(Re)syllabification Across Word Boundaries: Psycholinguistic Evidence From Dutch Clitics

Monika Baumann
University of Hamburg[1]

1. Introduction

This paper deals with the question of how the syllabic structure of encliticized forms is produced during speaking. In connected speech, the right boundaries of lexical words do not always align with the end of a syllable. In Dutch, for instance, many function words have two forms, one being phonologically strong and the other phonologically weak. The phonologically strong form contains a full vowel (e.g., *het* "it" [h∈t], *hem* "him" [h∈m], *en* "and" [∈n]), while the corresponding weak form normally has only schwa as a vowel (e.g., *het* "it" [∂t], *hem* "him" [∂m], *en* "and" [∂n]). Note that prosodic words never start with a schwa in Dutch and never have exclusively schwa as a vowel (Booij, 1996), while many of the weak forms show these properties. To avoid a schwa-initial syllable, and in accordance with the general tendency of languages to avoid syllables that lack an onset, a schwa-initial weak form of a function word (henceforth "clitic") will prefer having a coda element of the preceding word in its onset position, as shown in (1). Following the analyses by Gussenhoven (1985), Lahiri et al. (1990), and Booij (1996), I'll assume that the clitic is prosodically integrated into the preceding prosodic word.[2]

(1)	(ko:)$_\sigma$(k∂t)$_\sigma$	*kook het*	"cook it"
	(daŋ)$_\sigma$(k∂m)$_\sigma$	*dank hem*	"thank him"
	(bo:)$_\sigma$(t∂n)$_\sigma$	*boot en*	"boat and"

[1]This paper is based on my dissertation *The production of syllables in connected speech* (1995, Unpublished Doctoral Dissertation, University of Nijmegen). The research was supported by a grant from the Max-Planck-Gesellschaft zur Förderung der Wissenschaften, Munich (Germany) and carried out at the Max Planck Institute for Psycholinguistics, Nijmegen (The Netherlands). I owe many thanks to my supervisors Pim Levelt and Antje Meyer. Herbert Baumann, Daan Broeder, Ger Desserjer, John Nagengast and Johan Weustink provided valuable technical support for the experiments. I thank Antje Meyer for useful comments on this paper. Part of the work has been published in A. Dainora et al. (eds.), *Papers from the 31st Regional Meeting of the Chicago Linguistic Society, 1995, Volume 2: The Parasession on Clitics*, 50-63.

[2]Whether the enclitic syllable is incorporated into the preceding Foot, or chomsky-adjoined to it, or is immediately linked to the prosodic word node, is not crucial for my argument (see Booij, 1996, for a detailed discussion). Important is that the clitic belongs to the preceding word. This claim can be defended, since the rules that apply obligatorily within prosodic words also apply obligatorily in host+clitic combinations. An example is the rule of homorganic glide insertion to avoid hiatus in Dutch, as in *Ruanda* [ruwαnda] or *knieën* [knij∂n] "knies" (Gussenhoven, 1980). This rule also applies to encliticized forms: *zie het* "see it" [zi:j∂t] (Booij, 1995).

While most models of speech production deal with the encoding of single words, Levelt's (1992, 1993) model of phonological encoding accounts also for the production of encliticized forms like those in (1). According to the model, a speaker does not produce the syllables of the single lexical words when generating an encliticized form like $(\text{ko:k})_\sigma$ and $(\partial\text{t})_\sigma$ for *kook het*. The only syllabic structure produced is the postlexical surface structure, see (1). The results of experiments on encliticized forms in Dutch reported below in section 4, however, suggest that this claim may need modification. The results are not only relevant for models of phonological encoding in language production, but also for phonological theory, as will become apparent from section 5.

2. Phonological Encoding in Speech Production

A speaker produces on average three to five syllables per second, perhaps even more (Lenneberg, 1967). How those syllables are produced during speaking is an intricate issue that has become more and more important in the literature. Before looking at the special case of syllabification in encliticized forms, we have to ask how a speaker produces syllabic structure in general. Syllabification is one aspect of the encoding of a word's phonological form, which is again part of the general issue of how words are accessed from the lexicon when we speak.

2.1 Speech Errors and Models of Phonological Encoding

In the last two decades several models of language production have been developed (for instance, Fromkin, 1973; Garrett, 1975; Shattuck-Hufnagel, 1979; Dell, 1986, 1988; Levelt, 1989). In all models of language production mentioned above, lexical access consists of two parts, although the terminology used may be different. First, appropriate lexical items are selected and semantic and syntactic relations are created on the basis of the intended message. Second, the lexical units are phonologically encoded. We will look at this second process in more detail.

All models of phonological encoding have been heavily influenced by speech error data. The idea is that the way a system breaks down can provide insight into the way it works. The data are either naturally occurring errors, or errors that were elicited experimentally. Most phonological speech errors involve segments, which are exchanged, deleted, added, substituted, or shifted like in the sound exchange ***heft lemisphere*** (intended utterance: *left hemisphere*, example taken from Fromkin, 1973). The occurrence of those errors has lead researchers to believe that the phonological forms of words are not stored in the mental lexicon as units. Another argument against lexical storage of syllabified forms is that we would expect more errors that involve whole syllables as units, e.g. syllable exchanges. Those, however, are rare. Speakers seem to retrieve independently a word's segments and slots specifying positions within a syllable or a word and then combine the segments with the slots. When this process goes wrong, we encounter a speech error (see Meyer, 1992 for a detailed overview on the sound error literature).

Most models of speech production have been designed for the encoding of single words. However, connected speech is the output we normally produce. As mentioned above, postlexical syllable structure does not always coincide with the syllables of the single lexical words. This poses problems for models that specify syllabic positions in frames and label segments for syllabic positions, e.g., Shattuck-Hufnagel (1979) or Dell (1986) or positions within lexical words (Shattuck-Hufnagel 1992; Dell, 1988). These models could not explain why the stop in the coda of *kook* surfaces in onset position of the second syllable (k∂t)$_{\sigma}$ in (1), since it is specified as a coda segment of the verb *kook* that has to go into a coda position in the frame.

2.2 Levelt's Model of Phonological Encoding

A model designed to account for these forms is Levelt's (1992, 1993) outline of the phonological encoding component. He raises the question of why speakers should first partition a word's stored phonological form in a syllable frame and a string of segments, when they later unify them again. There should be another function for this partition than to provide speakers with the possibility to produce speech errors. Indeed, the separation of segments and frames to which the segments are then associated seems to be a useful concept to account for the production of encliticized forms in connected speech, where the surface syllables clearly do not correspond to the boundaries of lexical units: From the point of view of processing, it does not seem to be useful to first construct fully syllabified forms for individual lexical items that never appear in the output of the production process, but have to be resyllabified to account for the connected speech output. Instead, one would prefer to produce the postlexical surface syllables immediately.

(2) The Phonological Encoding Component of Levelt's Model:

activation of lexemes: <kook>, <het>$_{WEAK}$

separate segmental and metrical spellout procedures:	segmental /k,o:,k,∂,t/	metrical ω | σ' σ
prosodic word formation:		ω | \\ σ' σ
segment to frame association:		[(ko:)$_{\sigma}$(k∂t)$_{\sigma}$]$_{\omega}$

Levelt therefore proposed that a speaker creates frames of the size of a prosodic word as the basis of syllabification instead of single frames for individual lexical units, see (2). Two separate procedures called "segmental" and "metrical" work on each prosodic word. The segmental spellout procedure delivers an ordered sequence of segments (e.g. /k,o:,k,∂,t/ for *kook het*). Crucially, these segments are not marked for syllable positions. The second procedure, metrical spellout, makes available information on the number of syllables a word consist of, its prosodic word status, and its stress pattern. The metrical information of the single lexical words (e.g.: "monosyllabic prosodic word" for *kook*, and "monosyllabic word, no prosodic word" for the enclitic *het*) is integrated into one metrical frame for the whole prosodic word, resulting in a structure like (σ' σ)$_{\omega}$. This frame is then combined with the *sequence of segments in a process of segment-to-frame association, the result of which is* the surface syllable structure.[3] Producing separately an ordered string of segments not specified for syllable positions and frames for prosodic instead of lexical words has thus the advantage that only those syllables are produced that surface in speech output. This is appealing and elegant from an economical point of view.

A different though related claim concerns the time course of phonological encoding. Levelt's model predicts that syllables are produced at a late point of the production process, only after the metrical frames and the phonological segments have become available.

In section 4 I will report on experiments testing the claim that speakers do not produce underlying lexical syllables, but exclusively surface syllables. Section 3 describes the attempts to prove that syllables occur at a late point in phonological encoding.

3. The Time Course of Phonological Encoding

As research has shown in the past decade, the picture-word interference paradigm can be used to explore the time course of language production processes. The paradigm is based on the Stroop-task: Participants have no problem in reading aloud color terms that are printed in incongruent colors, for instance, the word *blue* printed in red ink. They are as good at this as at reading the color terms printed in black. However, participants have great difficulties in naming the colors of incongruent words, for instance, saying "red" when the word *blue* is written in red ink. This takes much longer than naming the color of a color block or a row of colored symbols.

The classic Stroop-task has been varied in several ways (see MacLeod, 1991, for an overview on research on the Stroop effect). In picture-word interference experiments, picture naming replaced color naming. Participants have to name pictures while interfering verbal information is presented. Interfering stimuli (IS) are either presented as written words superimposed on the picture (e.g., Rayner & Posnansky, 1978; Glaser & Düngelhoff, 1984) or they are presented auditorily (e.g., Schriefers, Meyer, & Levelt, 1990; Meyer & Schriefers, 1991). The IS can exhibit various relations to the pictures' names which affect response latencies. As compared to a neutral baseline (like a row of Xes in the visual or a rustle noise in the auditory domain), it takes longer to name, for instance, a picture of a sheep that is accompanied by a semantically related stimulus like *goat*. In contrast to that, a phonologically related stimulus like *sheet* speeds up response latencies as compared to the unrelated baseline (= phonological priming).

In addition to the relation between target and IS, the timing of the IS is important. IS in picture-word interference experiments can be presented at different points in time with respect to the appearance of the picture on the screen (= Stimulus Onset Asynchrony, SOA). Schriefers et al. (1990) found that semantically related IS slow down participants' response latencies when they are presented 150 ms before picture onset, whereas responses to phonologically related IS do not differ from responses in the neutral baseline condition when they are presented that early. However, when the IS are presented later (150 ms after

---

[3]Meyer (1991) showed that segmental spellout is produced in a segment-by-segment, left-to-right manner. Segment-to-frame association works from left to right, too, as has been shown by Meyer (1990).

picture onset), a different pattern emerges. At this later point in time, semantically related IS do not differ from the neutral baseline condition, but phonologically related IS speed up reaction times as compared to the baseline. These results can be taken as evidence for the fact that semantic processing in language production starts before phonological encoding, as most models of speech production assume.

For investigating the time course of syllabification during phonological encoding, IS were either phonlogically related to the target word that had to be produced, i.e. they corresponded to the target's first phonemes, or they were unrelated, i.e. they corresponded to the first phonemes of a different target. Participants produced verb forms as targets. Since verbs cannot be easiliy depicted, a semantic-associate learning task was used to elicit the verb forms. In this task (developed by Meyer, 1988, 1990, 1991), participants receive a sheet of paper with a list containing pais of words that are semantically related. They are instructed to learn these pairs by heart, in such a way that they are able to produce the second member of a pair (e.g., the verb *koken* "to cook") as soon as the first member (e.g., *eten* "meal") appears on the screen. In different blocks of the experiments, participants were instructed to produce either the verb's infinitive form, or the past tense form, or an encliticized form, where the verb was followed by the schwa-initial weak form of the pronoun *het* "it", see (3). While naming the verbs, participants heard the phonologically related or unrelated IS.

(3) <u>The Materials of the Priming Study:</u>

Targets:

Infinitive:	*koken*	$(\text{ko:})_\sigma(\text{k}\partial\text{n})_\sigma$	"to cook"
Encliticized:	*kook het*	$(\text{ko:})_\sigma(\text{k}\partial\text{t})_\sigma$	"cook it!"
Past Tense:	*kookte*	$(\text{ko:k})_\sigma(\text{t}\partial)_\sigma$	"cooked"

Interfering Stimuli (IS):

	Short	Long
Phonologically Related:	ko:	ko:k
Phonologically Unrelated:	le:	le:r

On the basis of the results of the earlier picture-word interference studies, we expect a phonological priming effect: Participants should respond faster if the IS are related to the target word than if the IS are unrelated. Furthermore, with respect to the time course of syllabification within phonological encoding, IS that are presented early during phonological encoding should show effects of segmental, but not yet of syllable production. The latter should occur at a later point in time. To test this, the phonologically related IS differed in length by one segment. As a result, they differed in the kind of their relation to the target: The IS either did or did not correspond to the target's first syllable. As shown in (3), long related IS corresponded to the first syllable of past tense forms, while short related IS matched the first syllable of infinitive and encliticized forms. To measure the amount of phonological priming that related IS provide independently from a general length effect that might be caused by the mere difference in length of the IS, the reaction times achieved with the related IS were subtracted from the reaction times that were obtained with unrelated IS of corresponding length.

Given these targets and IS conditions, the following predictions can be formulated on the basis of Levelt's model:

When the IS are presented at an early point in time, i.e. shortly before picture onset, no syllables should have been constructed yet, since syllabification is a late process in the model of phonological encoding. Therefore, the target word's syllable structure should not influence the results. Instead, one could expect that the more segments of the target are included in the interfering stimulus, the higher the priming effect of related IS should be, as compared to an unrelated baseline containing the same number of segments. Long IS should thus show higher priming effects than short ones, irrespective of the target's syllable structure.

When the IS are presented at a later point in time after picture onset, a speaker should be producing syllables. For this point in time we expect what can be called a 'syllable match effect': Related IS that coincinde with the target's first syllable should speed up reaction times more than related IS that do not coincide with the target's first syllable, both again as copared to an unrelated baseline of similar length. In other words, participants who produce infinitive or encliticized targets, which have a short first syllable, should show higher priming effects with short IS, which correspond to the target's first syllable, than with long ones. Participants who produce past tense target verbs, on the other hand, should show higher priming effects with long than with short IS.

Several experiments along this line have been run. Among others, the proportion of trials with related IS was varied. Another variation concerned the acoustic quality of the stimuli. In some experiments, the IS were spoken as syllables, in other experiments they were cut out of the target verb form that the subjects had to produce. Unfortunately, the results of the priming experiments did not reveal a pattern as clear as predicted (for a detailed discussion see Baumann, 1995). As expected, participants named the targets faster when they heard IS that were phonologically related to the targets than when they heard unrelated IS. This phonological priming effect indicated that the experimental manipulations tapped into the process of phonological encoding. Furthermore, participants generally reacted the faster the shorter the IS, irrespective of phonological relatedness. Participants thus seemed to be sensitive to the one segment difference between short and long IS. These robust effects were obtained in all experiments. Two control experiments furthermore showed that the results were not caused by morphological or lexical variables.

With respect to the role of the syllable, however, only one experiment yielded a pattern that could be interpreted as a syllable match effect. This experiment included infinitive and past tense targets, but no encliticized forms. The other experiments failed to show a syllable match effect. In sum, the experiments could not solve the issue of whether syllables are produced only at a late point during phonological encoding. The time course of syllabification needs further research.

4. Levels of Syllable Structure: Final Devoicing in Encliticized Forms

The second prediction of Levelt's model concerned the number of levels of syllable structure involved in phonological encoding: The only syllables constructed during speaking are those that do actually surface. For encliticized forms like *kook het*, this implies that a speaker never produces syllables that correspond to the single lexical items *kook* and *het*, but only the surface syllables (ko:)$_\sigma$ and (kət)$_\sigma$. How can we test this claim empirically?

Many languages have constraints for segments in coda position that do not hold for segments in syllable onsets. If we found effects of coda constraints for an obstruent which surfaces postlexically in onset position of the syllable that contains the schwa-initial function word, we would have to conclude that this obstruent has been in a coda position at an earlier stage of processing before it became an onset in surface syllable structure.

As some other languages, Dutch has only voiceless obstruents in syllable codas, while both voiced and voiceless obstruents occur in onsets, as shown in (4).[4]

(4)	*nies*	(ni:s)$_\sigma$	"sneeze (1st pers. sg)"
	niezen	(ni:)$_\sigma$(zən)$_\sigma$	"to sneeze"
	rood	(ro:t)$_\sigma$	"red"
	rode	(ro:)$_\sigma$(də)$_\sigma$	"red (inflected)"

As shown in (4), voicing is maintained in inflected or derived forms where the suffix starts with a vowel. For encliticized forms, it has been argued that final devoicing applies on the single lexical items, preceding postlexical resyllabification. In those forms, word-final obstruents are devoiced although they surface in onset position of the following syllable (e.g., Kooij, 1980; Booij, 1995, 1996).[5] Following this account, the encliticized forms should surface like the forms in (5a). The inflected forms show that the obstruents are underlyingly voiced: *binden* (bIn)$_\sigma$(dən)$_\sigma$ "to bind", *vrienden* (vri:n)$_\sigma$(dən)$_\sigma$ "friends". A monostratal account like Levelt's model, on the other hand, predicts the forms in (5b).

(5a)	*bind het*	(bIn)$_\sigma$(tət)$_\sigma$	"bind it"
	vriend en	(vri:n)$_\sigma$(tən)$_\sigma$	"friend and"
(5b)	*bind het*	(bIn)$_\sigma$(dət)$_\sigma$	"bind it"
	vriend en	(vri:n)$_\sigma$(dən)$_\sigma$	"friend and"

Since the surface syllable structure is the only syllabic structure created during the production process, the first word's final obstruent is never in syllable-final position, where it could be devoiced. The experiments reported below were run to test which account makes the right predictions.

---

[4]Devoicing applies syllable-finally: *ABVA*, which is an acronym for *Algemene Bond van Ambtenaren* "General Union of Civil servants" is pronounced *A*[p.f]*A* (Booij, 1995). Final devoicing is productive as becomes obvious when Dutch speakers speak languages that allow voiced obstruents in codas or pronounce foreign names like *Sy*[t.n]*ey* (Booij, 1977).

[5]Final devoicing in cliticized forms has been a subject for discussion, since final devoicing does not seem to be obligatory in some combinations of modal verbs and clitics in Dutch (see, Berendsen 1986, Booij & Rubach, 1987). For instance, *heb ik* ("have I") has two possible pronunciations (hɛ)$_\sigma$(bIk)$_\sigma$ and (hɛ)$_\sigma$(pIk)$_\sigma$. Berendsen accounts for this by different prosodic structures for the two forms, while Booij assumes that the voiced variant is stored as a unit in the lexicon. This discussion is not crucial for the present argument, since only full verbs were tested.

The empirical investigation of final devoicing in encliticized forms consisted of two sets of two experiments. Each set contained a production and a perception experiment. The aim was to obtain information about the voice quality of the final obstruent in forms like *raad en* ("commission and"), where the schwa-initial weak form of the conjunction *en* encliticizes to the preceding noun and the obstruent surfaces in onset position of the second syllable. The stimuli consisted of 13 minimal noun pairs that only differed in the underlying voice quality of their final stop like *raad* ("advice") and *raat* ("honeycomb").[6]

(6)	Singular		Plural		
raad	(ra:t)$_\sigma$	"commission"	*raden*	(ra:)$_\sigma$(dən)$_\sigma$	"commissions"
raat	(ra:t)$_\sigma$	"honeycomb"	*raten*	(ra:)$_\sigma$(tən)$_\sigma$	"honeycombs"

This is the complete set of monosyllabic minimal noun pairs varying in the voice quality of their final stops in Dutch. As shown in (6), devoicing neutralizes this difference in singular forms. In the plural forms, however, the voice quality of the stops is maintained, since they surface in onset position of the second syllable.

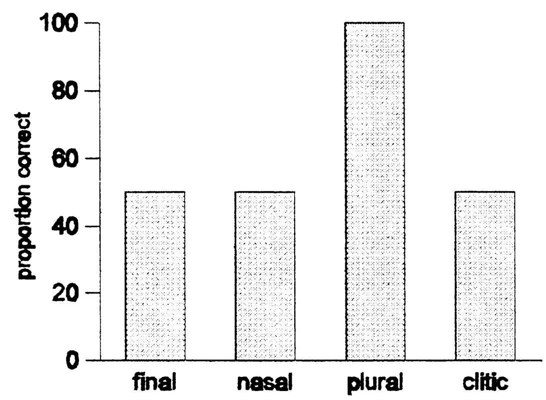

Figure 1a. Predictions of Resyllabification Theory for the Four Context Conditions

In the production experiment, participants produced sentences that contained the minimal pairs in different contexts. In the perception experiment, different participants heard the sentences that had been produced in the first experiment and had to perform a rating task on the voice quality of the critical obstruent.

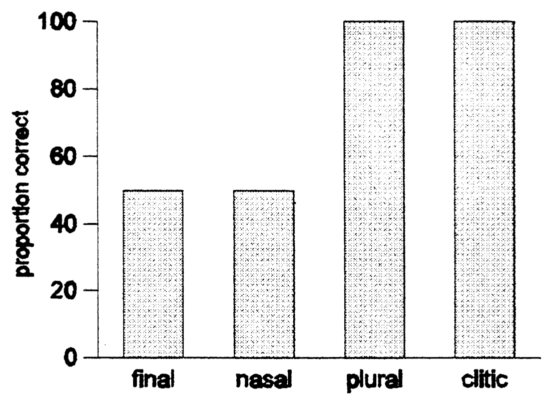

Figure 1b. Predictions of the Production Model for the Four Context Conditions

The noun could occur in four different contexts. In a 'final' context, the critical obstruent occurred sentence-finally. In a 'nasal' context, the obstruent was followed by a nasal consonant which could not form a phonotactically legal onset with the obstruent. In both contexts, final devoicing should apply obligatorily. Participants in the perception experiments were informed that in 50% of the cases they would hear the member of a pair ending in a "d" (or "b"), and in 50% of the cases the word

[6]Dutch does not have minimal pairs ending in fricatives, since voiced and voiceless fricatives are accompanied by different vowels: Vowels preceding underlyingly voiceless fricatives are lax, while vowels preceding voiced fricatives are tense.

ending in a "t" (or "p"). When the voice contrast is neutralized like in the final and nasal context, the percentage of assignment of "voiceless"- and "voiced"-responses should be at chance level. Figures 1a and 1b show the predictions that the two accounts make for the proportion of correct responses in the different context conditions.

In the nasal condition, in addition a phonetic voice assimilation effect from the nasal on the preceding stop was expected. The stop should be more voiced in the nasal context than in the final context in underlyingly voiced as well as in underlyingly voiceless targets. In a 'plural' condition, the minimal pairs should be clearly distinguishable, leading to a high percentage of correct responses. In a 'clitic' context condition, the nouns were followed by a schwa-initial function word. Monostratal accounts predict for this condition that participants in the perception experiment should be able to correctly distinguish nouns with underlyingly voiced from those with voiceless stops, because the stop is never in syllable-final position, where it could be devoiced. Following a theory that includes resyllabification, on the other hand, participants should perform at chance level, since final devoicing applies on the individual lexical words, preceding resyllabification in the clitic context.

4.1 The First Final Devoicing Study

4.1.1 The First Production Experiment

The aim of this experiment was to have participants produce stimuli that could later be used in the perception experiment. A delayed repetition task served to elicit the responses. A sentence appeared on the screen for a short period of time (1500 ms). Participants had to memorize it and to produce it in reaction to a visual prompt on the screen that was presented after a random pause of 500 to 1000 ms. The minimal pairs occurred in three different contexts (clitic, final, nasal), see (7).

(7) The Sentences to be Produced for the Minimal Pair *raad - raat*:

Context	Underlyingly	
clitic:	voiced	*Pien zegt "Ik zie een* ***raad en*** *een akker".* "Pien says 'I see a commission and a field'"
	voiceless	*Pien zegt "Ik zie een* ***raat en*** *een akte".* "Pien says 'I see a honeycomb and a file'"
nasal:	voiced	*Pien zegt "Ik zie een* ***raad naast*** *een akker".* "Pien says 'I see a commission near a field'"
	voiceless	*Pien zegt "Ik zie een* ***raat naast*** *een akte".* "Pien says 'I see a honeycomb near a file'"
final:	voiced	*Pien zegt "Ik zie een akker en een* ***raad****".* "Pien says 'I see a field and a commission'"
	voiceless	*Pien zegt "Ik zie een akte en een* ***raat****".* "Pien says 'I see a file and a honeycomb'"

Several means were introduced to distract participants' attention from the minimal pairs to avoid contrastive pronunciations. For instance, 50% filler trials were included with word pairs that differed in one segment either in onset, or nucleus, or coda position (like *kan* "pitcher" - *pan* "pan"; *rek* "rack" - *rok* "skirt"; *been* "leg" - *beer* "bear"). furthermore, the two members of a minimal pair (e.g., *raad* and *raat*) were combined with two different, but phonologically similar nouns in the sentence (e.g., *akker* and *akte*), see (7). A native speaker of Dutch controlled that the two nouns in the sentence were not semantically related and that all sentences were similarly odd. The phonological restrictions on the stimuli did not allow for the construction of semantically well-formed sentences. Participants were told that they performed in a syntax memorization task.[7] Since it is not easy to get subjects to produce encliticized forms in an experimental situation, the carrier sentences were rather long and had to be produced in a small amount of time (2000 ms).

4.1.2 The First Perception Experiment

Four participants of the production experiment provided the stimuli for the perception experiment. These were two men and two women, one of each came from the South and one from the North of the Netherlands. Their productions were slightly manipulated in the speech lab: The second noun, which had been different to distract the speakers from the minimal pairs, e.g., *akker* and *akte* was replaced by always the same noun *oom* "uncle" that had been taken from a filler sentence of the production experiment. Some of the minimal pairs of the production experiment had to be excluded, because in these pairs, one member has to be preceded by an article, while the other member is a mass noun and must not be preceded by an article. So the presence or absence of a determiner would inform the listeners which member of the minimal pair they heard. Seven minimal pairs were tested in the first perception experiment.

Participants were seated in a sound-proof booth in front of a monitor and a keyboard and heard a sentence over headphones, while the two members of the respective minimal pair were presented on the screen. One member of the minimal pair appeared on the right side, one on the left side of the screen. There was a scale between them with numbers from 1 to 5, for instance: *raat 1—2—3—4—5 raad.* Participants had to type a "1" when they were sure they heard the word appearing left on the screen (*raat* in the above example), a "2" when they thought they rather heard that word than the other one. They typed a "5" when they thought they heard the word presented on the right side of the screen (here *raad*), a "4" when they saw a trend towards this word. They typed a "3" when they could not make a decision between the words. The scores were automatically written to a result file. When they wanted to listen to the sentence again, they could do so maximally twice, using a push button device.

Looking at the results, the proportion of undecided responses was small in all contexts (about 10%). Speaker's region of birth and sex had no effect. Figure 2 shows the proportion correct responses. In the final and the nasal context, participants did not exceed chance level in deciding on the target word's final

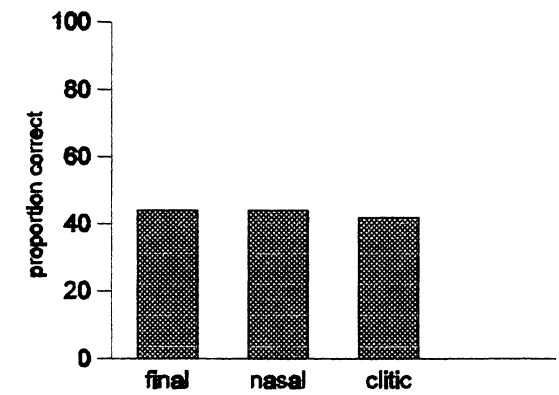

Figure 2. Proportions Correct Responses in the Three Context Conditions of Study 1

[7]There were six different carrier sentences with small syntactic variation. The sequence *Ik zie* "I see" in (7) was replaced by *Ik zag* "I saw", *Er is* "There is", *Er was* "There was", *Ik heb* "I have", and *Ik had* "I had".

obstruent, as expected. Crucially, however, also in the clitic context the proportions of correct responses turned out to be low. The responses given in this context did not differ from those in the other contexts, as predicted by an account that includes resyllabification.

However, encliticization is an optional process in Dutch. It might have been the case that although the tempo in the production experiment had been reasonably high, it was not high enough to guarantee that subjects produced encliticized forms all the time. If this was the case, also the model of phonological encoding would predict final devoicing, since in the absence of encliticization, the noun and the following phonologically strong function word are encoded seperately. The noun's final obstruent would then syllabify in coda position.

Two phonetically trained judges investigated the material auditorily and found that one speaker had not produced encliticized forms consistently. A reanalysis after excluding this speaker did not change the pattern of results. However, a revised set of experiments was run. The modified form also allowed for including a plural condition.

4.2 The Second Final Devoicing Study

4.2.1 The Second Production Experiment

The carrier sentences and the task differed from the first study. Instead of long sentences, the stimuli only consisted of the minimal pair noun, the conjunction *en* ("and") and the second noun, which was always a monosyllabic vowel-initial word. The structure of the new short carrier sequences allowed for the whole set of minimal pairs to be included. In addition to a clitic context (*raad en aar* "commission and are", *raat en aal* "honeycomb and ale") and a final context (*aar en raad* "are and commission", *aal een raat* "ale and commission"), a plural context replaced the nasal context condition. Plural forms could be produced for four minimal pairs of which both members have a regular plural form. Underlyingly voiced obstruents should remain voiced in the plural context (*raden en aren* "commissions and ares", *raten en aalen* "honeycombs and ales"). Importantly, the plural condition can serve as a proof that subjects are able to pick up differences in voicing from the signal.

The materials again contained 50% filler pairs. Participants performed a repeated articulation task. They were asked to memorize the target sequence that appeared in the center of the screen and to produce the target sequence as soon as they saw a cue signal on the screen. They had to have finished their response when they heard a beep over headphones. The visual cue reappeared again and again (11 times in total) and the time lag between cue onset and warning beep decreased stepwise by 70 ms until it was 430 ms short. This forced subjects to use speech of increasing speed.

4.2.2 The Second Perception Experiment

The task in the perception experiment remained unchanged. This time, subjects listened to 13 minimal pairs that occurred in clitic and final context, and four minimal pairs in plural context, spoken again by two men and two women, two from the South and two from the North of the Netherlands. Of the 11 repetitions that had been produced in the production experiment, the 5th was chosen for the perception task. Again, the utterances were manipulated by replacing the second noun always with the noun *olm* ("elm"). In contrast to the first study, the presentation of context was blocked: One group of subjects started with the clitic condition, followed by the final condition, the other group started with the final condition, followed by the clitic condition. Both groups ended with the plural condition.

As in the first perception experiment, the proportion of undecided responses turned out to be low. In the plural condition, subjects gave almost no such responses. As expected, in the plural context condition, both voiced and voiceless targets got above 90% accurate responses, and in the final condition, subjects again performed on chance level, see Figure 3. Most importantly, however, the rate of accurate responses was again about chance in the clitic context condition and did not differ from the final context. In both contexts, subjects assigned "voiceless"-responses to about half of both, the voiced and the voiceless targets. This indicates that although subjects are able to distinguish between the two members of a minimal pair in the plural condition, they cannot do this in the clitic and final condition.

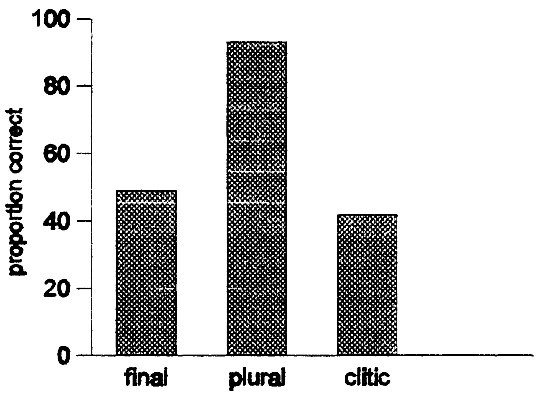

Figure 3. Proportions Correct Responses in the Three Context Conditions of Study 2

In addition, acoustic measurements were done on the materials of the second production experiment. The following cues to voicing were investigated: The duration of the vowel that preceded the stop, the length of the stop's closure and burst, and the absence or presence of voice activity during closure. The measurements confirmed the results of the perception experiment. In the plural forms, voiced stops clearly differed from voiceless stops. Voiced stops were accompanied by voicing during closure, whereas voiceless stops were not. Furthermore, voiceless stops had a clear burst, which was absent or minimal in voiced stops. In the final context condition, the durational values that were measured for underlyingly voiced and voiceless stops did not differ. This result indicates that final devoicing neutralized the voicing contrast not only phonologically, but also phonetically. Importantly, the measurements in the clitic context condition revealed the same pattern as in the final context condition. Underlyingly voiced and voiceless stops did not differ acoustically.[8]

5. Implications

Two perception experiments and acoustic measurements showed that the voice contrast of minimal pairs like *raad* - *raat* is neutralized when the stops occur at the end of an utterance, and hence in syllable coda position. In plural forms, on the other hand, where the stops are in onset position, subjects perceive voiced stops as voiced and voiceless stops as voiceless. Importantly, the voice contrast is neutralized in encliticized forms, although the stops preceding a schwa-initial clitic surface in onset position.

[8]The reader is referred to Baumann (1995) for a detailed description and discussion of the acoustic measurements.

This result is of interest for some issues in phonological theory. One ongoing debate is whether resyllabification or final extraprosodicity should be preferred in theories of syllabification (for a detailed discussion of this debate see Hall, 1994). Resyllabification is structure-changing: A consonant delinks from coda position, which in fact destroys the first syllable structure, before a new one is built. Therefore, some phonologists prefer extraprosodicity, where a root morpheme's final consonant is considered to be invisible for syllabification. When extraprosodicity turns off, the consonant participates in syllabification and accociates to the following onset if possible. In contrast to resyllabification, this procedure is structure-building only. According to Itô (1986), extraprosodicity turns off at word level, i.e., at the end of the derivational component. When a word-final consonant surfaces in the onset of a clitic's syllable postlexically, like in Dutch encliticized forms, she would have to assume postlexical resyllabification. Rice (1990), on the other hand, assumes that extraprosodicity holds by convention and remains active in phrasal phonology. The account is more appealing at first sight, since within- and between-word syllabification are treated by the same mechanism. But to account for final devoicing in Dutch encliticized forms, the obstruent has to occupy a coda position at some point, because underlyingly voiced stops surface voiceless in postlexical onset position. Dutch final devoicing in encliticized forms adds hence another case in favor of resyllabification to the debate of resyllabification versus extraprosodicity.

Furthermore, the experimental results are of interest for the question whether the phonological component includes intermediate levels of syllabification. A theory like Lexical Phonology, which distinguishes a lexical and a postlexical component, where the output of the former provides the input to the latter component, can account for the data (see also Booij, 1996): Final devoicing applies at the end of the lexical level, preceding the postlexical rule component. Theories that replace the traditional rules and derivations by other means can also explain the results. But they have to make additional assumptions to account for the data (as discussed in detail by Booij, to appear). For example, in Optimality Theory (McCarthy & Prince, 1993; Prince & Smolensky, i. pr.), the underlying (unsyllabified) representation of, say, a word is paired with a whole set of candidates for the word's surface structure, which is then evaluated by a set of ranked wellformedness constraints to determine the surface form. Postlexical phonological phenomena have not yet received much attention within the young Optimality framework. At the moment, the only option to account for the results seems to be to allow for two levels of constraint evaluation. First, surface candidates are evaluated at a lexical level, and the constraint that regulates syllable-final devoicing has to rank high within that level to make sure that candidate forms with voiced syllable-final obstruents are excluded from the set. The output of the lexical level is then further evaluated by constraints at a postlexical level that rule out candidates that are not encliticized. Importantly, these two levels of constraint evaluation have to be serially ordered. A simultaneous evaluation of the two levels does not provide the correct output.

To return to the production of syllables during speaking, the experimental results on final devoicing in Dutch enclititized forms present problems for Levelt's model of phonological encoding, according to which a speaker produces only one level of syllabic structure, and this is the postlexical syllable structure. Since in encliticized forms like *raad en* the final obstruent of the first lexical unit never surfaces in coda position, it should not be devoiced. however, the experimental results clearly showed that final devoicing applies on the stops that surface in onset position. One possibility to account for this could be to allow for resyllabification during phonological encoding. Resyllabification was included in earlier versions of Levelt's model (Levelt, 1989), which assumed that lexical items are syllabified seperately by associating the ordered sequence of their segments with the independently generated metrical frames. Final devoicing may then apply on the lexical words before encliticized forms combine into one prosodic word.

References

Baumann, M. (1995). *The production of syllables in connected speech.* (Unpublished Doctoral Dissertation). Nijmegen: University of Nijmegen.

Berendsen, E. (1986). *The phonology of cliticization.* Dordrecht: Foris.

Booij, G. (1977). *Dutch morphology. A study of word formation in Generative Grammar.* Lisse / Dordrecht: de Ridder / Foris.

Booij, G. (1995). *The phonology of Dutch.* Oxford: Clarendon Press.

Booij, G. (1996). Cliticization as prosodic integration: the case of Dutch. *The Linguistic Review.*

Booij, G., to appear. Lexical Phonology and the derivational residue. To appear in: J. Durand & B. Laks (Eds.), *Proceedings of the Conference on Models & Trends in Phonology, Royaumont, June 1995.*

Booij, G. & Rubach, J. (1987). Postcyclic versus postlexical rules in lexical phonology. *Linguistic Inquiry, 19*, 1-44.

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 of a connectionist model. *Journal of Memory and Language, 27*, 124-142.

Fromkin, V.A. (1973). Introduction and appendix. In: Fromkin, V.A (ed.), *Speech errors as linguistic evidence* (pp. 11-45, 243-269). The Hague: Mouton.

Garrett, M.F. (1975). The analysis of sentence production. In: G. Bower (ed.), *Psychology of learning and motivation: Vol. 9* (pp. 133-177). New York: AP.

Glaser, W.R., & Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 640-654.

Gussenhoven, C. (1980). Review of M. Trommelen & W. Zonneveld 1979 (Inleiding in de Generatieve Fonologie. Muiderberg: Dick Couthino). *Gramma, 4*, 174-183.

Gussenhoven, C. (1985). Over de fonologie van Nederlandse clitica. *Spektator. Tijdschrift voor Nederlandistiek, 15* (1985/1986), 180-200.

Hall, T.A. (1994). Extrasyllabicity and resyllabification. In R. Wiese (Ed.), *Recent Developments in Lexical Phonology (=Theorie des Lexikons, Arbeiten des SFB 282, Vol. 56)* (pp. 151-166). University of Düsseldorf and University of Wuppertal.

Itô, J. (1986). *Syllable theory in prosodic phonology.* (Doctoral dissertation). Amherst, MA: University of Massachussets.

Kooij, J.G. (1980). Morpheme boundaries and syllable boundaries: A case for natural phonology. In W. Zonneveld, F. van Coetsem, & O.W. Robinson (Eds.), *Dutch Studies. Vol. 4. Studies in Dutch phonology* (pp.61-78). The Hague: Martinus Nijhoff.

Lahiri, A. et al. (1990). The pronominal clitic [dər] in Dutch: A theoretical and experimental approach. *Yearbook of Morphology, 3*, 115-127.

Lenneberg, E. (1967). *Biological foundations of language.* New York: Wiley.

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. (1993). Timing in speech production: With special reference to word form encoding. *Annals of the New York Academy of Sciences, 682*, 283-295.

MacLeod, C.M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163-203.

McCarthy, J.J., & Prince, A.S. (1993). *Prosodic Morphology I: Constraint interaction and satisfaction.* (Ms.). University of Massachusetts, Amherst and Rutgers University.

Meyer, A.S. (1988). *Phonological encoding in language production. A priming study.* (Unpublished Doctoral Dissertaion). Nijmegen: University of Nijmegen.

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.

Meyer, A.S., & Schriefers, H. (1991). phonological facilitation in picture-word intereference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 1146-1160.

Prince, A.S., & Smolensky, P. (in press). *Optimality Theory: Constraint interaction in a generative grammar.* Cambridge, MA: MIT Press.

Rayner, K., & Posnansky, C. (1978). Stages of processing in word identification. *Journal of Experimental Psychology: General, 107*, 64-80.

Rice, K.D. (1990). Predicting rule domains in the phrasal phonology. In S. Inkelas & D. Zec (Eds.), *The phonology-syntax connection* (pp. 289-312). Chicago & London: University of Chicago Press.

Schriefers, H., Meyer, A.S., & Levelt, W.J.M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29*, 86-102.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial order mechanism in sentence production. In: W.E. Cooper & E.C.T. Walker (eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295-342). Hillsdale, NJ: Lawrence Erlbaum.

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