LANGUAGE AND COGNITIVE PROCESSES, 2002, *17* (1), 15–29

Length effects turn out to be syllable structure effects: Response to Roelofs (2002)

Julio Santiago

Departamento de Psicología Experimental y Fisiología del Comportamiento, Universidad de Granada, Spain

Donald G. MacKay

Department of Psychology, University of California, Los Angeles, USA

Alfonso Palma

Departamento de Psicología Experimental y Fisiología del Comportamiento, Universidad de Granada, Spain

Roelofs (2002) showed that by-item picture naming latencies in Santiago, MacKay, Palma, & Rho (2000) were linearly related to total number of segments across conditions, suggesting that structural effects of number of syllables and onset complexity might reflect a confound with phonological length. However, Roelofs failed to test the statistical reliability of this relationship with structural factors as covariates, and when we ran these and other analyses on our data, length effects were non-significant for two measures of length. We then discuss three additional Santiago et al. results favouring structural accounts but not length accounts, with number of syllables and onset complexity as the strongest structural factors, with a smaller effect (if any) of coda complexity, and no effect of vowel nucleus complexity. Finally, we argue that structure-sensitive phonological encoding mechanisms that may operate differently in different languages provide a better account of available evidence, including word production data that Roelofs (2002) claims support length accounts.

Requests for reprints should be addressed to Julio Santiago, Dept. de Psicología Experimental y Fisiología del Comportamiento, Facultad de Psicología, Universidad de Granada, Campus de Cartuja s/n, Granada, Spain. Email: santiago@ugr.es

http://www.tandf.co.uk/journals/pp/01690965.html DOI: 10.1080/01690960042000148

Syllable structure versus word length effects in language production are becoming important issues in evaluating alternative models of phonological encoding. Two types of models concern us here: Levelt-type models (e.g., Levelt, 1989, 1992; Levelt & Wheeldon, 1994; Levelt, Roelofs, & Meyer, 1999; Roelofs, 1997) and MacKay-type models (e.g., Burke, MacKay, Worthley & Wade, 1991; Dell, 1986; MacKay, 1987; but not Dell, Juliano, & Govindjee, 1993). A Levelt-type model accesses a word via phonological composition rules that create from left-to-right the phoneme sequence for the phonological syllables in the word. These phonological syllable strings serve to access a syllabary and retrieve phonetic syllable programs that get stored in an articulatory buffer. When phonetic programs for the entire word have been buffered, articulation can then begin.

Under this theory, onset times or vocal production latencies should increase with the total number of segments in a phonological word. Since syllables are created "on the fly" as the segment series unfolds from left to right, syllable structure and/or number of syllables per se should only affect onset latency indirectly due to concomitant increases in total number of segments in either the initial or final parts of the word.

MacKay-type models, hereafter, Node Structure theory (NST), postulate two types of phonological and phonetic components: hierarchically organised content nodes and sequence nodes. The phonological content nodes represent syllables and their constituents (onset nodes and rime nodes). An onset node represents the initial consonant or consonant cluster of the syllable. A rime node represents the vowel nucleus and coda of the syllable. A vowel nucleus node represents a simple or complex vowel. A coda node represents the final segment or segment cluster in the syllable.

Sequence nodes guide word production via tree-traversal processes that activate content nodes from top-to-bottom and left-to-right within each level of a content node hierarchy (see Santiago, Mackay, Palma, & Rho, 2000, for details). No articulatory buffering is required: Articulation starts as soon as the bottom-most left-most node in the phonetic or muscle movement hierarchy is activated. Subsequent segments are produced as their muscle movement nodes are activated in the tree-traversal process.

This model predicts that vocal production latency should increase with the number of tree-traversal operations (sequential decisions) that precede activation of the left-most bottom-most node in the hierarchy. This means that if phonetic or articulatory factors associated with the first speech sound are controlled across words, the main determinant of vocal latency should be complexity of the left branch of the hierarchical structure, which varies with the total number of syllables in a mono-morphemic word, and onset complexity of the first syllable (single consonant vs. consonant cluster). Total length of the words in segments should not affect latency directly but only indirectly due to concomitant increases in left-branch complexity. Finally, within a given hierarchical level, factors affecting encoding of initial parts of the word should affect latency more strongly than factors affecting final parts.

THE ROELOFS′ (2002) COMMENTARY

The main point of Roelofs' (2002) commentary and reanalysis of the Santiago et al. (2000) data is that length is difficult to disentangle from word/syllable structure and vice versa in a picture naming task. However, a careful reanalysis of our data indicates that length is not the sole, or even a major contributor to the Santiago et al. results.

Santiago et al. (2000) reported two experiments that provided evidence consistent with NST: In both experiments, number of syllables and onset complexity of the initial syllable (additively) increased latency, while complexity in vowel nucleus or coda of the initial syllable did not. In Experiment 1, mono- and bisyllabic words with simple vs. complex onsets were the responses in a picture naming task, with response latencies corrected (via subtraction from a base line condition) to factor out perceptual and conceptual differences across conditions (see Santiago et al., for details). Corrected latencies showed additive effects of onset complexity and number of syllables. In Experiment 2, the responses were monosyllabic words with simple vs. complex onsets, vowel nuclei, and codas (syllable structures were CV(C), CCVC, CVV(C), CVCC, CCVV(C), CCVCC).[1] Taking conditions CV(C) and CCVC together with CVV(C) and CCVV(C) in one partial factorial analysis, and with CVCC and CCVCC in a second, onset complexity had an effect on latency with no interactions, while neither vowel nucleus complexity nor coda complexity produced reliable effects or interactions. These results contradicted predictions derived from Levelt-type models, but supported the word-syllable structure predictions of MacKay-type models.

Against these conclusions, Roelofs (2002) argued that observed effects in Santiago et al. (2000) actually reflect differences in total number of segments across conditions. One argument was substantive in nature: An ANCOVA using length as covariate for by-items data in Santiago et al. eliminated the effect of onset complexity, and suggested a linear relation

[1] Because of the small number of words that could be found for each condition ($N = 6$), in some conditions Santiago et al. included words of the specified structure but varying in coda segments, as indicated by putting the coda among brackets in the condition label. By mistake, the CV(C) condition appeared as CVC, giving the misleading impression that all words in that condition had one coda segment (the exception is "tea").

between latency and length in both experiments (see Figures 1, 2, 3 and 4 in Roelofs). From this, Roelofs concluded that ‘‘the syllable structure effects turn out to be word length effects’’ (2002, p. 4).

Roelofs (2002) then suggested ways that length might account for other Santiago et al. (2000) results in four collateral arguments. One was that additive effects of onset complexity and number of syllables in Experiment 1 reflect additive increases in mean number of segments per word across the four conditions. A second argument was that the non-significant effect of coda complexity in Experiment 2 reflects lack of statistical power. A third was that non-differences in planned comparisons between CV(C) words vs. CVV(C) and CVCC words (taken together), and between CCVC words vs. CCVV(C) and CCVCC words (taken together), reflect the small differences in mean length (less than a segment) between the compared conditions. The fourth argument was that the non-significant effect of vowel nucleus complexity results from the unitary nature of complex nuclei (diphthongs) in phonological planning and in general (see Roelofs, 1997). Roelofs (2002) apparently ignored the evidence from experimentally induced speech errors that constituents of diphthongs (e.g., simple vowel plus glide) can act as independent units in speech production (see MacKay, 1978).

MEASURES OF LENGTH

Our first problem with Roelofs’ arguments was that we could not replicate his results for whatever length measure he applied to our stimuli in his analyses of our data. Roelofs did not describe his rules for measuring length, but in view of his fourth argument, we assumed that he intended to measure ‘‘phonological length’’, where both simple and complex vowel nuclei count as single segments. However, phonological length does not match his descriptions of our stimuli in Experiments 1–2. Roelofs (2002, p. 4) claimed that 9 of the 14 matched monosyllabic word pairs in Experiment 1 and 12 of the 14 matched bisyllabic word pairs were longer in the complex- than simple-onset condition, but this was true for 10 monosyllabic pairs (example exceptions are ‘‘bulb-bread’’, ‘‘fork-flock’’, and ‘‘ghost-grill’’) and 13 bisyllabic pairs (the sole exceptions was ‘‘castle-cradle’’).

Roelofs then claimed that 17 of the 18 matched word pairs in Experiment 2 were longer in the complex- than simple-onset condition, but this was only true for 10 of these pairs (example exceptions: ‘‘foam-fly’’, ‘‘pear-prow’’, and ‘‘cage-crow’’). Finally, Roelofs claimed a 0.4 segment mean length difference of between CCVC vs. CCVV(C)-CCVCC words in Experiment 2 but this mean was 0.25 segments by our calculations.

Perhaps then Roelofs measured "strict length" which is the same as phonological length except that diphthongs count as two segments. To be fair we therefore computed both strict and phonological length to test whether either measure showed clear and significant differences in the right direction for explaining the by-item latency data in Appendices A and C of Santiago et al. (2000) (thereby ensuring complete comparability with Roelofs' analyses).

Experiment 1 length differences

In Experiment 1, words differing in onset complexity and number of syllables differed significantly in both phonological length (onset complexity: $F(1,52) = 41.761$, MSe $= 0.465$, $p < .001$; number of syllables: $F(1,52) = 124.592$, MSe $= 0.465$, $p < .001$) and strict length (onset complexity: $F(1,52) = 34.848$, MSe $= 0.524$, $p < .001$; number of syllables: $F(1,52) = 114.481$, MSe $= 0.524$, $p < .001$), indicating length differences in the right direction to explain our observed latency differences. However, there were reliable interactions with both measures (phonological length: $F(1,52) = 4.64$, MSe $= 0.465$, $p = .035$; strict length: $F(1,52) = 4.9$, MSe $= 0.524$, $p = .031$): the difference in length between bisyllabic words with simple and complex onset was greater than between monosyllabic words with simple and complex onset. However, effects on latency were additive for onset complexity and number of syllables, so that length alone clearly cannot provide a full explanation for the latency results.

In a follow-up test, we compared effects of length on naming latencies in a between-items ANOVA with differing numbers of words in each length condition. Number of words and mean latency (with range) per length condition are shown for phonological and strict length in Table 1, showing a clear linear pattern except for the 8-segment words. However, this pattern was unreliable, for both phonological length ($F(6,49) = 1.833$, MSe $= 101006.24$, $p = .001$) and strict length ($F(5,50) = 1.968$, MSe $= 10132.87$, $p = .099$). To rule out the possibility that these analyses failed to reach significance due to noise introduced by conditions with very few items (length 2, 7, and 8 for the phonological length measure, and length 8 for the strict length measure), we reran the ANOVA excluding these conditions, but the effects remained non-significant (phonological length: $F(3,44) = 1.861$, MSe $= 10896.47$, $p < .150$; strict length: $F(4,49) = 2.394$, MSe $= 10288.23$, $p < .063$). Because phonological length, an essential measure for supporting Roelofs' criticisms and length hypotheses, did not have reliable effects on latency, despite our wide latency range (from 251 ms at length 3 to 362 ms at length 6), phonological length per se cannot explain the latency effects observed in Santiago et al. (2000, Experiment 1).

TABLE 1
Number of words, mean latency and latency range for the two types of length (phonological and strict) in Santiago et al. (2000, Experiment 1)

Phonological length	*N*	*Mean latency*	*Min. latency*	*Max. latency*
2	1	226	226	226
3	8	251	116	399
4	22	283	135	547
5	10	320	156	510
6	8	362	273	520
7	5	395	332	451
8	2	281	246	317
Strict length				
3	9	248	116	399
4	20	282	135	547
5	10	335	186	510
6	8	324	156	520
7	7	387	280	453
8	2	281	246	317

Curiously, however, effects of strict length approached significance, which motivated the ANCOVAs that we discuss later.

Experiment 2 length differences

Phonological and strict length did differ across conditions in Experiment 2 in a manner consistent with an effect of length on vocal latency. We analysed length differences across conditions using the same two partial factorial designs as Santiago et al., with conditions CV(C), CCVC, CVV(C), and CCVV(C) factored as an onset complexity $\times$ vowel nucleus complexity cross; and with conditions CV(C), CCVC, CVCC, and CCVCC factored as an onset $\times$ coda complexity cross. In the first partial factorial (onset $\times$ vowel nucleus complexity), phonological length varied reliably in the same direction as the onset complexity factor ($F(1,20) = 32.727$, MSe $=$ 0.183, $p < .001$), while it did not vary between words with simple versus complex nuclei because it counts complex nuclei as single segments ($F(1,20) = 3.636$, MSe $= 0.183, p = .07$; the trend was toward shorter words in the complex nucleus condition). Strict length showed differences across conditions in both the onset complexity factor ($F(1,20) = 32.727$, MSe $=$ 0.183, $p < .001$) and the vowel nucleus complexity factor ($F(1,20) = 14.545$, MSe $= 0.183$, $p = .001$). There were no interactions for phonological length ($F < 1$) or strict length (also $F < 1$). The second partial factorial (onset $\times$ coda complexity) only included words with a simple vowel (which collapses our phonological and strict length measures), and also gave differences across conditions in the right direction to explain our latency

TABLE 2
Number of words, mean latency and latency range for the two types of length (phonological and strict) in Santiago et al. (2000, Experiment 2)

Phonological length	*N*	*Mean latency*	*Min. latency*	*Max. latency*
2	3	240	125	404
3	12	249	107	393
4	15	295	58	492
5	6	347	166	433
Strict length				
2	1	404	404	404
3	7	180	107	260
4	19	298	58	492
5	9	321	160	433

results (onset complexity: $F(1,20) = 169$, MSe $= 0.041$, $p < .001$; coda complexity: $F(1,20) = 169$, MSe $= 0.041$, $p < .001$) with no interactions between the factors ($F(1,20) = 1$, MSe $= 0.041$, $p = .329$).

To actually test the hypothesis that length differences across conditions can explain the Santiago et al. (Experiment 2) results, we next analysed latencies as a function of phonological and strict length. Table 2 shows number of words in each length condition with their mean and range latencies for phonological and strict length. Once again, latency seemed to vary linearly with length (except for words of length 2 due to the outlier "tea": see Roelofs, 2002), but the phonological length effect did not approach statistical reliability ($F(3,32) = 1.25$, MSe $= 12216.67$, $p = .305$) and did not become significant in a second ANOVA that excluded the one condition with very few items (length 2, $F(2,30) = 1.704$, MSe $= 11613.71$, $p = .198$).

However, strict length showed a different pattern of results, reaching significance in the analyses including all conditions ($F(3,32) = 3.467$, MSe $= 10305.61$, $p = .027$) or excluding the length-2 condition ($F(2, 32) = 4.487$, MSe $= 10305.61$, $p = .019$). The question that remains is therefore whether these reliable differences in strict length can explain the Santiago et al. results (even though adopting a strict length measure would nullify Roelofs' length account for the non-effects of vowel nucleus complexity in Experiment 2).

STRICT LENGTH CANNOT EXPLAIN THE SANTIAGO ET AL. (2000) RESULTS: ANCOVA ANALYSES

To test whether strict length per se accounts for Experiment 1 results, we examined strict length effects via ANCOVAs that corrected for onset complexity and number of syllable effects. Strict length (conditions from 3 to 7 segments only) had no effect on latency either with onset complexity as covariate ($F(4,48) = 1.917$, $MSe = 10451.77$, $p = .122$) or with number of syllables as covariate ($F < 1$). These results indicate that length per se cannot explain effects of syllable and word structure in Santiago et al. (2000, Experiment 1).

Strict length also had a significant effect on latencies in Experiment 2 results. However, this effect disappeared when onset complexity was introduced as covariate ($F(2,31) = 1.901$, $MSe = 10333.64$, $p = .166$). In contrast, the effect of strict length persisted (although to different extents) when the covariate was vowel nucleus complexity ($F(2,31) = 4.562$, $MSe = 10410.01$, $p = .018$) or coda complexity ($F(2,31) = 3.520$, $MSe = 10636.47$, $p = .041$). This suggests that the seeming effects of strict length reflect correlations between strict length and three structural factors (onset, vowel nucleus, and coda complexity), such that onset complexity exerts the strongest influence in producing this seeming effect, followed by coda complexity and finally vowel nucleus complexity. This conclusion fully agrees with that of Santiago et al. (2000, Experiment 2).

To summarise, Santiago et al. (2000) found effects of structural factors on latency, and strict length (but not phonological length) differed across conditions in a way that could account for these effects. However, strict length cannot account for these effects since it was unrelated to latency when structural factors were partialled out. This suggests that Figures 1 and 2 in Roelofs provide a misleading visual impression produced by correlations between length and structural factors.

DISCUSSION AND CONCLUSIONS

Roelofs (2002) suggested that structural effects (onset complexity and number of syllables) in Santiago et al. (2000) were due to a confound with phonological length, a measure that counts complex vowels as single segments. The basis for this conclusion was that structural effects disappear in Santiago et al.'s by-item data if phonological length is partialled out in an ANCOVA. However, analogous ANCOVAs with phonological and strict length as covariates indicated that Roelofs' length effects were by-products of structural effects.

So do we just call it a stand-off? Three additional results suggest not. First, phonological and strict length interacted with onset complexity and number of syllables in Santiago et al. (Experiment 1) whereas no corresponding interactions occurred for latency: Onset complexity and number of syllables had additive effects on latency. This contrast suggests that phonological and/or strict length cannot fully explain effects of onset complexity and number of syllables on latency. Second, although superficial analyses suggested a linear relation between latency and phonological and/or strict length in Experiments 1 and 2, phonological length neither reached nor approached significance in predicting latencies, even for length conditions with a reasonably large number of items and the strongest linear relation with latency. Third, strict length did reach significance in restricted Experiment 1 analyses and in both restricted and unrestricted Experiment 2 analyses, but these effects disappeared with structural factors partialled out in covariate analyses. However, it is important to note that only the onset complexity covariate succeeded in eliminating the strict length effect in Experiment 2: Vowel nucleus or coda complexity reduced without eliminating this length effect. If the non-effect of length with a structural factor as covariate is due only to the high correlation of length and the structural factor, then all three structural factors should have eliminated the length effect to the same extent, because all three are equally correlated with strict length. Because only onset complexity eliminated the length effect, structural factors must be doing the real work in latency effects, albeit to different degrees, vowel nucleus complexity with no effect, coda complexity with a weak (and unproven) effect, and onset complexity with the strongest effect.

The preceding only holds, of course, for the strict length measure, which was surely not the measure that Roelofs had in mind because it nullifies his account of why vowel nucleus complexity had no effect in Santiago et al. (Experiment 2). The fact that phonological length did not reliably affect latencies is therefore important, indicating that neither phonological nor strict length alone affects vocal latency, but only as a by-product of structural factors. This conclusion counters Levelt-type models such as WEAVER++ (Roelofs, 1997), and favours MacKay-type models (MacKay, 1987).

Santiago et al. (2000) chose not to control for total length in their experiments for two reasons: the small number of picturable English words with the necessary structural characteristics, and failures of published studies (e.g., Bachoud-Lévi, Dupoux, Cohen, & Mehler, 1998; Levelt & Wheeldon, 1994) to find effects of length per se on vocal latency. However, Roelofs (2002) suggests that clear and significant length effects do exist in the literature and that published failures to observe these effects are readily explained. We will argue that his first assertion is wrong: Evidence for length effects in the literature is weak or non-existent, and the weak evidence may reflect structural rather than length effects. We will also argue that Roelofs' proposed account of conditions under which length effects occur cannot explain several basic facts and cannot fully explain the pattern of data in the literature. Like Roelofs (2002), we will not address the full range of relevant evidence in this brief reply, but will only discuss data from word production tasks such as picture naming and implicit priming.

Roelofs (2002) claims that Roelofs (1998) and Roelofs and Meyer (1998) reported length effects. This is a surprising claim. First, although participants in Roelofs and Meyer produced words that differed in number of segments and syllables, Roelofs and Meyer did not report a length effect, either for implicit primes or for target words. Also, even if these participants did produce words of differing lengths with different latencies, the between-item variation of length in Roelofs and Meyer prevents firm conclusion as to the basis for this hypothetical effect: An implicit priming task where the word responses are just-learned paired associates only allows clear conclusions for within-item differences involving, e.g., the same item in homogeneous vs. heterogeneous blocks. Adding to these difficulties, the main determinant of latency in this task (the strength of the cue-response link) was free to vary across stimuli in these studies (see Santiago, 2000, for further discussion of implicit priming tasks).

Second, neither Roelofs (1998) nor Roelofs and Meyer (1998) are strictly relevant because unprepared rather than primed or prepared latency is the issue at hand. Roelofs (1998), as Meyer (1990, 1991) before, found that priming magnitude varies with how many initial segments the prime and target response share in an implicit priming paradigm: Meyer (1990, 1991) made this point for initial segments shared within the first (and only) phonological word, and Roelofs (1998) made the same point for initial segments that extended into a second phonological word. The only difference was that Meyer (1991) observed the same amount of facilitation for CV and CVC primes for response words with a shared first syllable, suggesting that participants were preparing syllable-sized units, while Roelofs (1998) found that the total number of shared segments was all that mattered. Ferrand, Segui and Grainger (1996) and Schiller (1998) found similar results for masked priming in picture naming.

However, length effects in *priming* word production may be different from length effects in word production per se. In priming, what matters is the amount of prime-target overlap, whereas in word production what matters is the length of the target itself. Although a post hoc connection between the two kinds of effects can be suggested, further research is needed before empirical identity can be claimed. If length effects in priming and word production per se really are identical, then (a) we should expect different vocal latencies for targets of different length; (b) shorter latencies as more initial segments are primed, and, (c) fully prepared reaction time will be identical for words of different length. Meyer (1990, 1991) and Roelofs (1998) have only shown that (b) may be true. Evidence for (a) and (c) comes from studies that manipulate target length in both standard naming and fully prepared naming, as in Eriksen, Pollack, and Montague (1970), Klapp, Anderson, & Berrian (1973), Santiago (1997), and Santiago et al. (2000).

However, several facts complicate the picture. Up to now, no direct evidence for a relation between (a–c) and (b) has been presented. Moreover, other studies, e.g. Levelt and Wheeldon (1994) and Bachoud-Lévi et al. (1998), provide evidence against (a), and Bachoud-Lévi et al. (1998) have criticised the early work of Klapp et al. Finally, Sternberg, Monsell, Knoll and Wright (1978) provide evidence against (c), reporting length effects on latencies for fully prepared responses. The whole issue is clearly controversial, and given the differences in manipulated variables and experimental paradigms, assuming identity of underlying processes seems premature.

Moreover, the structural hypothesis actually provides a better account of target length effects in the extant literature than the length hypothesis. Klapp et al. (1973) controlled for total length in segments, ruling out a length account of their structural (number of syllables) effects. Unfortunately, however, Klapp et al. did not control for frequency-familiarity, which may have contaminated their results (see Bachoud-Lévi et al., 1998). The effects of length on prepared naming in Sternberg et al. (1978) may likewise reflect structural factors (number of stress groups and syllables within stress groups), together with memory search processes, but not the total number of segments within each stress group (although this factor was not explicitly controlled). Finally, as we have shown, structural factors such as number of syllables and onset complexity provide a better account of the Santiago et al. (2000) data than length in segments. Given the problems of prior studies, and the negative results of others such as Bachoud-Lévi et al. (1998), our data clearly constitute the best evidence to date for structural effects on vocal production latency.

Finally, Roelofs (2002) argues that the failures to find effects of length per se in the literature, particularly Levelt & Wheeldon (1994) and Bachoud-Lévi et al. (1998), can be explained as reflecting variability in the criterion to start pronunciation. This criterion would apply at a phonetic level, after preparation of the phonological word. The idea is that based on strategic factors, participants may decide to start speaking as soon as they have accessed and buffered the first phonetic syllable program of the word or they may wait until they finish accessing and buffering subsequent phonetic syllables. Roelofs (2002) bases this suggestion on unpublished data of Meyer, Levelt, and Roelofs (in preparation) who claimed a number-of-syllables effect in replicating Bachoud-Lévi et al. (1998), but only when the conditions were blocked with monosyllables in one block and disyllables in another or only when participants received feedback as to their performance.

This suggestion is problematic on many different dimensions. We consider first the problems of a variable criterion for explaining prior data, and then for explaining some new, as yet unpublished, data. Finally, we raise some issues as to how a variable criterion might work from a theoretical point of view.

Roelofs (2002) suggests that a variable criterion may explain why Levelt and Wheeldon (1994, Experiment 4) failed to find a length effect for words with identical initial syllable but different second syllable, which varied in length from 3 to 5 segments. This suggestion is problematic unless it explains what changed from Experiments 1–3 to Experiment 4 in Levelt and Wheeldon (1994). Experiments 1–3 reported a second syllable frequency effect in producing bisyllabic words, allegedly due to frequency-sensitive access of the second phonetic syllable before pronunciation onset for the word. If the whole word must be buffered before beginning to pronounce its first syllable, second syllable frequency will clearly have the strongest impact on latency, and participants must have been using a complete word criterion to start pronunciation. But since the experiments were identical in procedural details, including general characteristics of the materials, it is unclear why participants in Experiment 4 would decide to set a different onset criterion from participants in Experiments 1–3.

A similar problem arises when Roelofs (2002) explains the differing result of Bachoud-Lévi et al. (1998) versus Santiago et al. (2000) by suggesting that participants used an initial criterion in Bachoud-Lévi et al. versus a complete word criterion in Santiago et al.: What factor or factors shifted the criterion? One possibility concerns the ease and simplicity of the Bachoud-Lévi et al. task relative to the Santiago et al. task. Participants in Bachoud-Lévi et al. only had to name the object as soon as it appeared, using whatever name came to mind first, whereas participants in Santiago et al. had to produce pre-specified names, following an initial training block, resulting in much longer uncorrected latencies than in Bachoud-Lévi et al. Perhaps the more difficult conditions in Santiago et al. encouraged participants to start pronunciation only after monitoring the complete word. This being the case, however, the initial criterion must be the routine criterion for everyday speech production, with the complete word criterion reserved for special occasions, an important modification to Levelt-type theories suggested by Bachoud-Lévi et al. (1998).

However, new (unpublished) data from the Bachoud-Lévi and Santiago laboratories have suggested cross-language differences that cannot simply reflect changes in onset criteria. When English speakers received extensive over-training in the Santiago et al. (2000) procedure, that reduced their correct response latencies to the range of Bachoud-Lévi et al. (1998), they still showed onset complexity and number of syllable effects. However, Spanish speakers did not show a number of syllables effect either with extensive over-training or with the original Santiago et al. (2000) conditions. And French speakers failed to show a number of syllables effect with extensive over-training or with further increases in task difficulty. It is difficult to see how a variable criterion to start pronunciation (which would work in basically the same way across languages) can account for these cross-linguistic differences.

Another problem for the variable criterion account concerns the relation between the data of Meyer, Levelt, and Roelofs (in preparation) versus Roelofs and Meyer (1998). Santiago (2000) has recently demonstrated implicit priming effects using pictures as prompts rather than recently learned paired associates. Picture naming latencies were shorter when picture names in a block contained shared initial segments than when they did not. Indeed, error variability was less for the implicit picture naming task than for the standard task, at least for speakers of Spanish. A block of picture naming trials with names containing the same number of syllables but no segments in common can therefore be considered analogous to a standard implicit priming block with words homogeneous as to number of syllables. However, because Roelofs and Meyer (1998, Experiment 4) showed that purely metrical implicit priming effects for number of syllables do not occur, the variable criterion account is vacuous unless factors that lead to changes in criterion from one experimental paradigm to the other are specified (and tested).

Finally, having complete word planning at a phonological level together with a variable criterion only at the phonetic level is only a way of having one's cake and eating it too. These theoretically unparsimonious assumptions cannot enable WEAVER++, or any other similar model, to explain the missing length effects in specific circumstances. If a whole phonological word must be planned before starting pronunciation, delays introduced before phonetic encoding should cause longer latencies for longer words. Phonetic encoding may increase these processing delays even further (access of all phonetic syllables), or may introduce a constant delay (due to accessing the first syllable) for words of different lengths. However, phonetic encoding surely cannot supersede prior delays introduced at the phonological encoding level.

To conclude, Santiago et al.'s (2000) data are best explained in terms of structural factors (mainly number of syllables and onset complexity) rather than length in segments. The length hypothesis also fails to give a unified account of prior findings in the literature and of recent unpublished results, even supplemented with a variable criterion to start pronunciation. In our opinion, the bulk of the evidence available up to now suggests that the ultimate solution will involve structure-sensitive mechanisms for phonological encoding that differ across languages. Future evidence may also suggest a variable criterion to start speaking, but the factors affecting the setting of this criterion must be isolated if this account is to avoid circularity. Until then, however, it seems safe to assume that length per se has little direct effect on unprepared word production latencies if structural factors are controlled.

Manuscript received August 2000
Revised manuscript accepted September 2000

REFERENCES

Bachoud-Lévi, A.-C., Dupoux, E., Cohen, L., & Mehler, J. (1998). Where is the length effect? A cross-linguistic study of speech production. *Journal of Memory and Language*, *39*, 331–346.

Burke, D.M., MacKay, D.G., Worthley, J.S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language*, *30*, 542–579.

Dell, G.S. (1986) A spreading activation theory of retrieval in language production. *Psychological Review*, *93*, 283–321.

Dell, G.S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, *17*, 149–196.

Eriksen, C.W., Pollack, M.D., & Montague, W.E. (1970). Implicit speech: Mechanism in perceptual encoding? *Journal of Experimental Psychology*, *84*, 502–507.

Ferrand, L., Sgui, J., & Grainger, J. (1996). Masked priming of word and picture naming: The role of syllabic units. *Journal of Memory and Language*, *35*, 708–723.

Klapp, S.T., Anderson, W.G., & Berrian, R.W. (1973). Implicit speech in reading, reconsidered. *Journal of Experimental Psychology*, *100*, 368–374.

Levelt, W.J.M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W.J.M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*, 1–22.

Levelt, W.J.M., Roelofs, A., & Meyer, A, (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38.

Levelt, W.J.M., & Wheeldon, L.R. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*, 239–269.

MacKay, D.G. (1987). *The organisation of perception and and action.* New York: Spinger-Verlag.

MacKay, D.G. (1978). Speech errors inside the syllable. In A. Bell & J. Hooper (Eds.), *Syllables and segments* (pp. 201–212). New York: North Holland.

Meyer, A. (1990) The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, *29*, 524–545.

Meyer, A. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, *30*, 69–89.

Roelofs, A. (1997). The WEAVER model of word from encoding in speech production. *Cognition*, *64*, 249–284.

Roelofs, A. (1998). Rightward incrementality in encoding simple phrasal forms in speech production: Verb-particle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 904–921.

Roelofs, A. (2002). Syllable structure effects turn out to be word length effects: Comment on Santiago et al. (2000). *Language and Cognitive Processes*, *17*, 1–13.

Roelofs, A., & Meyer, A. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 922–939.

Santiago, J. (1997). *Estructura Silábica y Activación Secuencial en Producción del Lenguaje.* [Syllable Structure and Sequential Activation in Language Production]. Unpublished doctoral dissertation, Dept. Psicología Experimental y Fisiología del Comportamiento, Universidad de Granada, Spain.

Santiago, J. (2000). Implicit priming of picture naming: A theoretical and methodological note on the implicit priming task. *Psicológica, 21*, 39–59.

Santiago, J., MacKay, D.G., Palma, A., & Rho, C. (2000). Sequential activation processes in producing words and syllables: Evidence from picture naming. *Language and Cognition Processes*, *15*, 1–44.

Schiller, N. (1998). The effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory and Language*, *39*, 484–507.

Sternberg, S., Monsell, S., Knoll, R.L., & Wright, C.E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G.E. Stelmach (Ed.), *Information Processing in Motor Control and Learning* (pp. 117–152). New York: Academic Press.