

Monitoring syllable boundaries during speech production

Bernadette M. Jansma^{a,1} and Niels O. Schiller^{a,b,*}

^a *University of Maastricht, Faculty of Psychology, Department of Neurocognition, The Netherlands*

^b *Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

Accepted 3 December 2003

Available online 31 January 2004

Abstract

This study investigated the encoding of syllable boundary information during speech production in Dutch. Based on Levelt's model of phonological encoding, we hypothesized segments and syllable boundaries to be encoded in an incremental way. In a self-monitoring experiment, decisions about the syllable affiliation (first or second syllable) of a pre-specified consonant, which was the third phoneme in a word, were required (e.g., *ka.No* 'canoe' vs. *kaN.sel* 'pulpit'; capital letters indicate pivotal consonants, dots mark syllable boundaries). First syllable responses were faster than second syllable responses, indicating the incremental nature of segmental encoding and syllabification during speech production planning. The results of the experiment are discussed in the context of Levelt's model of phonological encoding.

© 2003 Elsevier Inc. All rights reserved.

Keywords: Speech production; Phonological encoding; Syllable boundaries; Self-monitoring; Dutch

1. Introduction

A basic question in phonological encoding is how the metrical frame of a word form is computed. Is metrical information also encoded incrementally? We tried to answer this question in another study in which we required participants to decide as fast as possible on the stress position of a word corresponding to a visually presented picture (Schiller, Peters, Jansma, & Levelt (submitted)). What we found was that initial stress yielded shorter self-monitoring latencies in bisyllabic words than final stress. Furthermore, in trisyllabic words, monitoring latencies were shortest for stress on the first syllable, followed by stress on the second syllable, followed by stress on the third syllable. That is, the encoding of stress follows the same rightward incremental pattern as the encoding of segments. Here, however, we will focus on the time course of syllable boundary en-

coding using a speech production task (implicit picture naming).

Before we describe our experiment, we will give the reader some background information about phonological encoding. Word form encoding or phonological encoding in speech production can be divided into a number of processes. Levelt, Roelofs, and Meyer (1999) present the most fine-grained model of phonological encoding to date. According to this model, phonological encoding can start after the word form (e.g., *kano/kano* 'canoe') of a lexical item has been accessed in the mental lexicon. First, the phonological encoding system must retrieve the corresponding segments and the metrical frame of a word form. Segmental and metrical retrieval are assumed to run in parallel. During segmental retrieval an ordered set of segments (phonemes) of a word form is retrieved (e.g., /k/, /a/, /n/, and /o/), while during metrical retrieval the metrical frame of a word is retrieved, which consists at least of the number of syllables and the location of the lexical stress (e.g., for *kano* this would be a frame consisting of two syllables the first of which is stressed, i.e., / ' _ _/; for discussion see Levelt et al., 1999). At this point, Levelt's theory assumes that stress is only stored for words with a non-default stress

* Corresponding author. Fax: +31-43-3884125.

E-mail addresses: n.schiller@psychology.unimaas.nl, niels.schiller@mpi.nl (N.O. Schiller).

¹ The first author, i.e., Bernadette M. Jansma, published under her maiden name Schmitt before 2003.

pattern. In Dutch, this would be non-initial stress. However, recent experiments (Schiller, Fikkert, & Levelt, 2004) showed that even exceptional stress patterns might not be stored in the lexicon as long as they can be derived by rule.

Then, during segment-to-frame association previously retrieved segments are combined with their metrical frame. The retrieved ordering of segments prevents them from being scrambled. They are inserted incrementally into slots made available by the metrical frame to build a so-called phonological word. This incremental syllabification process respects universal and language-specific syllabification rules, e.g., *ka.no*.² Evidence for the incremental ordering during *segmental* encoding comes from a number of studies using different experimental paradigms (e.g., Meyer, 1990, 1991; Schiller, in preparation; Van Turenhout, Hagoort, & Brown, 1997; Wheeldon & Levelt, 1995; Wheeldon & Morgan, 2002). The reason for “spelling out” lexical words only to rebuild them again into phonological words lies in the necessity to form maximally pronounceable syllables (see Levelt et al., 1999 for details). In connected speech, phonological words often have syllable structures deviating from the canonical syllable structures of the lexical words (see example in footnote 2). The domain of syllabification is the phonological word and this may be larger (clitics) or smaller (compounds) than the lexical words themselves (Booij, 1995). Segment-to-frame association is the process that lends the necessary flexibility to the system to cope with varying phonological contexts.

After the segments have been associated with the metrical frame, the resulting phonological syllables may be used to activate the corresponding phonetic syllables in a mental syllabary (Levelt & Wheeldon, 1994). Syllables in the syllabary may possibly be represented in terms of gestural scores (Browman & Goldstein, 1992) specifying articulatory motor programs for syllable-sized chunks. Although there is very little on-line evidence for the use of syllables in speech production (Ferrand, Segui, & Grainger, 1996; Ferrand, Segui, & Humphreys, 1997; but see Brand, Rey, & Peereman, 2003; Schiller, 1998, 2000; Schiller, Costa, & Colomé, 2002), the idea of having precompiled syllabic motor programs is very attractive because it decreases the computational load of the phonological/phonetic encoding component (Cholin, Schiller, & Levelt, 2004; Crompton, 1981; Levelt & Wheeldon, 1994; for lexico-statistical support see Schiller, Meyer, Baayen, & Levelt, 1996).

One idea is that syllables in the syllabary are activated through their segments and selected on the basis of

Luce’s choice rule (for details see Levelt et al., 1999; Roelofs, 1997). In case there is no corresponding syllable in the syllabary, it has to be computed on the fly by concatenating individual segments. Once the syllabic gestural scores are made available, they can be translated into neuro-motor programs, sent to the articulators, and then be executed resulting in overt speech. Levelt’s theory does not assume that the exact articulatory movement trajectories are programmed, but rather his theory assumes neuromuscular speech tasks to be achieved by the articulators (Fowler, Rubin, Remez, & Turvey, 1980; Kelso, Saltzman, & Tuller, 1986).

Evidence for the piecemeal nature of phonological encoding comes from a study by Wheeldon and Levelt (1995). They asked participants to monitor for pre-specified segments when generating the Dutch translation of an English word. This task can be seen as a production equivalent of the phoneme-monitoring task (Connine & Titone, 1996). They found that participants were faster in monitoring for the first consonant in a $C_1VC_2C_3VC_4$ (where C stands for consonant and V for vowel) word, such as *lifter* (‘hitchhiker’), than for the second consonant. Furthermore, they were faster in monitoring for C_2 than for C_3 and C_3 was faster than C_4 , although this last difference did not reach significance. Wheeldon and Levelt (1995) took their result to confirm the incremental encoding of segments during phonological encoding in speech production. They argued that their monitoring effect occurred at the phonological word level, i.e., when a fully syllabified phonological representation of a word is generated.³ Interestingly, there was no correlation between participants’ monitoring latencies for the target phonemes and the spoken duration of the carrier words (see also Schiller, in preparation). This suggested that the code being monitored must specify the constituent phonemes (the targets) but that the code is neither phonetic nor articulatory in nature. Recently, Wheeldon and Morgan (2002) replicated this Dutch result in English.

Interestingly, Wheeldon and Levelt (1995) found a significant 56 ms difference between the second and the third consonant in $C_1VC_2C_3VC_4$ words, i.e., at the syllable boundary (see below). They interpreted this effect as having to do with computing the syllable boundary, which delays the insertion of the segments in the second syllable. That is why C_3 yielded significantly longer monitoring latencies than C_2 . However, Wheeldon and Morgan (2002) could not exactly replicate this syllable boundary effect in English. They also found that the difference between the consonants at the syllable boundary (63 ms) was significant but the relative mag-

² A phonological word is not necessarily identical to the syntactic word because some syntactic words such as pronouns or prepositions, which cannot bear stress themselves, cliticize onto other words forming one phonological word together, e.g., *gave + it* → /gɑt.vɪt/.

³ A phonetic representation could be excluded as the locus of the effect because results remained the same when an articulatory suppression task, i.e., counting aloud, was added during monitoring (see Wheeldon and Levelt, 1995, Experiment 1b).

nitude of the effect was smaller than in the Wheeldon and Levelt (1995) study (compared to the difference between C₁ and C₂). They showed, however, that their syllable boundary effect was compromised by carrier words with ambisyllabic word-medial consonants. Wheeldon and Morgan (2002, pp. 516–517) concluded that, “the carrier word syllabification might indeed contribute to the size of the monitoring difference between the word medial consonant targets.”

In the present study, native speakers of Dutch were required to generate internally the corresponding phonological word form for a given picture and press a key when the word fulfilled a certain phonological criterion and withhold the key press when the word did not fulfill the criterion. By using tacit naming plus a minimal push-button response, we were able to investigate phonological and/or phonetic encoding in a direct way. The correctness of push-button responses suggested that participants came up with the correct and intended names of the pictures.

2. Experiment: syllabic decision with bisyllabic targets

A question, which has not been answered conclusively so far, has to do with the role of the syllable boundary. Wheeldon and Levelt (1995) found approximately the same difference in monitoring latencies between the first and the second consonant of a bisyllabic word (55 ms) as between the second and the third consonant (56 ms). However, in the former case there was an intervening vowel between C₁ and C₂, whereas in the latter case there was no vowel, but a syllable boundary in between. Wheeldon and Levelt (1995) accounted for this constant effect by proposing two different factors: (a) intervening segments, i.e., the vowel between C₁ and C₂, and (b) syllable boundaries, i.e., the boundary between C₂ and C₃. However, the fact that C₃ also occurred at a later position in the word than C₂ is confounded with the syllable boundary position. Therefore, it is unclear whether the longer monitoring latencies for C₃ compared to C₂ have anything to do with the preceding syllable boundary or whether they are simply an effect of serial order of encoding. Definitely, syllable boundaries should

have some effect if Levelt et al. (1999) are correct in assuming that segment monitoring occurs at the level of the fully syllabified phonological word. Therefore, we decided to investigate the role of syllable boundaries in monitoring. We did this by asking participants to tacitly name pictures and to determine whether the first post-vocalic consonant of a bisyllabic picture name was affiliated with the first or the second syllable.

3. Method

3.1. Participants

Eighteen undergraduate students from the University of Maastricht took part in the experiment (mean age 21.3, 10 women). They all had normal or corrected-to-normal vision and were paid for their participation. All participants were right-handed and native speakers of Dutch.

3.2. Materials

The materials consisted of 96 bisyllabic, monomorphemic Dutch nouns. Line drawings of the corresponding objects were either taken from the picture database of the Max Planck Institute for Psycholinguistics or drawn by a professional artist. Items could be divided into four groups of equal size depending on the consonant–vowel structure of their first syllable (CV vs. CVC) and the location of their lexical stress (initial vs. final). All items were between four and seven phonemes long and all were of low to moderate frequency as determined by CELEX (see Baayen, Piepenbrock, & Gulikers, 1995; see Table 1 for details). A complete list of all items can be found in Appendix A.

3.3. Procedure

Participants were tested individually. They were seated behind a computer screen and asked to place their right index finger on the right shift key of a keyboard that was placed in front of them. For each experimental trial, participants were asked to press the right shift key

Table 1
Lexico-statistical characteristics of the target picture names

Stress location	CV structure of the first syllable	Example	Mean CELEX frequency (per one million words)	Mean length in segments
Initial	CV	<i>kano</i>	31.1	5.1
Initial	CVC	<i>kansel</i>	19.5	6.2
Final	CV	<i>kanon</i>	19.0	5.0
Final	CVC	<i>kalkoen</i>	15.6	6.2

Note. The mean CELEX frequency for the CV items with initial stress is slightly higher than for the other three categories because one item, i.e., *tafel* ‘table,’ has a frequency of 247.4 per one million words, by far the highest frequency of all items. Discarding the item *tafel*, this category has a mean frequency of 21.7 per one million words.

when the pivotal consonant belonged to the first syllable (e.g., *kaN.sel* ‘pulpit’) and withhold the key press when it belonged to the second syllable (e.g., *ka.No* ‘canoe’). In a second block, they received the same stimuli, but the instructions were switched so that they actively responded if the target consonant belonged to the second syllable, but not if it belonged to the first. An experimental trial consisted of the following events: first, a fixation point appeared for 500 ms in the center of the screen, which participants were asked to fixate. Then, after 300 ms, a picture appeared around the same location on the screen. Pictures were of approximately equal size. They all fitted into a 7 × 7 cm square. As soon as possible after the picture appeared participants had to give their response. Reaction times (RTs) were registered automatically. The picture disappeared from the screen when participants responded or after 2000 ms. The following trial began after an inter-trial interval of 1000 ms. Trial sequencing was controlled by the Experimental Run Time Software (ERTS).

Before the experimental trials started, participants were familiarized with the pictures. Each picture was shown individually with the picture name underneath until the participant pressed the space bar and the next picture appeared. After picture familiarization, each picture was shown again to the participants who were asked to name the pictures aloud as fast and as accurately as possible. The practice block served the purpose of demonstrating whether or not participants knew the name for each picture. In the experimental trials, participants were asked to suppress overt naming of the pictures and—if necessary—press the key as fast and as accurately as possible after a picture appeared on the screen.

3.4. Design

The experiment started with a familiarization and a practice block. Then two test blocks followed with reversed instructions. After each block there was a short break. The order of trials was randomized for each block and each participant individually. Half of the participants started with a block in which they had to actively respond to picture names with the target consonant in the first syllable and withhold response for names with the target consonant in the second syllable. Then they received a second block with the same material in which the response contingencies were reversed. The other half of the participants was presented with the reversed block order.

4. Results

Incorrect responses and time-outs were counted as errors (14.8%) and discarded from the RT analysis. Our hypothesis was that syllabic encoding should take place

incrementally. We expected to see longer RTs for target consonants located in second compared to first syllables. Descriptively, this expectation was confirmed by the data. Mean RTs for the two conditions were 1017 ms ($SD = 98$) for the first syllable condition and 1056 ms ($SD = 106$) for the second syllable condition. Mean decision latencies for first syllable affiliation were 39 ms faster than for second syllable affiliation. One tailed t tests revealed that RTs were significant by participants and items ($t_1(17) = 2.25$, $p < .05$; $t_2(94) = 2.16$, $p < .05$). Error data support this trend and revealed more errors for the second syllable condition (17.8%) than for the first syllable condition (11.7%). This difference also turned out to be significant, based on paired sample t tests on arc-sin transformed error proportions ($t_1(17) = 2.16$, $p < .05$; $t_2(94) = 2.60$, $p < .05$).

5. Discussion

To be able to make a syllabic decision, participants had to phonologically encode the name of the picture presented on the screen. Only after syllabifying the word, they could make the decision about the syllable affiliation. Levelt et al. (1999) argued that this is the phonological word level. We found a clear advantage for the first syllable over the second. This advantage cannot be attributed to the position of the target segment in the word, because the pivotal consonant was almost exclusively the third segment.⁴ The only difference between first and second-syllable condition is the location of the syllable boundary, i.e., either before or after the pivotal segment. When the segment was before the syllable boundary (as, for instance, in *kaN.sel*) participants were 39 ms faster to make their syllabic decision than when the segment occurred after the syllable boundary (as, for instance, in *ka.No*).

Since syllabification and segmental encoding presumably run incrementally, this effect can be explained in a straightforward fashion. In the case of *kaN.sel*, the pivotal segment is encoded before the syllable boundary has been inserted. In contrast, for words like *ka.No*, the segment only occurs after the syllable boundary has been inserted. Therefore, the decision about the syllabic affiliation is also slightly delayed in the latter condition as compared to the former. The difference between the two conditions (i.e., 39 ms) might be due to computing the syllable boundary. To compare, the syllable boundary effect found by Wheeldon and Levelt (1995) with a similar self-monitoring task was 56 ms.

⁴ In three second-syllable words (e.g., *sleutel*, *spijker*, and *vlieger*) the pivotal segment was in fourth position. However, even in five first-syllable words (e.g., *pleister*, *scalpel*, *tractor*, *trompet*, and *vlinder*) the pivotal segment was the fourth segment.

However, there might possibly be an alternative account for the results: If the majority of the pictures for second syllable words (e.g., *ka.No*) took longer to recognize than pictures for first syllable words (e.g., *kaN.sel*), this could be the reason for the observed syllable affiliation effect. This was tested in an object/non-object decision experiment: Ten participants, all students from the University of Nijmegen, saw either one of the 96 pictures of existing objects (e.g., persons, animals, natural, and artificial objects) or one of the 48 pictures of nonsense objects (taken from Kroll & Potter, 1984) and were required to press with their preferred hand side as fast and as accurately as possible the *YES* button on a button box if they thought the picture was denoting an existing object and the *NO* button otherwise. The trial sequencing was similar to the main experiment reported above. Participants visually inspected all the pictures of existing objects and nonsense objects before the object/non-object experiment started. The experiment was run in two blocks. Each block contained 12 pictures of existing objects from each of the four experimental categories plus the 48 pictures of nonsense objects. The same nonsense objects were presented in both blocks. Between the two blocks there was a short break. The order of trials was randomized individually for each block and participant. The mean decision latencies for the two syllable affiliation conditions of the first post-vocalic consonant (first vs. second syllable) were 433 ms ($SD = 29$) for picture names with first syllable consonants (e.g., *kaN.sel*) and 429 ms ($SD = 38$) for picture names with second syllable consonants (e.g., *ka.No*). The 4 ms difference between the second and first syllable items was not significant ($t_1(9) < 1$; $t_2(94) < 1$), which means that pictures whose names had the pivotal consonant in the first or second syllable were recognized equally fast.

Another potential criticism is that the results do not reflect incremental phonological but rather differences in lexical access time. If lexical access for targets with initial stress was faster than lexical access for targets with final stress, e.g., because of the computation or retrieval of regular vs. irregular stress patterns, then this might account for the effect reported above. In order to test this potential confound, we carried out another control experiment employing a picture-naming task. Thirty new participants were tested in a standard picture naming experiment. The same 96 pictures used in the stress decision task appeared one at a time on a computer screen and the participants' task was to name the pictures as fast and as accurately as possible. The experiment started with a familiarization block in which each participant saw each picture on the screen one at a time. Each trial in the picture naming part started with a fixation point that was visible for 500 ms in the center of the screen and followed by a blank screen for 300 ms. Then the picture appeared in the center of the screen and

remained in view until a verbal response was given. At picture onset, a clock was started. Verbal responses were registered with a microphone in front of participants. The microphone was connected to a voice key, which stopped the clock when it was triggered. After 1000 ms the next trial started. Presentation of the trials was controlled by NESU. Errors (wrong responses, voice-key failures, etc.) and time-outs were discarded from the RT analysis (4.1%). Also, we only took into account RTs between 300 and 1500 ms. The mean naming latencies for picture names with initial stress was 823 ms ($SD = 56$) while it was 787 ms ($SD = 69$) for picture names with final stress. This 36 ms advantage for picture names with final stress over picture names with initial stress was significant by participants but not by items ($t_1(29) = 5.33$, $p < .01$; $t_2(94) = 1.74$, n.s.). Error rates showed no significant effect. The naming advantage of final over initial stress pictures showed that monitoring latencies and picture naming latencies were not confounded in our stress decision experiment.

6. General discussion

In this paper, we modified a methodology introduced by Wheeldon and Levelt (1995) to investigate the time course of phonological encoding during language production. We were especially interested in syllabification. The results of Wheeldon and Levelt's study demonstrated that the representation on which the monitoring response is based is phonological and syllabified in nature. Participants are monitoring an internal abstract code, i.e., the output of the process that assigns segments (phonemes) to a syllabified prosodic frame.

Meyer (1990, 1991), Wheeldon and Levelt (1995), and Van Turenout et al. (1997) showed that the segmental encoding of speech is essentially an incremental process. Of course, overt speech is a sequential process and necessarily has to proceed from beginning to end. But the studies mentioned above investigated the phonological *planning stage* of word generation and found strict serial ordering effects.

In general, our results fit into the picture of phonological encoding drawn by Levelt and Wheeldon (1994). These authors proposed that the prosodic frame of a word is made available syllable-by-syllable and that the corresponding segments are subsequently assigned to their slots in the syllable frame. Wheeldon and Levelt (1995) made the additional assumption that the computation of a syllable boundary delays the initiation of segmental assignment in the following syllable. This delay at the syllable boundary might be the reason for the outcome of the main experiment reported here. Monitoring for a segment that follows a syllable boundary takes longer than monitoring for a segment in the same segmental position before a syllable boundary

because computing and inserting the syllable boundary presumably takes time.

We interpreted the effects obtained in the experiments reported above as genuine speech production effects. Moreover, we were able to refute two alternative accounts through control experiments, namely a visual perceptual account and a lexical access account (see above). However, it is theoretically possible that the effect we measured is a perception and not a production effect. Assume that speakers generate internally the name of a given picture. Instead of phonologically encoding the picture name and monitoring it at the same time (production monitoring) it is conceivable that participants first encoded the target word and afterwards scanned the encoded word for syllable boundaries (perception monitoring). Theoretically, we cannot disentangle those two possibilities because both would yield incremental results. However, we know from segmental monitoring studies (Schiller, in preparation; Wheeldon & Levelt, 1995; Wheeldon & Morgan, 2002) that the acoustic characteristics of the target words (e.g., the acoustic distance between the to-be-monitored segments) exhibit a pattern different from the monitoring results. For instance, Wheeldon and Morgan (2002) found that the interval of monitoring latencies between word initial and final phonemes were significantly shorter than the corresponding interval of articulatory

duration. Also, similar to Wheeldon and Levelt (1995), they did not find a correlation between the differences in monitoring latencies and the corresponding speech measurements. Furthermore, when the target words were presented overtly and participants were asked to make a decision about the presence or absence of certain segments in the acoustic signal (external monitoring), weak but significant correlations were observed between internal and external monitoring latencies. These results might be attributed to similarities in the processes that monitor both codes and taken as evidence against a perceptual monitoring account, i.e., retrieving and then scanning a phonological code (Morgan & Wheeldon, 2003; Wheeldon & Morgan, 2002). Our interpretation is that the effects we were measuring have their basis in speech production but since self-monitoring involves the comprehension system (Levelt et al., 1999; Levelt, 2001), perceptual characteristics might be reflected in the data as well.

Acknowledgments

Niels O. Schiller is supported by the *Royal Dutch Academy of Arts and Sciences* (KNAW). The authors thank Mart Bles for his assistance in running the experiment.

Appendix A. Materials (target pictures) used in the experiment

Targets with initial stress		Targets with final stress	
CV	CVC	CV	CVC
bezem ('broom')	banjo ('banjo')	banaan ('banana')	balkon ('balcony')
boter ('butter')	borstel ('brush')	beha ('bra')	biljart ('pool')
hamer ('hammer')	bunker ('bunker')	bureau ('desk')	bonbon ('candy')
jager ('hunter')	cactus ('cactus')	citroen ('lemon')	dolfijn ('dolphin')
kabel ('cable')	cirkel ('circle')	fabriek ('factory')	garnaal ('shrimp')
kano ('canoe')	dokter ('doctor')	gebit ('dentures')	gordijn ('curtain')
kegel ('bowling pin')	gondel ('gondola')	geweer ('rifle')	harpoen ('harpoon')
ketel ('kettle')	halter ('weight')	giraf ('giraffe')	kalkoen ('turkey')
koning ('king')	herder ('shepherd')	gitaar ('guitar')	karkas ('skeleton')
lepel ('spoon')	kansel ('pulpit')	kameel ('camel')	kasteel ('castle')
molen ('wind mill')	lifter ('hitch hiker')	kanon ('canon')	kompas ('compass')
motor ('motor bike')	masker ('mask')	karaf ('pitcher')	lantaarn ('lantern')
nagel ('finger nail')	panter ('panther')	konijn ('rabbit')	magneet ('magnet')
navel ('navel')	parfum ('perfume')	libel ('dragonfly')	pastoor ('priest')
ratel ('rattle')	pinguin ('penguin')	loket ('counter')	penseel ('brush')
robot ('robot')	pleister ('band aid')	matras ('mattress')	pincet ('tweezers')
sleutel ('key')	scalpel ('scalpel')	meloen ('melon')	pistool ('gun')
spijker ('nail')	stempel ('stamp')	piraaf ('pirate')	pompoen ('pumpkin')
tafel ('table')	tempel ('temple')	piloot ('pilot')	portret ('portrait')
tijger ('tiger')	tractor ('tractor')	raket ('rocket')	sandaal ('sandal')
toren ('tower')	varken ('pig')	rivier ('river')	soldaat ('soldier')

Appendix A (continued)

Targets with initial stress		Targets with final stress	
CV	CVC	CV	CVC
vlieger ('kite')	vlinder ('butterfly')	sigaar ('cigar')	tampon ('tampon')
vogel ('bird')	wortel ('carrot')	tomaat ('tomato')	trompet ('trumpet')
zebra ('zebra')	zuster ('nurse')	toneel ('stage')	vampier ('vampire')

References

- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (CD-ROM)*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- Booij, G. (1995). *The phonology of Dutch*. Oxford: Clarendon Press.
- Brand, M., Rey, A., & Peereman, R. (2003). Where is the syllable priming effect in visual word recognition? *Journal of Memory and Language*, 48, 435–443.
- Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49, 155–180.
- Cholin, J., Schiller, N. O., Levelt, W. J. M. (2004). The preparation of syllables in speech production. *Journal of Memory and Language*, 50, 47–61.
- Connine, C. M., & Titone, D. (1996). Phoneme monitoring. *Language and Cognitive Processes*, 11, 635–646.
- Crompton, A. (1981). Syllables and segments in speech production. *Linguistics*, 19, 663–716.
- Ferrand, L., Segui, J., & Grainger, J. (1996). Masked priming of word and picture naming: The role of syllabic units. *Journal of Memory and Language*, 35, 708–723.
- Ferrand, L., Segui, J., & Humphreys, G. W. (1997). The syllable's role in word naming. *Memory & Cognition*, 35, 458–470.
- Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production: Speech and talk* (vol. 1, pp. 373–420). New York, NY: Academic Press.
- Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, 14, 29–59.
- Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning and Verbal Behavior*, 23, 39–66.
- Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceeding of the National Academy of Sciences*, 98, 13464–13471.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75.
- Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50, 239–269.
- Meyer, A. S. (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. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, 30, 69–89.
- Morgan, J. L., & Wheeldon, L. R. (2003). Syllable monitoring in internally and externally generated English words. *Journal of Psycholinguistic Research*, 32, 269–296.
- Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, 64, 249–284.
- Schiller, N. O. (1998). The effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory and Language*, 39, 484–507.
- Schiller, N. O. (2000). Single word production in English: The role of subsyllabic units during phonological encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 512–528.
- Schiller, N. O. (in preparation). *The incremental nature of phonological encoding in speech production*.
- Schiller, N. O., Costa, A., & Colomé, A. (2002). Phonological encoding of single words: In search of the lost syllable. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 35–59). Berlin: Mouton de Gruyter.
- Schiller, N. O., Fikkert, P., & Levelt, C. C. (2004). Stress priming in picture naming: An SOA study. *Brain and Language*.
- Schiller, N. O., Meyer, A. S., Baayen, R. H., & Levelt, W. J. M. (1996). A comparison of lexeme and speech syllables in Dutch. *Journal of Quantitative Linguistics*, 3, 8–28.
- Schiller, N. O., Jansma, B. M., Peters, J., & Levelt, W. J. M. (submitted). *Monitoring metrical stress in polysyllabic words*.
- Van Turenhout, M., Hagoort, P., & Brown, C. M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 787–806.
- Wheeldon, L., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, 34, 311–334.
- Wheeldon, L., & Morgan, J. L. (2002). Phoneme monitoring in internal and external speech. *Language and Cognitive Processes*, 17, 503–535.