Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Memory and Language 50 (2004) 47–61
Journal of Memory and Language
www.elsevier.com/locate/jml

The preparation of syllables in speech production

Joana Cholin,[a,*] Niels O. Schiller,[a,b] and Willem J.M. Levelt[a]

[a] *Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands*
[b] *University of Maastricht, The Netherlands*

Received 5 March 2003; revision received 11 August 2003

Abstract

Models of speech production assume that syllables play a functional role in the process of word-form encoding in speech production. In this study, we investigate this claim and specifically provide evidence about the level at which syllables come into play. We report two studies using an *odd-man-out* variant of the *implicit priming paradigm* to examine the role of the syllable during the process of word formation. Our results show that this modified version of the implicit priming paradigm can trace the emergence of syllabic structure during spoken word generation. Comparing these results to prior syllable priming studies, we conclude that syllables emerge at the interface between phonological and phonetic encoding. The results are discussed in terms of the WEAVER++ model of lexical access.

Keywords: Language production; Phonological encoding; Syllables; CV-structure; Implicit priming paradigm; Form preparation

The role of the syllable as a functional unit in the speech production process has been investigated in several psycholinguistic studies (Baumann, 1995; Chen, Chen, & Dell, 2002; Chen, Lin, & Ferrand, 2003; Ferrand, Segui, & Grainger, 1996; Ferrand, Segui, & Humphreys, 1997; Levelt, 1992; Levelt & Wheeldon, 1994; Meijer, 1996; Schiller, 1997, 1998, 1999, 2000; Schiller, Costa, & Colomé, 2002). Many of the off-line studies suggest the existence of the syllable as a production unit (Fromkin, 1971; Schiller, Meyer, & Levelt, 1997; Shattuck-Hufnagel, 1987, 1992; Treiman, 1983; Treiman & Danis, 1988). For example, speech error data suggest that segmental errors such as exchanges of segments only take place for identical syllable internal positions, i.e., onsets exchange with onsets, nuclei exchange with nuclei, etc. (Berg, 1988; MacKay, 1970; Nooteboom, 1969; Shattuck-Hufnagel, 1979; Stemberger, 1982). This is referred to as the *syllable position constraint*. However, a quantitative analysis showed that the majority of such errors occurs in the onset position. Thus, the syllable onset constraint may be a word-onset constraint (Shattuck-Hufnagel, 1987, 1992; see Meyer, 1992; for a critical review). Evidence from metalinguistic tasks suggests that syllables play a role at some level of processing in speech production (Schiller et al., 1997; Treiman, 1983; Treiman & Danis, 1988; see Bagemihl, 1995 for a review) but makes no strong claim about where. Despite the (limited) off-line support for the syllable and the relevance of syllables to linguistic phenomena, on-line experiments do not provide evidence that the syllable is a production unit (Brand, Rey, & Peereman, 2003; Evinck, 1997; Schiller et al., 2002).

The majority of prior on-line studies used some form of priming as their experimental method. The experiments reported here use a different paradigm to investigate the syllable as a processing unit, i.e., the *implicit priming paradigm* (Meyer, 1990, 1991). The existence of syllabic units is assumed by two influential models of speech production, i.e., the Levelt, Roelofs, and Meyer model (1999) on the one hand and the model proposed by Dell (1986, 1988) on the other hand. Despite this general agreement, these models differ in the status of syllabic units in phonological encoding. Dell's (1986,

* Corresponding author. Fax: +31-24-3521-213.
E-mail address: Joana.Cholin@mpi.nl (J. Cholin).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2003.08.003

1988) model includes word forms that are already syllabified when retrieved from the mental lexicon, i.e., an abstract phonological representation which is specified not only for its segmental composition but also for its internal syllabic structure. In contrast, the model of spoken word production proposed by Levelt et al. (1999) assumes syllables play a crucial role at the interface of phonological and phonetic encoding. At this interface, abstract phonological syllables are generated which are subsequently mapped onto phonetic syllables. We will return to the issue of when syllables are predicted to play a role in speech production periodically throughout the paper. Here, we introduce a version of the implicit priming paradigm that specifically taps into the preparation of syllable structure. The Levelt et al. (1999) model of lexical access, and its computer simulation WEAVER++, will be taken as the theoretical framework for the interpretation of our findings. This requires a short introduction to the model's phonological and phonetic encoding parts.

Phonological and phonetic encoding in WEAVER++

According to the WEAVER++ model (Levelt et al., 1999; Roelofs, 1997b), the preparation of a spoken word proceeds through a number of stages. After conceptually driven selection of the appropriate lemma from the mental lexicon, the target word is first phonologically encoded, which largely consists of computing its syllabification and prosody. This is incrementally followed by phonetic encoding, which includes the computation of the articulatory gestures for the target word's syllables in their phonetic context. Finally, the execution of these gestural scores by the laryngeal and supralaryngeal muscle systems produces the acoustic realization of the spoken word. The present paper exclusively concerns the stages of phonological and phonetic encoding.

Phonological encoding

The first operation in phonological encoding is the retrieval of the target word's *phonological code* from the mental lexicon. The code consists of an ordered set of phonemic segments. For stress-timed languages such as English and Dutch the model also assumes the existence of sparse metrical markers in phonological codes. More specifically, the stress position is marked for those words whose stress does not appear in default position (but see Schiller, Fikkert, & Levelt, in press for a different position). For English, the default position is defined as the first full-vowel syllable of the word. Different from other models of spoken word production (in particular Dell, 1986, 1988), Levelt et al.'s retrieved phonological codes are not syllabified. The main argument for this assumption derives from the phenomenon of *resyllabification*. In connected speech, syllable boundaries often differ from a word's canonical syllabification. The domain of syllabification is the phonological word which can be smaller or larger than the lexical word due to morpho-phonological processes like inflection or cliticization (Booij, 1995). If, for instance, the stored phonological code for the word *predict* would be syllabified (i.e., as *pre-dict*), then the speaker must 'resyllabify' the word when used in a different context, such as past tense (*pre-dic-ted*) or cliticization (*predict it—pre-dic-tit*). The ubiquity of such 'resyllabifications' in the normal use of English (or Dutch for that matter), would make this a highly inefficient procedure.[1] For a language like Mandarin Chinese, which has a small set of syllables and limited resyllabification processes, the story might be different. The issue of cross-linguistic differences will be revisited later in the paper.

The alternative assumption is, therefore, that a word's syllabification is not retrieved, but generated 'on the fly,' dependent on the context in which the word appears. During this process, called 'prosodification,' spelled-out segments are incrementally combined to form successive syllables. Also, these successive syllables are incrementally assigned the appropriate metrical properties, either following default stress, or otherwise the retrieved non-default stress marking feature. The incremental composition of syllables follows, on the one hand, universal syllabification constraints (such as maximization of onsets and sonority gradations) and, on the other hand, language-specific rules, e.g., phonotactics. Together, these rules create maximally pronounceable syllables. The output of phonological encoding is a phonological word, specified for its metrical, syllabic, and segmental properties.

Before turning to the next processing step we will briefly describe the assumptions Dell's (1986, 1988) model makes with respect to the phonological encoding process. As already mentioned above, this model includes abstract phonological representations that are specified for internal syllabic positions, i.e., the word form retrieved from the mental lexicon activates not only segmental information but also syllabic frames. These syllabic frames serve as placeholders into which the retrieved segments are inserted during the process of segment-to-frame-association.

Phonetic encoding

These fairly abstract, syllabified phonological words are incrementally translated into articulatory-motor programs. These programs consist in large part of

[1] Schiller, Meyer, Baayen, and Levelt (1996) estimated the occurrence of resyllabifications in a running text for Dutch. Approximately once every six words speakers of Dutch would have to resyllabify their lexical forms.

specifications for subsequent syllabic gestures. One assumption of the theory is that speakers have access to a repository of syllabic gestures. This repository, coined the 'mental syllabary' (Levelt, 1992; Levelt & Wheeldon, 1994), contains the articulatory scores for at least the high-frequency syllables of the language. Schiller has computed that English speakers do some 85% of their talking with no more than 500 different syllables (out of some 12,000, see Schiller et al., 1996; Schiller, 1997). Hence, for normal speakers, the corresponding articulatory gestures may have become highly over-learned motor actions. The model assumes that as soon as a syllable emerges during incremental syllabification, the corresponding syllabic gesture will be selected from the repository in Broca's area or a pre-motor area (Dronkers, 1996; Indefrey & Levelt, 2000; Kerzel & Bekkering, 2000). Proposing the notion of a mental syllabary, however, was not intended to deny the existence of a mechanism for the generation of low-frequency or entirely new syllabic gestures. That mechanism is still to be modeled in detail within the framework of WEAVER++, but irrelevant for the present discussion.

In summary, the theory proposed by Levelt et al. (1999) takes syllables, rather than segments, to be basic programming units of speech articulation. This is entirely in line with traditional notions of speech generation. According to Fujimura and Lovins (1978) as well as Lindblom (1983) only the syllable can form the appropriate source for late phonological processes, such as allophonic variation, coarticulation and (as a result of this) assimilation. Phenomena such as word-initial aspiration of plosives in English or word-final devoicing in Dutch or German can be described conveniently with reference to the syllable as a unit (see Kenstowicz, 1994).

The WEAVER++ model specifies where syllabic patterns emerge in speech generation. First, syllables are not stored in the mental lexicon; they are not specified in the phonological codes speakers retrieve from their form lexicon. This predicts the absence of syllable-specific effects in priming paradigms because syllables are not represented as units in long-term memory. Below, we will discuss in more computational detail the basis for this prediction as well as the relevant evidence (and counter evidence). Second, phonological syllables first arise during incremental phonological encoding, i.e., during context-sensitive 'syllabification.' Third, as phonological syllables arise, they trigger the retrieval of syllabic articulatory gestures ('phonetic syllables') from a repository of articulatory motor actions, the 'mental syllabary.' The purpose of the present paper is to trace the emergence of syllables in word generation by means of a paradigm which manipulates the speaker's ability to do advance preparation of a syllable. It can provide the speaker with a head-start in syllabification and in retrieving a word's first syllabic gesture.

Priming studies of syllable access

Several cross-linguistic studies were conducted to investigate whether syllables could be primed and thereby identified as an independent unit in the process of speech production (for Dutch: Baumann, 1995; Schiller, 1997, 1998, 1999; for Mandarin Chinese: Chen et al., 2003; for French: Brand et al., 2003; Evinck, 1997; Ferrand et al., 1996; Schiller et al., 2002; for English: Ferrand et al., 1997; Schiller, 2000; Schiller & Costa, submitted; for Spanish and an overview see Schiller et al., 2002). One of the first studies that was conducted in order to test whether syllables can be primed in speech production was Baumann (1995). She investigated the time course of syllabification during phonological encoding in Dutch. In a series of priming experiments using a semantic-associate learning task, she tested whether a syllable priming effect could be obtained. A first finding of her experiments was that phonologically related primes, whatever their syllabic relation to the target word, facilitated the response relative to unrelated control primes. A second result was that, in all related conditions, CVC-primes were more effective than CV-primes. But, thirdly, no specific syllable priming effects were obtained.

Several subsequent studies have failed to find a syllable priming effect but rather confirmed the finding of a segmental length effect. Much discussion has been given to the results of the apparent syllable priming effect in French (Ferrand et al., 1997, Experiment 5). However, Brand et al.'s (2003) failure to replicate the Ferrand effects suggests that this should not be taken as strong evidence for the syllable effect (see also Evinck, 1997; and for a review Schiller et al., 2002). In sum, the evidence from syllable priming tasks may indicate this method is not tapping into the appropriate level of processing to reveal potential syllable effects.

Syllable frequency studies

Syllabic effects are, however, predicted for access to the hypothesized mental syllabary. Gestures for high-frequency syllables should be more accessible than gestures for low-frequency syllables. (The argument is further spelled out in the next section). In order to find empirical evidence for this, Levelt and Wheeldon (1994) investigated naming latencies for words consisting of high- versus low-frequency syllables. The prediction in three naming tasks was that, under the assumption of the existence of a mental syllabary, onset latencies for words that consist of high-frequency syllables should be shorter than those for words consisting of low-frequency. This expectation was inspired by the finding that word form access is sensitive to word frequency (Jescheniak & Levelt, 1994; Oldfield & Wingfield, 1965), although that involves a different level of processing. Levelt and Wheeldon's core finding was that, when word frequency was controlled for, words with high-frequency syllables were named faster than words with low-frequency syllables. If syllables are computed on-line rather than retrieved from a repository, their frequency of use should be irrelevant. The obtained syllable frequency effects therefore seemed to support the notion of the mental syllabary, where syllables are stored separately from words.

One potential problem with this conclusion is that syllable frequency was correlated with segment frequency in some of Levelt and Wheeldon's experiments. It is hardly possible in Dutch to disentangle these effects. A replication of this syllable frequency effect with carefully controlled experimental material would be desirable to allow for any strong claims.

The WEAVER++ predictions in more detail

The WEAVER++ model provides an account for the absence of a syllable priming effect and for the presence of a syllable frequency effect. So far, the model has been more successful in the former case than in the latter. Here the computational rationale is discussed in some more detail, because it provides at the same time the motivation for the present experiments. *WEAVER* (*W*ord-form *E*ncoding by *A*ctivation and *VER*ification) is the spreading activation based computer network model developed by Roelofs (1992, 1996, 1997a, 1997b, 1998, 1999), which is based on Levelt's (1989, 1992) theory of speech production. WEAVER++ adopts Dell's (1986) assumption of word form retrieval by the spread of activation and Levelt's (1992) on-line syllabification and access to a syllabary (Levelt & Wheeldon, 1994).

In accordance with Levelt and Wheeldon (1994), WEAVER++ (Roelofs, 1997a, 1997b) assumes that the syllabification of a word is computed on-line during the speech production process. In the WEAVER++ model, segments in the retrieved phonological code are not specified for their syllable position, but only for their serial order within a word. The actual syllabic position of a segment is determined by the syllabification process. Each retrieved segment in the phonological code spreads activation to all syllabic gestures in which it partakes. Hence, upon retrieval of a phonological code, there are always multiple phonetic syllable programs in a state of activation. How is the appropriate syllable program selected? There are, first, selection conditions. The crucial one is that the syllable matches the phonological syllable that is incrementally composed; this involves a procedure of verification. Second, each syllable in the syllabary has a frequency dependent selection threshold. This causes the predicted syllable frequency effect on naming latencies. Notice, however, that the threshold assumption is a 'modular' one. Removing it does not affect the architecture of the system. Third, selection is subject to Luce's (1959) choice rule. During any smallest interval, the probability of selecting the (verified) target syllable equals the ratio of its activation to the summed activation of all syllable nodes. Given the choice ratio, the expected selection latency can be computed.

WEAVER++ does not predict any syllable priming effects, at least not for Dutch and English for the following reasons: first, there is no syllable structure in the phonological code. The code's retrieval cannot be specifically primed by a string of segments that matches the word's canonical syllable structure. Phonologically related primes in the masked priming paradigm (pre-) activate the phonological segments retrieved from the mental lexicon during segmental spellout. As a consequence, the longer the prime the more segments get (pre-) activated during segmental spellout and the shorter the phonological code's selection latency. Notice that primes never get articulated in priming tasks. Hence, there is no need to transform incoming segments into syllables, i.e., no syllable structure is imposed on the input (it is free to map onto all compatible syllables).

In addition, incremental syllabification is not specifically facilitated by syllable matching primes. Take a CV.CVC target word such as *lotus*. A masked visual CV-prime (LO) will activate the first two segments and all syllable programs in which they partake, including the syllable program [lo] but also the syllable program [lot]. A CVC-prime (LOT) will activate the first three segments of the phonological code, the syllable program [lo] to the same amount that the CV prime (LO) did, and in addition the syllable program [tus] to some extent. Hence, it primes the relevant syllable programs despite the fact that it does not correspond to the first syllable of the target word. Therefore, there will only be a number-of-segments effect. The same holds when the target word has a CVC.CVC structure, such as *cactus*. Here the prime (CA) will be less effective than the prime (CAC), for similar reasons. Thus, WEAVER++ predicts an effect of prime length or the so-called segmental overlap effect (Schiller, 1998, 1999, 2000) but no interaction of prime and target syllabic structure.

Following the arguments of the model, in order to identify the syllable as a processing unit, we have to investigate the late syllabification process with a method that involves the advance construction of a phonological word's first syllable and the corresponding advance syllabary access. The implicit priming paradigm (Meyer, 1990, 1991; Roelofs, 1996, 1998; Roelofs & Meyer, 1998) provides access to exactly these late steps in spoken word encoding. Whereas (explicit) priming is sensitive only to early stages of phonological encoding, the implicit priming paradigm exhibits effects that emerge at these early stages but also comprise later stages at the interface of phonological and phonetic encoding, i.e., on-line syllabification, possibly including syllabary access.

In the implicit priming paradigm, participants repeatedly produce a syllable shared by several response words (as in *lotus*, *local*, and *loner*). As speakers know of the shared properties between response words they can prepare the first phonological syllable and the corresponding syllable program of the target word. The empirical issue is then whether a prepared syllable of the target word is a more effective preparation than a non-syllabic string of prepared segments, other factors being controlled. To answer this question, we designed a special version of the implicit priming paradigm.

The implicit priming paradigm

The basic paradigm

The implicit priming paradigm involves the production of words that are part of a list of previously learned paired associates. Participants learn a small set of prompt–response pairs. The response words are either phonologically related or not. Each experiment using the implicit priming paradigm consists of two types of sets, called the homogeneous and heterogeneous sets. In the homogeneous set, the response words share part of their form, e.g., the first syllable in *loner*, *local*, *lotus*, or the first syllable in *beacon*, *beadle*, *beaker*, or the first syllable in *major*, *maker*, and *maple*. The heterogeneous sets are created by regrouping the pairs from the homogeneous sets, e.g., *loner*, *beacon*, *major* (etc.). Each word is thus tested under both the homogeneous and the heterogeneous conditions, hence each word is its own control in the experiment. Production latency (the time between onset of prompt and speech onset, measured by voice key) is the dependent variable. A preparation effect is said to have occurred if production latencies in the homogeneous condition are shorter than in the heterogeneous condition. Meyer (1990, 1991) reported such a preparation effect only when the response words in the homogeneous sets shared one or more word-initial segments. No effect was found for shared word-final segments demonstrating the incrementality of the process of syllabification. The preparation effect was found to increase with the length of the shared initial stretch.

Using this paradigm, Meyer (1991) reports that sets with open initial syllables (CV) that share only those two initial segments produced preparation effects that were equivalent to effects produced for sets with closed syllables (CVC) that shared three initial segments. This result was surprising because a pure segmental length effect would predict larger preparation effects in the CVC sets since they comprise one more shared segment. This finding supports the possibility of syllabic effects that are independent of segmental length.[2]

The paradigm with an odd-man-out

To investigate a specific syllable preparation effect, we opted for a slightly different variant of the original implicit priming paradigm, i.e., the implicit priming paradigm with an odd-man-out (Janssen, Roelofs, & Levelt, 2002). The term odd-man-out labels an item in the response list of a homogeneous set that has (compared to the other words in that list) a different feature, such as another syllabic structure. The homogeneous set containing an odd-man-out is the so-called *variable* set; the homogeneous set without an odd-man-out is called the *constant* set.

In a constant set, the response-words consist of items which share *two* phonological properties. One is always the shared word-initial segments. The other can, for instance, be the word's syllable structure. The constant set can be *beacon*, *beadle*, *beaker*, sharing both the initial CV and the initial syllable. In comparison with the constant set, the variable sets contain only one of these two phonological properties, namely the shared word onset. A related variable set could be *beacon*, *beatnik*, *beaker*, where all items share the initial CV, but where they do not share the initial syllable (*bea* versus *beat*).

We chose the odd-man-out variant of the implicit priming paradigm rather than the classic version for its ability to keep one phonological property constant while systematically manipulating the other. In the present studies, the number of shared initial segments is held constant but the underlying syllabic structure differs. Thus, we can investigate what knowledge about the to-be-produced word speakers need in order to (successfully) prepare for it. If segmental information about syllabic structure is sufficient to prepare for the target word, then it should make no difference whether or not the response words in a set additionally share the syllabic structure. For this scenario, response times should be equally fast independent of syllabic structure; thus no effect of the manipulation should be observed. If speakers need the information of the current syllabic structure in addition to the information of the shared initial segments, then they should be faster in preparing for constant items than for variable items. In the experiments reported below, we argue that participants are only able to successfully prepare for the target word in those cases in which they have both types of information from the constant sets. In variable sets, in which the number of shared segments is invariable but the syllabic

[2] However, contrary to Meyer's (1991) results, Roelofs (1996, Experiment 6), showed that the size of the preparation effect depends on the length of the shared syllable in terms of number of segments.

structure is different for one item in the response set, we argue that participants are not able to prepare for *any* members of that response set as they cannot predict the next upcoming syllable.

To ensure that it is in fact the case that not only the odd-man-out might be excluded from a preparation mechanism, the odd-man-out itself is excluded from the analysis. As already mentioned, we predict that the odd-man-out hinders participants from fully preparing for any of the first syllables within the response set; thus we should be able to find the effect even after exclusion of the odd-man-out.

Experiment 1: Production of CVV targets

In order to test if the emergence of the syllable is traceable in the preparation of spoken Dutch words, two experiments were carried out using the odd-man-out variant of the implicit priming paradigm. Under the assumption that the syllable indeed represents a processing unit at the phonology/phonetics interface, the odd-man-out should reduce the preparation effect in the variable set as compared to the constant set. The odd-man-out, with a syllable structure that differs from the other members in the set, is expected to spoil the preparation effect, not only for the odd-man-out, but also for the other set members. However, if speakers just need the segmental information for preparation, the reaction times for constant and variable sets should show no difference.

Method

Participants

Twenty-four native speakers of Dutch participated in the first experiment. They were randomly taken from the pool of participants of the Max Planck Institute in Nijmegen, The Netherlands and were paid for their participation.

Materials

Eight different Dutch verb stems served as base for constructing eight different experimental blocks (four verb stems for the constant and four verb stems for the variable sets). The verb stems within each set were presented in three different inflectional forms plus the corresponding noun, e.g., the infinitive form of the verb stem 'leid-' *lei.den* (CVV; [to] lead), the corresponding noun *lei.der* (CVV; leader), the gerund *lei.dend* (CVV; leading), and the past tense form *lei.dde* (CVV; led). As all response words are derived from the same verb stem, segmental overlap within each set was assured. In constant sets, the first syllable of all four words in the set had the same syllable structure, as in the '*lei*.den'-set quoted above. The variable sets were constructed in the same way, but the past tense form of the verb stems selected for these sets had a different syllable structure compared to the other members in their sets, e.g., *ro.ken* (CVV; [to] smoke), *ro.ker* (CVV; smoker), *ro.kend* (CVV; smoking), but *ROOK.te* (CVVC; smoked).[3] Thus, all items in the constant and the variable sets have the same syllable structure in the first syllable with the exception of the past tense form in variable sets; this item with the deviating syllable structure served as odd-man-out. A full list of items is given in Appendices A and B. Prompt words for each target word in each condition were derived from the strong verb *staan* ([to] stand), *staander* (stand), *staande* (standing), and *stond* (stood). The use of this irregular verb with its four (irregular) inflectional forms guaranteed that there was no form overlap between prompt and target (as would have been the case by using a regular verb, e.g., wer.*ken*—ro.*ken*). The fact that the same prompt was used for all target words allowed for maximum comparability within and between sets and also simplified the participant's learning task.

Design

Constant and variable sets were presented in a four and in a three-item condition. The four-item sets contained all of the items mentioned above, whereas in the three-item sets the past tense form, i.e., the one that caused the odd-man-out in the variable set but not in the constant set, was excluded (see Table 1). Although the syllable structure is therefore constant in the three-item sets, we will denote the pair of a constant four-item set and its three-item derivative set the 'constant condition,' and the pair of a variable four-item set and its three-item derivative the 'variable condition.' Hence, we crossed two factors: a factor "Word Type" (opposing the constant and variable conditions) and a factor "Item Set" (opposing the four- and three-item sets). In the data analysis, the past tense form in the four-item sets was also excluded. The voice onset latencies of the remaining three forms were compared to those of their corresponding three-item sets. As the odd-man-out is expected to spoil the preparation effect for the whole set, the original four-item set in the variable conditions should show larger latencies compared to their three-item sets as well as to the other constant sets. The resulting advantage of presenting the response words in different set sizes, i.e., three- versus four-item sets rather than in heterogeneous and homogeneous sets, as in the original version of the implicit priming paradigm, is that the two types of sets are more comparable to each other because we did not present different lemmas in the heterogeneous sets (as would happen by regrouping items from the homogeneous sets) but only items which are

[3] Vowels in open syllables, as in *ro*.ken and vowels marked twice in orthography, as in *rook.te*, both have a long pronunciation in Dutch.

Table 1
Response set and prompts within a constant and a variable set in a three- and a four-item set

	Word Type			
	Constant sets		Variable sets	
Prompts	Four-item set	Three-item set	Four-item set	Three-item set
staan ([to] stand)	lei.den ([to] lead)	lei.den	ro.ken ([to] smoke)	ro.ken
stond (stood)	lei.dde (led)		rook.te (smoked)	
staander (stand)	lei.der (leader)	lei.der	ro.ker (smoker)	ro.ker
staande (standing)	lei.dend (leading)	lei.dend	ro.kend (smoking)	ro.kend

derived from the same stem. The two-by-two design was within-subjects. Each participant was presented four out of eight experimental sets, two sets being variable and two being constant.

Procedure and apparatus

The participants were tested individually in a quiet room. They were given detailed written instruction specifying that they had to respond as accurately and as quickly as possible. The experiment consisted of alternating learning and test phases. In the learning phase, participants were shown the four (or three) pairs of prompt–response words of a set on the computer screen (NEC Multisync3FG). When they indicated that they had studied the pairs sufficiently, the experimenter started the practice phase in which participants saw all four prompts together on the computer screen and they had to produce the corresponding responses in a row. When they failed, the learning phase was started again and the session was rehearsed to ensure that they learned the sets accurately. When they successfully completed the practice phase, the experimenter started the test phase. Each trial started with an attention sign (asterisk) marking the position of the prompt. The asterisk was displayed for 500 ms and after a pause the prompt was presented. Four different presentation times (after a 350/600/850/1300 ms pause) equally distributed across conditions were chosen to prevent speakers from producing the prepared target onsets before the prompt was actually displayed. Simultaneously with prompt presentation the voice key was activated for 1500 ms. The prompt disappeared after the response with a delay of 500 ms. The asterisk of the next trial appeared after 100 ms. Prompts within each set were repeated five times in a random order, resulting in response sets of 15 items for the three-item sets, and 20 items for the four-item sets.

The presentation of the stimuli and the measuring of the reaction times were controlled by the NESU software package. The spoken reactions were registered by a Sennheiser MD211N microphone, which fed into a NESU-box voice key device and a DAT recorder (Sony DTC-55ES). The experimenter sat in the same room and took note of hesitations, voice key errors, wrong naming responses, and time outs. After the completion of each item set, the number of successful trials and the corresponding mean reaction time were displayed on the participant's screen. The total duration of the experiment varied as a function of participant's learning time. On average, an experimental session lasted for 30 min.

Results

Only those test items for which a correct response was obtained were included in the reaction time analysis. Test items leading to wrong or invalid responses were not included (all wrong naming responses, voice key errors, and hesitations). Time outs ($>$1500 ms) and extreme outliers (i.e., naming latencies shorter than 300 ms) were also removed. Two participants were excluded from the analysis because of high error rates (more than 20% errors). The mean voice onset latencies, standard deviations, error rates, and preparation effects are summarized in Table 2.

Analyses of variance were run with Set Size (three-item sets versus four-item sets) and Word Type (constant versus variable) as independent variables. As mentioned before, we excluded the past tense form in the four-item sets from the analyses and compared only the remaining three forms (infinitive, noun, and gerund) to their corresponding three-item sets. The term 'four-item sets' always refers to the original four-item sets with the excluded past tense form to distinguish them from the corresponding three-item sets.

Error rates

In this experiment, there were 3.1% trials excluded altogether. As none of the main effects or interactions were significant, the error analysis is not reported.

Reaction times

As a first result, we expected an effect of Set Size, i.e., the three-item sets were expected to be produced faster compared to their four-item sets. The reduced set-size was expected to make it easier for the participants to recall the items, thus resulting in shorter voice onset latencies. This is confirmed by the data. Participants responded on average 58 ms faster in the three-item sets compared to their four-item sets. The main effect of Set

Table 2
Mean voice onset latencies (in ms), standard deviations, percentage errors (in parentheses), and preparation effects (in ms) in Experiment 1

Word Type	Set Size						Preparation effects
	Three-item sets			Four-item sets			
	M	SD	% Err	M	SD	% Err	
Constant	606	152	(5.0)	631	163	(2.4)	25
Variable	595	170	(0.3)	686	182	(4.7)	91

Size was significant ($F_1(1,21) = 53.12$, $MS_e = 1463.17$, $p < .01$; $F_2(1,22) = 147.89$, $MS_e = 281.82$, $p < .01$). The main effect of Word Type (variable versus constant sets) was significant by participants but not by items ($F_1(1,21) = 8.87$, $MS_e = 1873.97$, $p < .01$; $F_2(1,22) < 1$).

The crucial prediction tested in this experiment was that of a significant interaction between the factors Word Type and Set Size. More precisely, beside the fact that the voice onset latencies in the four-item sets were larger because of the additional item, the variable four-item sets were predicted to be slower in comparison to the constant four-item sets due to the odd-man-out. We predicted the difference between the three-item sets and the four-item sets to be larger in the variable condition than in the constant condition. This prediction was confirmed by the data. The preparation effect (the difference between the production latencies of the three- and the four-item sets within one condition) was much larger in the variable condition (686 – 595 ms = 91 ms) than in the constant condition (631 – 606 ms = 25 ms). The interaction of Word Type and Set Size was marginally significant for participants ($F_1(1,21) = 4.14$, $MS_e = 4925.09$, $p = .055$) and highly significant for items ($F_2(1,22) = 45.48$, $MS_e = 281.89$, $p < .01$).

Discussion

The results of Experiment 1 show that there is in fact a preparation effect for the syllable. The effect of 66 ms can be attributed to a syllable structure effect since the overlap of initial segments is the same in constant and in variable sets. Thus, the larger onset latencies in the variable four-item sets can only be explained by the change of the syllabic structure in those sets. As we excluded the past tense forms from the four-item sets and analyzed only the remaining three forms, the effect cannot be due to the odd-man-out itself, the item with a deviating syllable structure in variable sets. As we expected, the odd-man-out spoils the preparation effect for the whole (variable) set.

The syllable structure, CV(V), that we investigated in this experiment is the most basic and simplest syllable structure universally. According to phonological theory, syllables universally 'prefer' to have simple onsets and no coda (Hooper, 1972; Selkirk, 1982; Vennemann, 1988). Therefore it is possible that the preparation effect we found is specific to CV-syllables and will not generalize to typologically more complex syllables. That is why we decided to try and replicate the obtained preparation effect for syllables with different properties, long vowels and complex onset clusters. Specifically, we constructed sets of target items beginning with CCVV-syllables.

Experiment 2: Production of CCVV targets

The second experiment tested the same predictions as the previous experiment with different participants and materials. Materials in this experiment also consisted of bisyllabic Dutch words but contained word stems which had a different syllable structure. In this experiment, all words except the odd-man-out shared a CCVV-structure as first syllable as in *pra.ten* (CCVV; [to] speak) or *sle.-pen* (CCVV; [to] drag), the syllable structure of the first syllable for the odd-man-out was CCVVC as in *sleep.te* (dragged).[4] Design, procedure, and apparatus of Experiment 2 were the same as in Experiment 1.

Method

Participants

Twenty-four undergraduate students from the same population described in the previous experiment participated in Experiment 2.

Results

Only those test items for which a correct response was obtained were included in the reaction time analysis. Test items leading to wrong or invalid responses were not included (all wrong naming responses, voice key errors, and hesitations). Time outs (>1500 ms) and extreme outliers (i.e., naming latencies shorter than 300 ms) were also removed. The mean voice onset latencies, standard deviations, error percentages, and preparation effects are summarized in Table 3.

[4] Please note, that as in the roken/rookte example, vowels in open syllables, e.g., *sle.pen* and vowels marked twice in orthography, e.g., *slee.pte*, both have the same long pronunciation in Dutch.

Table 3
Mean voice onset latencies (in ms), standard deviations, percentage errors (in parentheses), and preparation effects (in ms) in Experiment 2

Word Type	Set Size						Preparation effects
	Three-item sets			Four-item sets			
	M	SD	% Err	M	SD	% Err	
Constant	601	160	(2.9)	633	174	(3.2)	32
Variable	582	164	(3.6)	675	197	(4.7)	93

Analyses of variance were run again with Set Size (three-item sets versus four-item sets) and Word Type (constant versus variable) as independent variables.

Error rates

In this experiment, 3.6% of the trials were errors. None of the main effects or interactions were significant.

Reaction times

There was a difference of 62 ms between the three-item sets and the four-item sets. This difference was significant for the factor Set Size by participants and by items ($F_1(1,23) = 48.29$, $MS_e = 1966.95$, $p < .01$; $F_2(1,22) = 128.3$, $MS_e = 369.51$, $p < .01$). The main effect of Word Type was not significant ($F_1(1,23) = 3.4$, $MS_e = 1139.52$, n.s.; $F_2(1,22) < 1$).

The preparation effect (the difference between the production latencies of the three- and the four-item sets within one condition) was much larger in the variable condition (674 − 582 ms = 93 ms) than in the constant condition (633 − 601 ms = 32 ms). The interaction of Word Type and Set Size was just significant for participants ($F_1(1,23) = 4.27$, $MS_e = 5806.93$, $p = .05$) and highly significant for items ($F_2(1,22) = 30.7$, $MS_e = 369.51$, $p < .01$).

Discussion

Experiment 2 again confirms the prediction of a preparation effect and replicates the effect of Experiment 1. Participants produced longer voice onset latencies in the variable four-item sets than in the three-item sets. As in the first experiment, this result can be interpreted as a syllable structure effect, because the overlap of the beginning segments was the same for variable and constant sets. The difference between the constant and variable three-item sets in this experiment showed the same pattern as in the first experiment namely that the variable three-item sets yield faster reaction times than the constant three-item sets. In Experiment 1, the reaction times in the variable three-item sets were on average 11 ms faster in comparison to the constant three-item sets (606 − 595 ms); in Experiment 2 this difference was even stronger. Here, participants produced the variable three-item sets on average 19 ms faster than the constant three-item sets. Thus, the longer response latencies in the variable four-item sets cannot be attributed to the fact that the variable item sets were more difficult to learn or to produce. Consequently, the larger reaction times in the variable four-item sets must be explained by the syllabic change in those sets. In other words, the odd-man-out spoiled the preparation effect for the whole set.

As the crucial effect was only marginally significant for participants in the first experiment, we decided to enhance test power by combining the data of both experiments in one ANOVA, with Experiment as a between-subjects factor. This collapsed analysis shows that the effect is reliable, also for subjects. The main effect of Set Size was significant ($F_1(1,45) = 102.04$, $MS_e = 1691.207$, $p < .01$; $F_2(1,46) = 283.42$, $MS_e = 313.99$, $p < .01$). The main effect for Word Type is significant by participants but not by items ($F_1(1,45) = 11.89$, $MS_e = 1512.63$, $p < .01$; $F_2(1,46) = 1.65$, $MS_e = 4408.40$, n.s.). The interaction of Word Type and Set Size was significant for both participants ($F_1(1,45) = 8.57$, $MS_e = 5267.09$, $p < .01$) and items ($F_2(1,46) = 76.88$, $MS_e = 313.99$, $p < .01$). Neither the main effect nor any of the interactions with the factor Experiment were significant.

Experiments 1 and 2 demonstrate that response times in the implicit priming paradigm are faster when speakers have advanced knowledge about both segmental and syllabic content of an upcoming word. Prior studies suggest that increased segmental overlap leads to larger preparation effects (Meyer, 1991) and thus we believe that the variable sets in the present experiments were being prepared by participants but not to the same extent as the constant sets. However, this needs to be empirically demonstrated.

To test this claim, we carried out two control-experiments in which we contrasted the constant and variable (four-item) sets both comprising initial segmental overlap to sets where there was no overlap of initial segments across items. This experimental design corresponds to the classic version of the implicit priming paradigm (Meyer, 1991) where homogeneous sets are compared to heterogeneous sets. The same materials as in Experiments 1 and 2 were used to conduct Control-Experiments

Table 4
Example for a homogeneous and a heterogeneous set within a constant and a variable set

Word Type			
Constant sets		Variable sets	
Homogeneous set	Heterogeneous set	Homogeneous set	Heterogeneous set
lei.den ([to] lead)	lei.den ([to] lead)	ro.ken ([to] smoke)	ro.ken ([to] smoke)
lei.dde (led)	haa.tte (hated)	rook.te (smoked)	huil.de (cried)
lei.der (leader)	po.ter (person who plants)	ro.ker (smoker)	boe.ner (person who polishes)
lei.dend (leading)	wa.dend (wading)	ro.kend (smoking)	ha.kend (croching)

Note. Heterogeneous sets are created by regrouping the response words from the four different homogeneous sets (for a full list of homogeneous item sets please see item lists of Experiment 1 for Control-Exp. 1 and item list of Experiment 2 for Control-Exp. 2, respectively, in Appendices A and B).

1 and 2. The former constant and variable four-item sets served as (constant and variable) homogeneous sets. The heterogeneous sets were created by regrouping the items from the homogeneous sets such that no two words in a set shared word onsets. See Table 4 for an example of a homogeneous and a heterogeneous set.

A preparation effect for the homogeneous sets compared to their heterogeneous sets (independent of whether they are constant or variable) is expected as all items share phonological properties, i.e., the first segments, and allow for advanced construction of the phonological word. No preparation is possible in heterogeneous sets as they consist of all different items.

To summarize the results of these control-experiments, we found an overall preparation effect for homogeneous versus heterogeneous sets in both experiments, confirming that segmental overlap, independent of syllable structure, leads to a preparation effect. The two homogeneous sets yielded in both control-experiments faster reaction times than their corresponding heterogeneous sets (for Control-Exp. 1: constant and variable *homogeneous* sets: 414 and 456 ms, constant and variable *heterogeneous* sets: 846 and 818 ms; for Control-Exp. 2: constant and variable *homogeneous* sets: 424 and 445 ms, constant and variable *heterogeneous* sets: 853 and 829 ms).

By subtracting the mean response time for the homogeneous *variable* set from the heterogeneous *variable* set (for Control-Exp. 1: 456–818 ms; for Control-Exp. 2: 445–829 ms) we can see the magnitude of the segmental preparation effect in the absence of syllabic overlap (362 ms for Control-Exp. 1; 384 ms for Control-Exp. 2). The same calculation for the constant sets (homogeneous sets minus their corresponding heterogeneous sets: for Control-Exp. 1: 414–846 ms; for Control-Exp. 2: 424–853 ms) shows the preparation effects in case of segmental *and* syllabic overlap (for Control-Exp. 1: 432 ms; for Control-Exp. 2: 429 ms). Thus, if we further subtract the preparation effect for the constant sets from the preparation effect for the variable sets we can determine the additional preparation benefit provided by the constant syllabic structure (for Control-Exp. 1: $432 - 362\,\text{ms} = 70\,\text{ms}$; for Control-Exp. 2: $429 - 384\,\text{ms} = 45\,\text{ms}$). These syllabic effects replicate the syllable preparation effects of Experiments 1 and 2.

The effect for the segmental preparation is much larger than the effect for syllabic preparation (approximately 400 ms versus approximately 58 ms). However, it is unlikely that these large differences are due to segmental overlap alone. Please note that there are four different lemmas in heterogeneous sets which have to be learned and recalled whereas in homogeneous sets there is only one single verb stem in four different inflectional forms. Thus, the learning and processing load in heterogeneous sets was four times as great. Furthermore, responses in heterogeneous sets were additionally hampered by the rotation of the different inflectional forms of a single verb stem across sets (e.g., in constant set 1, participants learn *leiden* as the infinitive form but in the subsequent set, they learn that the same verb must now be produced in the past tense, *leidde*).

Finally, we can additionally extract from the data of the control-experiments the following: under the assumption that a shared phonological property within one response set should lead to faster reaction times, one could predict that the heterogeneous constant sets should be faster compared to their variable counterparts because they share (as in their homogeneous sets) an abstract syllabic structure (not the segments) and in variable sets they do not. But the opposite is the case: heterogeneous *variable* sets (818 ms for Control-Exp. 1; 829 ms for Control-Exp. 2) are in both control-experiments faster than the *constant* sets (847 ms for Control-Exp. 1; 852 ms for Control-Exp. 2). This shows once again (see also Roelofs & Meyer, 1998) that a shared abstract syllabic structure (CV-structure) without segmental overlap does not give rise to benefit in preparing for a target word.

General discussion

Two experiments were reported that investigated the role of the syllable in the process of spoken word production. The implicit priming paradigm seems to be an appropriate method to tap into the late processes of syllabification and syllabary access. In the experiments, participants produced previously learned target words repeatedly within an experimental block. As we chose item sets in which the response words were all derived from the same word stem, segmental overlap within all sets was provided. The homogeneity of segmental overlap in constant and in variable sets was a crucial requirement to test whether, in addition to segmental information, speakers use information about the syllabic structure in order to prepare the response. If speakers only prepare for the segmental structure of the target word, there should have been no difference between the reaction times in constant and variable sets. However, this is not what we found. In Experiment 1, we investigated items with a constant syllable structure of the form CVV in the first syllable, while the odd-man-out in variable sets had a deviating syllable structure in the first syllable, namely a CVVC-structure. The 'variable' three-item sets—which contained the items of the variable four-item sets, minus their odd-man-out—yielded reaction times that were on average shorter than the 'constant' three-item sets. Thus, the larger reaction times in the variable four-item sets (in comparison to constant four-item sets) cannot be explained by the nature of the items in variable sets themselves, i.e., those items being more complex or more difficult to learn or produce than those in constant sets. The same holds for the findings in Experiment 2, where we investigated items with a constant CCVV-structure in the first syllable and a variable syllabic structure, which consisted of a CCVVC-structure. Here, variable three-item sets were on average even shorter compared to their counterparts in constant sets.

Another difference between the items in the constant and the variable sets lies not in the syllable structure of the past tense forms of the verbs but in the complexity of the segmental transition between the two syllables. Past tense forms in variable four-item sets consist of a consonant–consonant between-syllable sequence, e.g., *klaa***g.d***e*, that may be inherently more difficult to articulate than the vowel–consonant sequence found in the constant four-item sets, e.g., *knee.***dd***e*. This difference is an unavoidable characteristic of the nature of the present stimulus set. But while it may be that the words containing the more complex consonant clusters are harder to articulate than the words with the simpler transition and that this difference might lead to a response time difference between the two types of past tense forms, this possible difference cannot account for the longer reaction times of the *whole* (four-item) variable set. Remember that all the past tense forms are excluded from the analysis and therefore response time differences to the two types of past tense forms did not contribute to the response time difference between the variable and constant sets.

Another possibility is that the reported effects could be attributed to general memory retrieval processes rather than processes specific to speech production. Specifically, since items in the constant sets have overlap at multiple tiers (i.e., the segmental and the syllabic levels), they may be easier to retrieve from memory than items in the variable sets. However, in contrast to findings using the implicit priming paradigm, findings from immediate serial recall tasks report *slower* response times when items share phonological or phonetic features (see Baddeley, 1997, for a review). Thus, it is unlikely that the present results are due to memory effects.

Rather, the most plausible explanation for the present results is that the odd-man-out, the one variable item with the deviating syllable structure, spoiled the preparation effect for its whole (variable four) item set. Speakers could no longer prepare for the target word's first syllable. Still, we have to consider what explicitly happens in the implicit priming paradigm and which mechanisms of phonological and phonetic encoding (possibly involving syllabary access) contribute to the observed effect. In both constant and variable (four-item) sets, initial segments, which are constant across all items within a set, can be spelled-out from memory even before the prompt is displayed. However, only in the constant sets, in which the syllabic structure is also constant across items in a set, can the syllabification process begin to incrementally put the segments together and create the first syllable. This abstract phonological syllable can then be fed into the mental syllabary and activate its motor program. Thus, in constant sets, the first syllable is fully prepared for articulation. For variable sets, which lack a consistent syllable structure across items, the preparation cannot go beyond the retrieval of the initial segments. Only with the appearance of the prompt word does the further needed information about the appropriate syllable structure become available and only then the segments can be assembled for the first syllable.

Accordingly, we can conclude that, for constant sets, the first syllable can be fully prepared for articulation, including the retrieval of the corresponding gestural score from the mental syllabary. Thus, in constant sets, preparation can even go beyond on-line syllabification. The variable (four-item) sets can only be partially prepared for articulation (relative to the heterogeneous baselines), as can be concluded from the results of the two control-experiments. While advanced phonological encoding of the segmental information can occur, the absence of predictive syllabic structure prevents completion of phonetic encoding, namely, the retrieval of the corresponding gestural score from the mental syllabary.

An additional component of the observed effects may reside in the residual activation of the syllables in the mental syllabary. The retrieval time for the one gestural score required for the constant set may be faster due to the repeated selection of the same representation for all items within a set. But in variable sets, residual activation can only speed retrieval of the syllable half of the time (both the odd-man-out and the item immediately following the odd-man-out cannot benefit from the residual activation). While this may account for some of the observed differences, post-hoc analyses demonstrated that this is not the whole story. If only items which share the initial syllable with the immediately preceding trial are considered (thus allowing for a benefit of residual activation), the difference between the constant and variable sets remains. The remaining items in the constant four-item sets still show faster reaction times (633 ms, Experiment 1; 626 ms, Experiment 2) than the remaining items in the variable four-item sets (676 ms, Experiment; 670 ms, Experiment 2). Thus, even when the differences in residual activation of syllable representations in the mental syllabary are maximally matched, a syllable preparation effect is still observed. This post-hoc analysis demonstrates that all items in the constant sets benefit from the syllable overlap while none of the items in the variable do.

Using the implicit priming paradigm we successfully identified syllables as functional units in speech production. However, why was this task successful in finding a syllable effect when a decade of syllable priming studies failed? We would like to argue that the crucial difference lies in the explicit articulation of each word. Note that in all prior syllable priming studies, the prime was never overtly produced. If indeed syllables emerge late in speech production, primes that are not articulated may not reach the relevant stage in production where syllables are encoded, resulting in no priming effect.

Thus, our results confirmed the predictions made by WEAVER++. This model predicted that the two different tasks (priming versus preparation) affect different aspects of the process of phonological encoding. In syllable priming studies, the primes are assumed to speed up the segmental spell out, i.e., the moment when segments become available as part of the stored word form, which is retrieved from the mental lexicon. At the stage of segmental spell-out there is only segmental, but no syllabic information available according to the theory of lexical access proposed by Levelt et al. (1999; but see Dell, 1986, 1988). There can be no primed syllable retrieval. The finding that the magnitude of the priming effect increases with an increase of the number of shared segments, independent of a syllable match or mismatch with the target's first syllable, confirms the assumption that only shared segments can be primed. Actual syllable information such as syllable-internal positions for the current segments, i.e., which segment is assigned to the onset, the nucleus, or the coda position, only becomes available when phonological syllables induce the retrieval of the corresponding phonetic syllables from the mental syllabary.

The implicit priming paradigm allows for maximum preparation given the item set. In constant sets, where initial segments and also the syllabic structure of the first syllable are shared between items in one response set, the first syllable is fully prepared for articulation. Thus, all stages prior to articulation, including segmental spell-out, on-line syllabification and possibly access to the mental syllabary, can contribute to the preparation effect. In variable sets, responses can be prepared up to on-line syllabification or prosodification. Thus, all stages preceding on-line syllabification contribute to the preparation effect. In the current task, syllabic information is relevant in the sense that any deviating syllabic structure, i.e., an item with a different syllable structure, reduces the preparation effect. This account is confirmed—at least for a language like Dutch—by the results of the present study. While Dell's model can account for the results presented here, it can not account for the absence of syllable priming effects reported repeatedly in the literature (Schiller, 1997, 1998, 1999, 2000; Schiller et al., 2002).

However, as argued above, the assumption that syllabification is a late process may not hold cross-linguistically. In a recent study, Chen et al. (2002) investigated the process of word-form encoding in Mandarin Chinese by means of the (standard) implicit priming paradigm. In their experiments, they tested whether they could find preparation effects when the first syllable was shared, with or without shared tone. Indeed, they obtained preparation effects in both conditions, though the preparation effect was substantially stronger in case the tone was shared. The latter interaction is in line with findings reported by Roelofs and Meyer (1998). They tested for Dutch whether the standard first syllable priming effect requires that the target words share their stress pattern. Using iambic words (the non-default stress pattern in Dutch), they showed that an odd-man-out item (with trochaic stress pattern) entirely annihilated the preparation effect. This is in line with the Chinese data, except that if tone is not shared there is still a small preparation effect remaining. Chen et al. (2002) rightly raise the question what is exactly prepared when syllables are primed that do not share their tones. It cannot be the syllable's full articulatory gesture, because that must involve the tone. This can be constructed as another argument for whole syllable retrieval in Chinese. The remaining preparation effect would then be due to facilitation of retrieving a syllabified phonological code. Notice that this presupposes that in Mandarin Chinese syllables stored with the lexical item are not specified for tone; an item's tone pattern is, in some way, independently specified in lexical form memory. The retrieved tone pattern gets assigned to the appropriate phonological syllables during incremental syllabification. This is all highly speculative, but it shows that it is worth exploring the mechanism of syllabification in much more depth cross-linguistically.

To summarize, the different results found in Mandarin Chinese and Dutch (as well as in other Indo-European languages) could be due to the contrasting properties of the respective languages. In Mandarin Chinese, with a syllabary inventory of a much smaller size and no need for resyllabification, the syllabification process may be different in the sense that syllabic units are represented earlier in the process of speech production.

Conclusions

The results of the odd-man-out variant of the implicit priming task reported in this study show specific preparation of syllabic articulatory units is possible. Taking the results of prior syllable priming studies into account, we conclude that the used paradigm taps into the right level of processing where syllables are in fact encoded, i.e., the interface of phonological and phonetic encoding. This is in agreement with predictions from the spoken word production model by Levelt et al. (1999) and its computer simulation, WEAVER++ (Roelofs, 1996, 1997a, 1997b, 1998, 1999). The results from the Mandarin Chinese study (Chen et al., 2002) do not contradict the proposed syllabification process in Dutch, but rather suggest that a language user's word form encoding architecture will, at least in part, be tuned to specific requirements of target language phonology.

Acknowledgments

The present study emerged from an intensive two-day tutorial discussion meeting with Aditi Lahiri, which is still affecting much of our work on phonological and phonetic encoding. The authors also wish to thank Alissa Melinger and three anonymous reviewers for helpful comments.

Appendix A

Materials for Experiment 1

Four-item sets		Three-item sets	
Constant sets	Variable sets	Constant sets	Variable sets
lei.den ([to] lead)	hui.len (to cry)	lei.den	hui.len
lei.dde (led)	huil.de (cried)		
lei.der (leader)	hui.ler (person who cries)	lei.der	hui.ler
lei.dend (leading)	hui.lend (crying)	lei.dend	hui.lend
ha.ten ([to] hate)	boe.nen ([to] polish)	ha.ten	boe.nen
haa.tte (hated)	boen.de (polished)		
ha.ter (hater)	boe.ner (person who polishes)	ha.ter	boe.ner
ha.tend (hating)	boe.nend (polishing)	ha.tend	boe.nend
po.ten ([to] plant)	ro.ken ([to] smoke)	po.ten	ro.ken
poo.tte (planted)	rook.te (smoked)		
po.ter (person who plants)	ro.ker (smoker)	po.ter	ro.ker
po.tend (planting)	ro.kend (smoking)	po.tend	ro.kend
wa.den ([to] wade)	ha.ken ([to] crochet)	wa.den	ha.ken
waa.dde (waded)	haak.te (crocheted)		
wa.der (person who wades)	ha.ker (person who crochets)	wa.der	ha.ker
wa.dend (wading)	ha.kend (croching)	wa.dend	ha.kend

Note. Vowels in open syllables, e.g., ro.ken and vowels marked twice in orthography, e.g., roo.kte, both have a long pronunciation in Dutch.

Appendix B

Materials for Experiment 2

Four-item sets		Three-item sets	
Constant sets	Variable sets	Constant sets	Variable sets
kne.den ([to] knead)	kla.gen ([to] complain)	kne.den	kla.gen
knee.dde (kneaded)	klaag.de (complained)		
kne.der (kneader)	kla.ger (complainer)	kne.der	kla.ger
kne.dend (kneading)	kla.gend (complaining)	kne.dend	kla.gend
spui.en ([to] drain)	spoe.len ([to] rinse)	spui.en	spoe.len
spui.de (drained)	spoel.de (rinsed)		
spui.er (person who drains)	spoe.ler (person who rinses)	spui.er	spoe.ler
spui.end (draining)	spoe.lend (rinsing)	spui.end	spoe.lend
pra.ten ([to] speak)	dwa.len ([to] wander)	pra.ten	dwa.len
praa.tte (spoke)	dwaal.de (wandered)		
pra.ter (speaker)	dwa.ler (wanderer)	pra.ter	dwa.ler
pra.tend (speaking)	dwa.lend (wandering)	pra.tend	dwa.lend
plei.ten ([to] argue)	sle.pen ([to] drag)	plei.ten	sle.pen
plei.tte (argued)	sleep.te (dragged)		
plei.ter (arguer)	sle.per (person who drags)	plei.ter	sle.per
plei.tend (arguing)	sle.pend (draging)	plei.tend	sle.pend

Note. Vowels in open syllables, e.g., *sle.pen* and vowels marked twice in orthography, e.g., *slee.pte*, both have a long pronunciation in Dutch.

References

Baddeley, A. D. (1997). *Human memory: Theory and practice. (Revised ed.)*. Hove, UK: Psychology Press.

Bagemihl, B. (1995). Language games and related areas. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 697–712). Cambridge, MA: Blackwell.

Baumann, M. (1995). *The production of syllables in connected speech*. Ph.D. dissertation, Nijmegen University.

Berg, T. (1988). *Die Abbildung des Sprachproduktionsprozesses in einem Aktivationsflußmodell. Untersuchungen an englischen und deutschen Versprechern [The representation of the speech production process in a spreading activation model: Studies of German and English speech errors]*. Tübingen: Niemeyer.

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.

Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word form encoding in Mandarin Chinese as assessed by the implicit priming paradigm. *Journal of Memory and Language, 46*, 751–781.

Chen, J.-Y., Lin, W.-C., & Ferrand, L. (2003). Masked priming of the syllable in Mandarin Chinese speech production. *Chinese Journal of Psychology, 45*, 107–120.

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

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language, 27*, 124–142.

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature, 384*, 159–161.

Evinck, S. (1997). *Production de la parole en français: Investigation des unités impliquées dans l'encodage phonologique des mots* [Speech production in French: Investigation of the units implied during the phonological encoding of words]. Unpublished Ph.D. dissertation, Bruxelles University.

Ferrand, L., Segui, J., & Grainger, J. (1996). Masked priming of word and picture naming: The role of syllable 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, 25*, 458–470.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*, 27–52.

Fujimura, O., & Lovins, J. B. (1978). Syllables as concatenative phonetic units. In A. Bell & J. B. Hooper (Eds.), *Syllables and segments* (pp. 107–130). Amsterdam: North-Holland.

Hooper, J. B. (1972). The syllable in phonological theory. *Language, 48*, 525–540.

Indefrey, P., & Levelt, W. J. M. (2000). The neural correlates of languages production. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 845–865). Cambridge, MA: MIT Press.

Janssen, P. D., Roelofs, A., & Levelt, W. J. M. (2002). Inflectional frames in language production. *Language and Cognitive Processes, 17*, 209–236.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 824–843.

Kenstowicz, M. (1994). *Phonology in generative grammar*. Cambridge, MA: Blackwell.

Kerzel, D., & Bekkering, H. (2000). Motor activation from visible speech: Evidence from stimulus–response compatibility. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 634–647.

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. 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.

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 217–245). New York: Springer.

Luce, R. A. (1959). *Individual choice behavior*. New York: Wiley.

MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia, 8*, 323–350.

Meijer, P. J. A. (1996). Suprasegmental structures in phonological encoding: The CV-structure. *Journal of Memory and Language, 35*, 840–853.

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.

Meyer, A. S. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition, 42*, 181–211.

Nooteboom, S. (1969). The tongue slips into pattern. In A. G. Sciarone, A. J. von Essen, & A. A. van Raad (Eds.), *Nomen: Leyden studies in linguistics and phonetics* (pp. 114–132). The Hague: Mouton.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology, 17*, 273–281.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*, 107–142.

Roelofs, A. (1996). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language, 35*, 854–876.

Roelofs, A. (1997a). Syllabification in speech production: Evaluation of WEAVER. *Language and Cognitive Processes, 12*, 657–693.

Roelofs, A. (1997b). The WEAVER model of word-form 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. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes, 14*, 173–200.

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

Schiller, N.O. (1997). *The role of the syllable in speech production. Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography*. PhD dissertation, Nijmegen University (MPI series; 2).

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

Schiller, N. O. (1999). Masked syllable priming of English nouns. *Brain and Language, 68*, 300–305.

Schiller, N. O. (2000). Single word production in English: The role of subsyllabic units during speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*, 512–528.

Schiller, N.O., & Costa, A. (submitted). The role of the syllable in phonological encoding: Evidence from masked priming?

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.), *Papers in Laboratory Phonology 7* (pp. 35–59). Berlin: Mouton de Gruyter.

Schiller, N.O., Fikkert, P., & Levelt, C.C. (in press). 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., Meyer, A. S., & Levelt, W. J. M. (1997). The syllabic structure of spoken words: Evidence form the syllabification of intervocalic consonants. *Language and Speech, 40*, 103–140.

Selkirk, E.O. (1982). The syllable. In: H. van der Hulst & N. Smith, (Eds.), *The structure of phonological representations*. Part II. Dordrecht: Foris, pp. 337–383.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial ordering mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing* (pp. 295–342). New York: Halsted Press.

Shattuck-Hufnagel, S. (1987). The role of word onset consonants in speech production planning: New evidence from speech error patterns. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processing in language* (pp. 17–51). Hillsdale, NJ: Erlbaum.

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

Stemberger, J. P. (1982). The nature of segments in the lexicon: Evidence from speech errors. *Lingua, 56*, 235–259.

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition, 15*, 49–74.

Treiman, R., & Danis, C. (1988). Syllabification of intervocalic consonants. *Journal of Memory and Language, 27*, 87–104.

Vennemann, T. (1988). *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin*. Berlin: Mouton de Gruyter.