Introduction to the symposium on the role of the word, foot, and syllable in speech production and perception: The role of the syllable in speech production

Alice Turk
Department of Theoretical and Applied Linguistics, University of Edinburgh
E-mail: turk@ling.ed.ac.uk

ABSTRACT

This paper introduces word-sized and smaller constituents that are proposed to play a role in speech production and perception: The Prosodic Word, Foot, Syllable, and Mora, and examines the role of the syllable in speech production in some detail. Nine possible roles for the syllable are discussed, along with supporting evidence. Although support for the general influence of the syllable on speech production is clear, it is not the case that all nine roles are supported unequivocally, nor is it clear to what extent the syllable plays similar roles in all languages of the world.

1. INTRODUCTION

The purpose of this session is three-fold: 1) to re-examine the role of the syllable in speech production and perception, 2) to raise awareness of other types of word-sized and smaller units that may act instead of or in addition to syllables in influencing speakers' productions or listeners' behavior, and 3) to discuss some of the challenges in determining the role these constituents play in speech.

I begin the session by introducing possible word-sized and smaller constituents that may play a role in production and perception, and discuss possible roles for the syllable in production. The first invited speaker, Uli Frauenfelder, will address the role of the syllable in speech perception, the second speaker, Katherine Demuth, will address the role of the foot in children's early utterances, and the third speaker, Linda Wheeldon, will discuss the role of the prosodic word in speech production planning. We do not discuss the role of the mora; possible roles of this constituent are likely to be similar to those of the syllable.

2. PROPOSED WORD-SIZED AND SMALLER CONSTITUENTS

The nested prosodic constituents similar in size and smaller than lexical words that have been proposed in current theories of prosodic phonology (e.g. [1]) are the Prosodic/Phonological Word, the (within-word) Foot, the Syllable, and the Mora. These constituents are thought to be nested underneath larger, phrasal constituents within the prosodic hierarchy, including the Minor Phrase/Accentual Phrase, Major Phrase/Intermediate Intonational Phrase, and the Full Intonational Phrase; for a review of these, see Shattuck-Hufnagel and Turk [2].

Brief definitions:

•Prosodic Word: According to Selkirk [3], the prosodic word is at least as large as a stem, but can contain monosyllabic function words, such as *going+ to* in the contracted form *gonna,* or in a connected speech version of *need him* [nidɪm].

•The within-word foot is delimited by lexically stressed syllables: e.g. [*cata*][*strophic*]

•The syllable contains a sonority peak called the nucleus (usually a vowel) and optional surrounding consonants. Preceding consonants form the syllable onset if and only if they form permissible word-onset clusters, final consonants form the syllable coda if and only if they do not form part of a following syllable's onset. The nucleus and coda together form the syllable rhyme.

•The mora is a sub-syllabic unit which consists of either 1) an onset consonant + short vowel, 2) short vowel, 3) short vowel + final consonant(s), or 4) coda consonant(s). Whether coda consonants form moras on their own varies from language to language, and depends on the patterning of C_0VCC_0 syllables with respect to unequivocally bi-moraic CVV syllables.

The nesting of prosodic constituents is thought to behave differently from the nesting of syntactic constituents: While syntactic trees may show many levels of recursivity (the nesting of one type of constituent under a constituent of the same type), recursivity, if it occurs, is thought to be minimal in prosodic hierarchies. For example, Selkirk [3] proposes that a Pwd can be nested under another Pwd, as in $[[penni]_W less]_W$. This type of structure contrasts with the prosodic structure for $[courageous]_W$, which is a single, un-nested prosodic word. The difference in structure can account for the fact that stress has not shifted from its location on the stem in *penniless*, but has in *courageous*. Another characteristic of prosodic hierarchies is that nodes of particular levels are parsed as far as possible into constituents one level down in the hierarchy, that is, a Prosodic Word such as $[Apa]_F[lachi]_F[cola]_F$is exhaustively parsed into three feet. Where this exhaustivity principle (Selkirk [3]) is violated, violations are again minimal, and tend to be of the type illustrated in the *penniless* example, where the

ISBN 1-876346-48-5 © 2003 UAB

Prosodic Word node attaches to a syllable node (*-less*) no more than two levels down from the Prosodic Word node.

3. POSSIBLE ROLES FOR THE SYLLABLE IN SPEECH PRODUCTION

1. Abstract units which predict phonotactic patterns.

Cross-linguistic evidence suggests that one of the roles of the syllable is to predict permissible sequences of segments word-initially and word-medially. For example, in English, the segment sequence /bd/ is disallowed word-initially, but allowed word-medially when surrounded by vowels: *bdom, but *abdomen*. This asymmetry can be explained if we assume that *bd is a constraint that applies to syllable-onsets only, and that /b/ in *abdomen* is a syllable coda. That this constraint refers to the syllable, and not just to the word, as it would if it were a word-onset constraint, is supported by the fact that e.g. /bd/ is not allowed in word-medial tri-consonantal sequences, e.g. *asbdom.

The language-specific nature of these constraints is evident: In Spanish, word-initial s-stop clusters are disallowed (e.g. [ɛspanɛr] is the standard pronunciation for the airline Spanair), whereas these clusters are permissible in other languages, e.g. English *span.*

An obvious question is how these constraints evolved. One possibility is that they evolved from perceptual requirements for listeners to recognise members of word-initial clusters. Indeed, an examination of cross-linguistic sonority profiles (cf. Clements [4]) shows that these are remarkably similar across languages: less sonorous elements tend to occur at syllable margins, and more sonorous elements tend to occur in syllable nuclei. The fact that syllables that begin with stop clusters are virtually absent cross-linguistically can be explained by the fact that cues to place of articulation of the first member of a two-stop cluster would be virtually absent utterance-initially.

Can we then replace these syllable-based constraints by more general perceptual or production constraints on phonotactic sequences? This may be the case for some languages, e.g. English, whose phonotactic constraints appear to be of a type that could be explained by general factors. Nevertheless, general perceptual or production factors cannot explain language-specific differences in phonotactic constraints, e.g. why *sstop clusters are disallowed in syllable-initial position in Spanish, but are perfectly legal elsewhere.

2. Are syllables elements that are serially ordered during speech production planning?

Research on speech errors showing that speakers often exchange sub-lexical elements suggests that lexical items are not stored as undivided wholes, but are made up of smaller elements (cf. References cited in [5]). In English, the elements which most often participate in exchanges are segment-sized or slightly larger; exchanges of whole syllables non-isomorphic to morphemes are rare (Shattuck-Hufnagel [5]). These findings suggest that the elements of which English lexical items are composed are not syllables, but are smaller elements, close in size to the single segment.

3. Part of a prosodic frame that segments and/or features/gestures are slotted into during speech production planning.

The fact that segments that participate in exchange errors come from the same structural position in the majority of cases suggests that serially ordered elements are slotted into a structural frame, and that information about the type of appropriate frame is present in the mental lexicon. For example, exchange errors such as *tar cowed* for *car towed,* where initial consonants exchange [5] are overwhelmingly common, where as *dar toak* for *car towed* (where an initial consonant has exchanged for a final consonant, and vice-versa) are virtually unattested. Although syllables do not appear to be the elements that are serially ordered during speech production planning, it is possible that they play a central role in the planning process as part of the prosodic frame that these segments slot into. However, the fact that the majority (60-90%) of exchanged segments are word-onset segments raises the possibility that the frame that segments are slotted into is based on units larger than syllables. Nevertheless, experimental evidence does suggest that onsets of English stressed syllables also participate in errors (Shattuck-Hufnagel [6]); it is the rarity of unequivocal exchange errors between subcomponents of unstressed syllables which makes the syllable view difficult to sustain in English. Error corpora from languages other than English are direly needed to evaluate the cross-linguistic status of the syllable in this role.

4. Constituents that are built as a final stage in the speech production planning process, after serial ordering has taken place.

Levelt et al. [7] argues on the basis of apparent syllabification across word boundaries, e.g. the apparent syllable-initial aspirated /t/ in Southern British English *escort us*, that information about syllabification is not stored in the mental lexicon, but that syllabification occurs as a relatively late stage in the phonological encoding process, when words are combined together to form larger units.

Costa and Sebastian-Galles [8] found that Spanish CVC syllables and CVCCV words presented 150 ms. After the presentation of a picture primed the production of CVCCV words with different segmental content as names of the target picture. For example, *salmo* and *bar* presented after the presentation of a target picture of a fly, caused faster production latencies for the target word *mosca* 'fly'. These results are consistent with the view that segments are structured into syllables at some stage in phonological encoding, although the stage at which this occurs isn't precisely clear. English results from Sevald et al. [9]

ISBN 1-876346-48-5 © 2003 UAB

which showed faster rates of production of structure-sharing, e.g. KEM TIL.FER and TIL TIL.FER pairs as compared to non-structure sharing KEMP TIL.FER and TILF TIL.FER pairs are often cited as evidence in support of the same view, but it is unclear whether slower speaking rates could be due in their entirety to additional phonemes in the non-structure sharing cases (e.g. the addition of /f/ in TILF TIL.FER as compared to TIL TIL.FER).

5. Constituents which influence the phonetic shape of utterances

Findings presented above suggest a possible role for the syllable during phonological encoding; findings discussed in this section suggest that the abstract structures proposed above are physically realized in the acoustic signal, although the way this realization manifests itself is still a matter of debate.

5a. Constituents whose edges are signalled by particular phonetic realisations (e.g. so-called syllable-conditioned allophones).

Many languages have characteristic realisations of phonemes that depend on position-in-syllable, e.g. syllable-final aspiration of /s/ in some varieties of Spanish (e.g. Tenerife, Nicaragua): *meses* [mɛseh], *mes* [mɛh], *escuela* [ehkwɛla], and glottalisation of syllable-final voiceless stops in some varieties of English: that [ð□a/], Atkins [aʔkɪnz]

In addition, Krakow [10] has shown that displacement magnitudes of supra-laryngeal gestures and intergestural timing are influenced by syllable structure: Patterns of velic and labial coordination differ qualitatively between e.g. *seem#ore* and *see#more*. While it could be the case that these patterns are influenced by word, rather than syllable structure, comparisons of /m/ in *helmet* vs. hell mitt, *Seymour* vs. see more (word-medial, syllable-initial vs. word-initial, syllable-initial), and /m/ in *seemly vs. seem Lee, homely vs. home Lee*, showed word-medial syllable-edge-related patterns that were quantitatively different, but qualitatively similar to the patterns of inter-gestural timing found at word-edges.

5b. Constituents whose component segments show articulatory cohesion

5.b.i. Constituents whose components share similar positions in distributions of phonetic parameters.

Campbell and Isard [11] propose that tautosyllabic segments in non-phrase-final syllables should show similar positions on distributions of segment durations. Their elasticity hypothesis is best supported for the shortest syllables in their corpus; the intermediate and longest syllables showed greater elasticity (magnitudes of lengthening in z-scores) for onsets and nuclei as compared to syllable codas. In addition, differences were found between syllables ending with voiced vs. voiceless stops, although these differences were most marked in word and phrase-final position.

A similar, more general idea is proposed in Fujimura [12], which proposes that tautosyllabic segments share phonetic strength (i.e., they share similar gestural magnitudes).

5.b.ii Constituents which influence patterns of inter-segment coordination

Browman and Goldstein [13] showed that word-onset consonant clusters (e.g. spl- in *pea#splats*, pl- in *pea#plats*, etc.) were more tightly coordinated to a following vowel than clusters which span word boundaries, e.g. s#pl in *peace#plats*. These results raise the possibility that the unit which predicts patterns of inter-segment coordination is the syllable. However, because the boundaries under investigation in these studies were word rather than syllable boundaries, crucial evidence for the syllable in this role is lacking.

5b.iii. Is the syllable the domain of coarticulation?

Early proposals that the syllable serves as a domain of coarticulation (e.g. Kozhevnikov and Chistovich [14]) were shown to be incorrect; coarticulation has been shown to extend over syllable, and even word boundaries. Nevertheless, the possibility remains that coarticulation is attenuated at syllable boundaries. However, Hardcastle and Byrd present evidence counter to this view [15][16]: They show that it isn't always the case that consonant clusters spanning word boundaries are less overlapped than onset or coda clusters. For example, #kl and #sk clusters appear to be less overlapped than k#l and s#k clusters.

5c. Stress and tone bearing units

Trubetzkoy proposed that the syllable is the bearer or tonal and accentual/stress features [17]. At the very least, the bearer of stress/accent, and tone must be as large or small as the syllable, as evidenced by minimal pairs such as 'digest vs. di'gest, and by polysyllabic words which bear distinctive tones on each syllable in tonal languages such as Mandarin Chinese. Nevertheless, the view that the bearer of these suprasegmental features is the syllable as opposed to the segment is difficult to establish empirically. Although Xu, Silverman and Pierrehumbert, Prieto et al and others show convincingly that all parts of syllables play a role in predicting the details of tonal alignment [18] [19][20], van Heuven found that listeners can reliably tell the difference between syllables with contrastive phrasal stress on a single segment (e.g. [be:n] vs. [bo:n]) as compared to words with contrastive phrasal stress on an entire syllable (e.g. [bo:n] vs. [ve:r]), although not all speakers appear to signal these differences. Acoustic analyses of the stimuli in his experiment showed that these words were distinguished primarily by F0 contour shape alignment properties, and not by segmental durational differences [21].

ISBN 1-876346-48-5 © 2003 UAB

The lack of segmental durational differences in different focus contexts raises the possibility that the durational correlates of phrasal stress could span a syllable-sized domain. However, Sluijter and van Heuven and Turk and White [22][23] show that durational correlates of stress can span more than a single syllable within disyllabic and trisyllabic words. On the other hand, what is still unclear from Turk and colleagues' results is whether the temporal extent of phrasal-stress related lengthening is due to the lengthening of a unit longer than a single syllable, or instead due to a combination of lengthening on the stressed syllable and word-final lengthening that tends only to occur on phrasally stressed words. Work in progress is designed to address this issue.

4. CONCLUSIONS

This paper identified nine possible roles for the syllable in speech production, but only a few seem to be unequivocally supported. Furthermore, the bulk of evidence in support of these roles comes from English and a few other well-studied languages such as Spanish; it is altogether unclear to what extent the syllable plays the same roles in other languages.

REFERENCES

[1] Selkirk, E.O. (1981). On prosodic structure and its to relation to syntactic structure, Nordic Prosody II, (pp. 111-140). Trondheim: TAPIR.

[2] Shattuck-Hufnagel, S., & Turk, A.E . (1996). A prosody tutorial for investigators of auditory sentence processing. Journal of Psycholinguistic Research, 25(2), 193-247.

[3] Selkirk, E.O. (1996). The prosodic structure of function words. In J.L. Morgan & K. Demuth (Eds.), Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition, (pp. 187-213). Mahwah, N.J.: Lawrence Erlbaum.

[4] Clements, G.N.(1990). The role of the sonority cycle in core syllabification. In J. Kingston & M.E. Beckman (Eds.), Papers in Laboratory Phonology I: Cambridge University Press.

[5] Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning, The Producion of Speech, (pp. 109-136). New York: Springer.

[6] Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. Cognition, 42, 213-259.

[7] Levelt, W.M.J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. Behavioral and Brain Sciences, 22, 1-75.

[8] Costa, A. & Sebastian-Gallés, N. (1998). Abstract phonological structure in language production: evidence from Spanish. Journal of Experimental Psychology: Learning, Memory, and Cognition, 24(4), 886-903.

[9] Sevald, C.A, Dell, G.S., & Cole, J.S. (1995). Syllable structure in speech production: Are syllables chunks or schemas? JML, 34, 807-820.

[10] Krakow, R.(1989). The Articulatory Organization of Syllables: A Kinematic Analysis of Labial and Velic Gestures. Unpublished Ph.D. thesis, Yale University.

[11] Campbell, W.N. & Isard, S.D. (1991). Segment durations in a syllable frame. J.Phon, 19, 37-47.

[12] Fujimura, O. (2000). The C/D model and prosodic control of articulatory behavior. Phonetica 57, 128-138.

[13] Browman, C.P., & Goldstein, L. (1988). Some notes on syllable structure in Articulatory Phonology. Phonetica, 45, 140-155.

[14] Kozhevnikov, V.A., & Chistovich, L.A.. (1966). Speech: Articulation and Perception. Joint Publ. Res. Serv., Washington, D.C.

[15] Hardcastle, W.J. (1985). Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. Speech Communication, 4, 247-263.

[16] Byrd, D. (1996). Influences on articulatory timing in consonant sequences. Journal of Phonetics, 24, 209-244.

[17] Trubetzkoy, N. (1939). Grundzüge der Phonologie. Travaux du cercle linguistique de Prague 7.

[18] Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. Phonetica, 55, 179-203.

[19] Silverman, K.E.A., & Pierrehumbert, J.B . (1990). The timing of prenuclear high accents in English. In J. Kingston & M.E. Beckman (Eds.), Papers in Laboratory Phonology I: Cambridge University Press.

[20] Prieto, P. , van Santen, J. & Hirschberg, J. (1995). Tonal alignment patterns in Spanish. Journal of Phonetics, 23, 429-451.

[21] Van Heuven (1994). What is the smallest prosodic domain? In P.A. Keating (Ed.), Papers in Laboratory Phonology III, Cambridge University Press.

[22] Sluijter, A.M.C., & Heuven, V.J. van. (1995). Effects of focus distribution, pitch accent and lexical stress on the temporal organization of syllables in Dutch. Phonetica, 52, 71-89.

[23] Turk, A.E., & White, L. (1999). Structural influences on accentual lengthening in English. Journal of Phonetics, 27, 171-206.