[8] Grabe, E. and Warren, P., "Stress shift: do speakers do it or do listeners hear it?", in Connell, B. and Arvaniti, A. (eds.), Phonology and Phonetic Evidence, Cambridge University Press, Cambridge, p. 95-110.

[9] Hayes, B., Metrical Stress Theory, Univ. of Chicago Press, Chicago, 1995.

[10] Rooth, M., "A Theory of Focus Interpretation", Naural Language Semantics 1, 1992, p. 75-116.

[11] Hayes, B., "The Phonology of Rhythm in English", Linguistic Inquiry 15, 1984, p. 33-74.

[12] Speyer, A., "A prosodic factor for the decline in Topicalisation in English", in Kepser, S. and Reis, M. (eds.), Linguistic Evidence, Mouton de Gruyter, Berlin, p. 485-506.