Probing early lexical and morphological processing in Dutch with the MMN response

Different responses to morphologically simple and complex words in Dutch

Labbé Grunberg, H.; Rispens, J.; Don, J.; Weerman, F.; Vet, D.J.

DOI
10.5117/NEDTAA2023.1.002.GRUN
Publication date
2023
Document Version
Final published version
Published in
Nederlandse Taalkunde

Citation for published version (APA):
Labbé Grunberg, H., Rispens, J., Don, J., Weerman, F., & Vet, D. J. (2023). Probing early lexical and morphological processing in Dutch with the MMN response: Different responses to morphologically simple and complex words in Dutch. *Nederlandse Taalkunde*, *28*(1), 36-71. https://doi.org/10.5117/NEDTAA2023.1.002.GRUN

Amsterdam
University
Press

NEDERLANDSE TAALKUNDE 28.1 (2023) 36-71
HTTPS://DOI.ORG/10.5117/NEDTAA2023.1.002.GRUN

Probing early lexical and morphological processing in Dutch with the MMN response

Different responses to morphologically simple and complex words in Dutch

Hernán Labbé Grunberg
Universiteit van Amsterdam
h.a.labbegrunberg@uva.nl

Judith Rispens
Universiteit van Amsterdam
j.e.rispens@uva.nl

Jan Don
Open Universiteit
jan.don@ou.nl

Fred Weerman
Universiteit van Amsterdam
f.p.weerman@uva.nl

Dirk Jan Vet
Universiteit van Amsterdam
d.j.vet@uva.nl

Abstract

The past tense inflection has been a popular phenomenon to study the representational status of morphologically complex words. While several experiments in the processing of past tense verbs across languages have shown these verbs are stored via their constituent morphemes, experiments in the processing of the Dutch past tense indicate that these words are lexically stored in their surface form and therefore not decomposed. However, the experiments in Dutch past tense processing have not made use of experimental paradigms that can tap the earliest stages of word processing, where some theories predict decomposition would take place. We used the mismatch negativity response to study the representational status of monomorphemic and morphologically complex Dutch words. We were able to obtain different responses for monomorphemic and morphologically complex words, suggesting these are processed by different mechanisms. We cannot, however, discard the possibility that some past tense forms in Dutch do have surface form representations.

Keywords: MMN, lexical processing, morphological processing, morphological decomposition

1. Introduction

Different linguistic analyses have been proposed regarding the storage and processing of (morphologically complex) words. The question behind those hypotheses is the same and concerns the role of our linguistic memory: are mental representations sensitive to the internal structure of words (morphology) or is it enough to store sound-meaning combinations that we hear often without morphological specifications? Assuming the latter, the more we hear certain combinations of sounds associated with a specific meaning, the more likely a mental representation will be created that links those sounds to that meaning (Bod et al. 2003). Models that adopt this framework assume all words are stored via their surface forms (Rumelhart & McClelland 1988; Bybee 1995; Baayen et al. 2011). The hypothesis that (repeated) exposure to certain stimuli is the basis for creating neural representations of said stimuli in memory is referred to as *entrenchment* in the theory of Cognitive Grammar (Langacker 2017). Assuming the former account, where the mechanisms of language processing are sensitive to morphological structure, our lexical memory would consist of individual morphemes only (Taft & Forster 1975; Marslen-Wilson et al. 1996). In terms of processing, incoming acoustic information that is morphologically complex would then be decomposed into its constituent morphemes and only then would these elements activate their corresponding representation in memory. For example, the Dutch monomorphemic noun *seconde* ‘second’ and the past tense verb *speelde* ’played’ have the exact same ending (-*de*). However, their internal structure may be processed differently at the neurocognitive level. A monomorphemic word like the Dutch noun *seconde* is assumed by most (psycho)linguistic theories to be stored as a single sound-meaning representation in lexical memory. If, however, the mechanism of word processing is sensitive to the compositionality of *speelde*, or other regular past tense verbs, then decompositional theories would predict a decompositional mechanism would be activated when processing this word. This mechanism would identify the compositionality of the incoming stimulus, and would separate the word into its constituent elements, retrieving them separately from lexical memory, and creating the surface form of the word online, rather than retrieving it from memory in its surface form. Such an account presupposes that the decompositional mechanism is sensitive to the morphological patterns of the language, and therefore, that some form of grammar must be part of the cognitive representation of language (Marantz 2013).

The two opposing hypotheses are combined in so-called dual-mechanism models. In these models, the regular past tense has been hypothesized to be stored through constituent elements and processed via decomposition. However, not all morphologically complex words are assumed to be decomposed by theories that postulate the existence of such a mechanism. The *Words & Rules Model* by Pinker (1999) and the *Parallel Race Model* (PRM) by Baayen et al. (1997) are both examples of models that claim the decompositional route demands a higher cognitive load, and is therefore used only if direct, surface form retrieval is not available for an acoustic or visual stimulus. A variable believed to facilitate complex words being stored in their surface form is frequency of occurrence. These models claim the more frequent a word is processed, the more likely it will be stored in its surface form, even if a decompositional route exists for its processing.

Another factor that has been linked to surface form storage is derivational morphology, which is seen as fundamentally different from inflectional morphology (Schriefers et al. 1992; Badecker & Caramazza 1989; Maury et al. 2015). The latter, which includes plurality and past tense marking, adds extra information to a stem without changing its word class or meaning. For example, the verb *spelen* ‘to play’ remains a verb in its inflected form *speelde* ‘played’. Within decompositional theories, inflected verbs are often hypothesized to be stored and processed through their constituent elements precisely because inflection does not change the word class or overall meaning of the base word. However, if *speel* receives the derivational suffix -*er*, the verb changes word category and changes its meaning to ‘the agent performing the action denoted in the verb’.. This is one of the reasons why words that undergo derivations in morphology are thought to be processed differently from words that undergo inflections.

In this study, we will use EEG recordings to compare monomorphemic words to two instances of morphological complexity: two regular past tense verbs (inflectional morphology) and a derived, nominalized adjective (derivational morphology). By testing these two types of complex words, we hope to shed light on the difference (if any) between the two types of morphological complexity, and, specifically, on whether these two types are processed differently from monomorphemic words, which we know must be stored in their surface forms. Additionally, a comparison of our results to those obtained by Baayen and colleagues using lexical decision tasks (Baayen et al. 1997; Schreuder & Baayen 1997; Bertram et al. 2000) will also shed light on the evidence that can be gathered with tasks that tap into different stages of word processing (early for MMN, after procesing has ended for behavioral responses). The use of the nominalized adjective will moreover allow us to test the hypothesis whether derivational morphology requires the use of surface form representation of the derived forms. If this is the case, we expect the derived noun to show similar responses to monomorphemic words.

The next sections will discuss the methods used and evidence gathered by studies on the processing and storage of morphologically complex words, and the implications of these findings on theories of morphological processing and storage (section 2). We will then focus on the evidence and methods used to study processing and storage of complex words (section 3.1) and of Dutch complex words in particular (section 3.2).,We will show that a neurophysiological approach to the study of Dutch complex word processing and storage is needed to answer pending questions in the field (section 3.3). We then introduce the present study (section 4), its methods (section 5) and analysis techniques (section 6). Finally, we present our results (section 7) and discuss their implications (section 8).

2. The study of storage and decomposition

Although models that assume decomposition of complex words are often grouped together, there are actually several accounts in the decomposition literature regarding which words go through one or the other mechanism. Full decomposition models hypothesize full decomposition applies to all complex words, regardless of regularity, frequency or other factors (e.g. Taft 2004). Dual mechanism models (Clahsen 1999; Pinker 1999; Pinker & Ullman 2002; Marslen-Wilson & Tyler 2008) claim some complex words are decomposed, while others are stored in their surface forms.

Different types of dual-mechanism models differ in the configuration of both mechanisms and the type of linguistic constructions that are decomposed. In general, most dual-mechanism models assign an important value to structural transparency (Baayen et al. 1997; Pinker 1999). This refers to the ease with which the transparency of a word's compositional structure can be seen by looking at its surface form. For example, a verb stem like the Dutch *werk* ‘work’ drops the infinitive *-en* ending in *werken* ‘to work’ and receives the suffix *-te* to produce the past tense form *werkte* ‘worked’. Such a past tense form is said to be structurally transparent because the presence of both morphemes (the verb stem and the suffix) is clear from the linear order of the sounds that make up the word. The opposite is the case with a verb stem like *loop* ‘walk’, where the past tense form also follows a rule, but the rule does not add an extra element *next to* the first constituent. Instead, the rule changes the vowels inside the stem, making the stem *loop* turn into the past tense *liep* ‘walked’. In such cases of strong inflection, the inflected form is said to be structurally opaque, and therefore, it is assumed that the decompositional mechanism cannot recognize the compositionality of the word. Consequently, the input is processed as a single unit and relies on surface form storage instead of decomposition.

Another factor assumed by dual-mechanism models to influence which words are decomposed is the frequency with which words are encountered. The *Parallel Race Model*, for example, based a large part of its evidence on frequency effects of simple and complex words (Baayen et al. 1997; Bertram et al. 2000; de Jong et al. 2000; Schreuder 2003). The frequency effect rests on the assumption that if a certain word is encountered (and therefore processed) very frequently, then it is more efficient for the language processing mechanisms to store the whole word in a surface form representation, even if the word is structurally transparent and morphologically complex. Consequently, highly frequent morphologically complex words, as well as formulaic expressions (such as *hit the road* or *kick the bucket*) are assumed to be stored as single sound-meaning representations.[1]

Within the debate about the mechanism(s) responsible for processing morphologically complex words,he addition of the suffix to verb stems is a paradigmatic example of the systematic and productive regularity which can be exploited by a decompositional mechanism that relies on structural transparency. Despite its transparency, derivational morphology, on the other hand, has been argued to be processed differently from inflectional morphology (see Maury et al. 2015 for behavioral evidence and Bozic & Marslen-Wilson 2010 for neurophysiological evidence). After decades of

---

[1] There is no consensus as to what is the frequency threshold that determines when a word is decomposed and when it is stored in its surface form (although certain authors, like Alegre & Gordon (1999) have calculated the threshold for surface form storage to be 6 per million). This is due to the fact that frequency is calculated as the number of occurrences of a word in a specific corpus.

research there is general (but not unanimous) agreement that both surface form storage and decomposition are part of language processing. The debate revolves around which (types of) words are decomposed. In this study, we aim at providing an experimental paradigm that can reliably assess the representation of words at the earliest stages of word processing.

To know that our responses are tapping the representation of stored words, and are not simply a brain reaction to sound, we aim to use brain responses to distinguish strings of sounds that are words, from strings of sounds that are not (i.e. that are pseudowords or ungrammatical strings). Next, we hope to elicit different responses by presenting participants with existing words that differ in their morphological complexity. While we expect past tense verbs to show evidence of decomposition, our derived noun should, according to the literature based on dual-mechanism models, to behave like a monomorphemic word, in that there should be a single surface form representation for it. Alternatively, theoretical accounts of morphological knowledge and representation (Embick & Noyer 2007) would predict all morphologically complex words, including derivations, to behave similarly to inflectional morphology. On this account, we could expect a derived noun to be decomposed as well.

3. Morphological Processing: Empirical Results And The Case Of Dutch

3.1 Lexical decision task and masked priming

The evidence about Dutch morphological processing that we will review here is based on the Lexical decision task (LDT), where participants are shown strings of letters or sounds, and are asked to classify them into categories (like words or pseudowords, nouns or verbs, etc.). Response times and accuracy are taken as proxies for the accessibility of the word (how easily their neural representations are activated). A commonly used variation of the LDT is known as *masked priming LDT*. Just as the LDT version just described, a masked priming LDT also requires a response to be made regarding the classification of a target into predefined categories. Unlike the normal LDT designs, however, a word is used to prime the recognition of the target word. Crucially, the participant is barely able to process the prime, usually with no conscious knowledge of having been shown a word at all. The presentation of a word before the target form is expected to exert influence on the accessibility of the target itself. By manipulating the relation between the prime and the target, masked priming designs allowed researchers to investigate the relation between morphologically related words (Forster 1999).

The competing hypotheses of surface form storage and decomposition have been assessed with both types of LDT. The key difference lays in the stage of word processing that both tasks are able to probe. While LDT elicit responses that reflect the end-result of lexical processing, masked priming LDT assess the influence of unconscious lexical processing on the response to the target word. Most evidence of Dutch morphological processing comes from unmasked LDT, where the latter stages of lexical processing are investigated (Baayen et al. 1997; Bertram et al. 2000; Schreuder 2003).

The LDT, in its different versions, are taken as an important source of evidence for surface form storage and decomposition in the psycholinguistic literature. The results that have accumulated over the years in the literature, across languages, suggest that regular, transparent cases of complexity, such as the regular past tense are likely decomposed when being processed (Lignos & Gorman 2012). Derivational morphology, on account of the change in word category and/or meaning of the derived stem, would be a case of morphological complexity that requires surface form storage (see Rastle & Davis 2008 for a review of empirical results using masked priming designs).

3.2 The case of Dutch

An early study that investigated surface form storage and decomposition in Dutch morphological processing (Bertram et al. 2000) used the visual LDT to test the processing of the suffix *-te* used as a regular past tense (*werk* – *werk**te*** 'work – work**ed**') and as a nominalizing suffix (*warm* – *warm**te*** 'warm – warm**th**'). These experiments relies on measures of base and surface frequency, and their effect on reaction times in lexical decision tasks. Base frequency is a measure of the productivity of a given word stem across all instances of inflection or derivation. When the base frequency of a word predicts its reaction time, this is taken as evidence the word is being processed by the activation of its stem rather than its surface-form. Surface frequency effects, on the other hand, are taken as evidence of surface-form processing. In experiment 1a Bertram and colleagues compared high and low surface frequency verbs which were matched for base frequency. They found no effect of surface frequency in regular past tense verbs. In experiment 1b, they compared high and low base-frequency past tense verbs while controlling for surface frequency. Their results showed a significant effect of base-frequency. Taken together, these results were interpreted as evidence that the past tense is processed through decomposition into its constituent elements.

However, later studies challenged the hypothesis that the frequency effects of the results by Bertram et al. (2000) suggest rule-based decomposition of the past tense verbs. Ernestus & Baayen (2001, 2004) used production experiments to test to what extent the production of both past tense allomorphs -*te* and -*de* could be explained by a single alternation rule, whereby stems ending in voiceless consonants take the -*te* allomorph, while all other stems take on the -*de* allomorph. Although this research dealt with morphophonology rather than morphological processing per se, and with production rather than comprehension, its premise did rest on the assumption that the surface forms of inflected past tense verbs influence the choice of allomorphs for other verbs. Such an assumption implies that past tense words must be stored in their surface forms, which is incompatible with the hypothesis that morphologically complex words are stored via their constituent elements and, therefore, not in their morphologically complex surface form.

Some studies point to the possibility that complex words are generated by analogical processes with similar words with which they share semantic and phonological similarities. Ernestus & Baayen (2004) found that production of complex words could be explained by analogical pressure from surface form representations of similar verbs rather than rule-based processes. Tabak et al. (2005) showed that regular past-tense verbs show surface-frequency effects that are correlated with the degree of rhyme regular verbs have with irregular verbs. These studies suggest that analogy with phonologically similar stems, and not the instantiation of abstract rules, is the mechanism through which verbs are inflected in Dutch.

Regarding the processing of derived words with the -*te* suffix, experiment 2 of Bertram et al. (2000) replicated their studies 1a and 1b, but with the same ending acting as a nominalizing suffix. Their results showed an effect of surface-frequency only, and no effect of base-frequency in the recognition latencies of derived nouns with the -*te* suffix. These results were interpreted as evidence that derived nouns with the -*te* suffix were stored in their surface forms, and, therefore, not decomposed. This evidence suggests that in Dutch the two mechanisms of surface form storage and decomposition are used, with surface form storage acting as the preferred mechanism, and decomposition being used when surface forms do not have the necessary representational strength to support direct storage.

3.3 Limitations of previous results and the need for neurophysiological evidence

Despite the fact that analogy between surface form representations is indeed a plausible scenario when thinking about access-to-word representations, Taft (2004) argues that such effects do not preclude the existence of decomposition-based processing of morphologically complex words. Even if words are decomposed into their constituent elements, there must still be a later stage when these constituents are recombined into the surface forms we are all familiar with (Taft 2004). Following this reasoning, experimental results based on analogical relations between surface form representations, like the ones described above in Dutch, do not preclude the possibility that there was a decomposition stage prior to the activation of a surface form representation.

If surface form representations are ubiquitous even in models of decomposition, then the time at which processing is probed becomes crucial. Event-related potential (ERP) responses are sensitive to the firing of large groups of neurons in the cerebral cortex, and can be easily measured by placing caps with electrodes on participants scalps. EEG recordings can be time-locked to specific moments during word processing experiments, and they are able to measure electrical activity from the brain with microsecond accuracy.

Studies of morphological processing based on neurophysiological responses have tended to produce more evidence favoring decomposition of complex words in regularly and irregularly inflected words and derivations (cf. Marslen-Wilson & Tyler 2007; Morris & Stockall 2012). This is likely due to the fact that ERP responses can be time-locked to the exact moment where a linguistic item has been processed, whereas behavioral responses like those obtained in the Dutch lexical decision task experiments reported above require the process to end, and the response to be planned and executed, before the experimenters can obtain information about the processing of the word.

Indeed, neurophysiological evidence using ERP responses shows that already before 400 milliseconds, where the first semantic effects show up in ERP responses, there are at least three components sensitive to morpho-syntactic structure: The mismatch negativity (MMN), the early left anterior negativity (ELAN) and the left anterior negativity (LAN). While all three ERP responses have been elicited in the context of syntactic processing of sentences, the MMN response has also been reported to be modulated by the complexity at the word-internal (morphological) level. If all these ERP components peak before 400 milliseconds, and the first semantic ERP peaks at 400 milliseconds, then it becomes obvious that behavioral responses, which require approximately 500 milliseconds before responses can be given by participants, are ill-suited to tap into the earliest stages of word processing.

ERP responses are measured as voltage change over time. This means that changes in the amplitude of an ERP response constitute the dependent variable by which theories about language processing can be compared. Amplitude changes in the MMN response are calculated by subtracting the response to a repeated stimulus (the standard stimulus) from the response to a less frequently presented stimimulus (the deviant stimulus). This difference, which results in a negative amplitude modulation (i.e the amplitude shows a decreased in microvolts compared to the baseline measurement), is known as the MMN amplitude. The amplitude of the MMN response has been linked to the strength/degree of consolidation of the memory trace being activated by the sounds presented (Pulvermüller & Shtyrov 2006). This means that, compared to an acoustic control, a string of sounds that corresponds to an existing word in the participant's language will yield a more negative MMN response. This difference in MMN amplitude between existing (words) and non-existing (pseudowords) strings of sounds is known as the lexical MMN, because it can distinguish the lexicality of the strings of sounds being presented.

Of relevance for our study, Shtyrov & Pulvermüller (2002) showed that not only monomorphemic words, but also suffixes can have memory traces. They presented participants with the deviant *come* and *comes* and found that both elicited MMN responses differed in latency and topographical distribution. The authors concluded these differences were due to the fact that two different representations had been tapped with the MMN responses: that of the verb stem *come* and that of the third person singular suffix *-s*. In a follow-up experiment, Pulvermüller & Shtyrov (2003) presented the same verb stem and inflection (*walk* and *walks*) but this time preceded by either the pronoun *we* or an acoustically matched non-linguistic sound. They found that the correct string formed by *we come* produced a less negative MMN response than the ungrammatical *we comes* and the verb *come* presented in isolation. The authors interpreted this finding by claiming that the brain uses specialized sequence detectors that are sensible to the co-occurrence of linguistic items. In other words, after hearing *we*, the brain was able to anticipate the presentation of *come* such that the MMN amplitude was reduced. This reduction of grammatically correct syntactic constructions was labeled *syntactic mismatch negativity* (sMMN).

Earlier studies on inflectional and derivational morphology as well as compounding have shown that the MMN may be a valuable measure to study the processing and storage of morphological complexity. Bakker et al. (2013) found the memory trace of regularly inflected past tense verbs was smaller than that of an ungrammatical word (an sMMN), suggesting that, in the early stages probed by the MMN, there is no memory trace for the surface form of the regularly inflected past tense verb. Leminen et al. (2013) showed that derivations in Finnish behaved like monomorphemic words, thereby suggesting these are stored in their surface forms. Likewise, Hanna & Pulvermüller (2014) showed that MMN responses for German-derived words were larger than those elicited by pseudoword controls, indicating that these derived words had pre-existing memory traces and that they were, therefore, stored as surface form representations. Regarding compounds, MacGregor & Shtyrov (2013) found that opaque compounds (e.g. *groundwork*) showed evidence of being stored in their surface form, while transparent compounds were not different from their control pseudowords, and were not sensitive to surface-frequency effects, suggesting constituent-based representation and processing of such compounds. Particle verbs (also a form of compounds) were shown to be processed as single words regardless of their semantic transparency (Cappelle et al. 2010) and even in syntactic contexts where they are separated by other words (Hanna et al. 2017).

4. The present study

Our experiment aims at probing mechanisms responsible for monomorphemic, inflected and derived word processing in Dutch. Specifically, we want to examine the earliest stages of processing by time-locking the MMN component to the presentation of the word's final sounds (-*de* and -*te*). Assuming that the final sounds of the word trigger the retrieval of the memory trace that corresponds to that word, we will compare the amplitude of the MMN response elicited by the same deviant sounds (-*te* and -*de*) instantiating different types of words: Monomorphemic nouns, past tense verbs, a derived noun, and, acting as acoustic controls, pseudowords and ungrammatical verbs. We will also report the latency of the MMN peak, since it is assumed by most models of word processing, such as the Parallel Race Model (Baayen et al. 1997), that processing a word through decomposition takes more time than retrieving its surface form from memory. In any case, our design prevents us from directly comparing singular nouns (which we assume are directly retrieved) and past tense verbs (which we assume are processed via decomposition of their constituent elements), because the MMN response is too sensitive to the acoustic features of each standard-deviant pair.

Our first research question is whether the MMN paradigm can be used to differentiate words from pseudowords in Dutch. We expect distinct patterns of response when comparing our Dutch words to their acoustic controls. Specifically, we expect monomorphemic words will elicit more negative MMN amplitudes than their pseudoword and ungrammatical controls (lexical MMN).

Our second research question concerns the processing of regularly inflected past tense verbs in Dutch. While behavioural experiments carried out by Ernestus & Baayen (2001, 2004) showed evidence for lexical storage of surface forms for regularly inflected past tense verbs, Bakker et al. (2013) showed evidence for the decomposition of regularly inflected past tense verbs in English using the MMN paradigm. We will therefore compare Dutch grammatical past tense verbs to an acoustically matched pseudoword in the case of the *-de* past tense allomorph, and to an ungrammatical past tense verb (inflected with the wrong allomorph) in the case of the *-te* suffix. The reasoning of the syntactic MMN is that a complex word that is decomposed should not have a memory trace for the surface form of the word, but only memory traces for the constituent elements. This is why the memory trace for the surface form of a complex word is expected to be smaller than the memory trace of a string of sounds that is not part of lexical memory. A pseudoword and an ungrammatical word are, therefore, equivalent for the purposes of this MMN paradigm. We therefore use both types of contrasts for our past tense condition, by comparing *stolde* 'solidified' to the pseudoword *spolde*, while comparing the past tense *sleepte* 'dragged' to the ungrammatical *speelte* 'played'.

The MMN paradigm allows for strict control over variables that influence ERP responses, such as stimulus variability (item), attentional modulations and task-related ERP activity (Pulvermüller & Shtyrov 2006). However, the strict acoustic control of stimuli greatly reduces the types of words that can be tested in any given word class. In this line, recent MMN experiments have allowed themselves more variation in the acoustics of standard stimuli, as long as they keep the deviant element the same (e.g. *-ie* and *-ions* deviants preceded by different standards: '*il cr-*', '*il tr-*', '*il br-*' and '*nous cr-*', '*nous tr-*' and '*nous br-*' in Hanna et al. 2014, and *sicherheit* ('security') versus *sauberkeit* ('cleanliness') in Hanna & Pulvermüller 2014). In order to allow ourselves comparisons of monomorphemic, regular past tense and a nominalized adjective, we sacrificed some of the strict acoustic controls among our stimuli, while making sure the deviant elements were always the same among compared items. Table 1 shows each Dutch word used, together with the pseudowords and ungrammatical verb used as controls. Tests were carried out after the experiment to know what the participants report when hearing the stimuli. We played each standard-deviant combination to participants and asked them to write down what they heard. In case they heard a known word, they were asked to write an example sentence where the word is used. For example, a participant that heard the standard-deviant combination *secon* and *de* and recognized the Dutch word *seconde* 'second' wrote the example sentence *Ik wacht een seconde* 'I wait one second'.

Finally, based on the different results for inflected and derived words in Finnish (Leminen et al. 2013) we compared the response to a derived noun

Table 1. Stimuli and conditions

Condition	Standard	deviant	Class	Expected outcome
Lexical MMN	secon	de	singular noun	seconde > spolde
	spol	de	pseudoword	
Lexical MMN	sek	te	singular noun	sekte > spalte
	spal	te	pseudoword	
Syntactic MMN	stol	de	past tense	stolde < spolde
	spol	de	pseudoword	
Syntactic MMN	sleep	te	past tense	sleepte < speelte
	speel	te	ungrammatical past tense	
Lexical MMN	zwaar	te	derived noun	zwaarte > spalte
	spal	te	pseudoword	

with the *-te* nominalizing suffix (*zwaarte* 'heaviness') to an acoustically matched pseudoword (*spalte*). If derivations are stored in their surface forms, we should see a larger response to the nominalized word than to its control pseudoword. If, however, the response to the derived word is smaller than that of the pseudoword, we should conclude the derived word is being decomposed. We expect a lexical MMN for the derived noun because both the behavioral results from Dutch as well as the MMN studies on derived words have shown these to be stored as surface form representations. However, there are even behavioral and neurophysioloigcal studies that have found evidence that native speakers decompose even derived words (Stockall & Marantz 2006; Diependaele et al. 2011; Clahsen & Neubauer 2010; Marantz 2013; Fruchter & Marantz 2015). We therefore do not rule out the possibility that our derived word could elicit a MMN response with a less negative amplitude than its pseudoword control, thereby resulting in a syntactic MMN response.

5. Methods

5.1 Participants

Twenty right-handed native speakers of Dutch (ages ranged from 18-45; mean age was 25.7, 8 males) with normal hearing and without any known attentional or language disorders were recruited via advertisements on the internet, university boards and posters in Amsterdam. Participants gave their written consent before the experiment and were paid a total amount of 30 euros for their participation (15 euros per session). The study was approved by the Ethical Committee of the Faculty of Humanities of the University of Amsterdam.

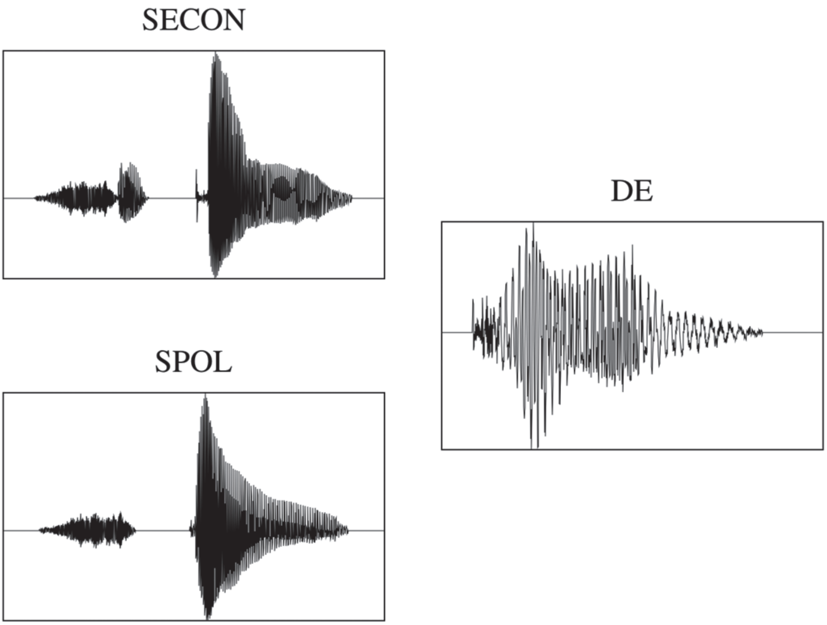

Figure 1. Construction of the singular noun *seconde* and its pseudoword control *spolde*. Both stimuli carry the exact same acoustic ending (-*de*)

5.2 Materials

We elicited MMN responses using the oddball paradigm, where a repeating standard sound is occasionally interrupted by the presentation of a deviant sound (-*te* or -*de* in this case). The standard stimuli corresponded to the words shown in Table 1, but without the deviant sounds. We randomized the presentation of deviant stimuli such that they would have a probability of occurrence of 17% and would total 120 per block. We instantiated two constraints on the presentation of our stimuli: no fewer than two and no more than eight standards could appear between two deviants, and each block started with 10 standards in a row, which were later discarded for the analysis. We also discarded the standard that came after each deviant from analysis.

Stimuli were recorded by a female native speaker of Dutch with a sampling rate of 44.1 Khz. Words containing the necessary sounds to create the stimuli were articulated several times and under different acoustic contexts to minimize co-articulatory bias. The recordings were selected in a way that they were most similar to each other in terms of length, pitch and intensity. Additionally, the exact same sound strings were used across words when possible (the -*de* and -*te* endings are the same across conditions, as is the

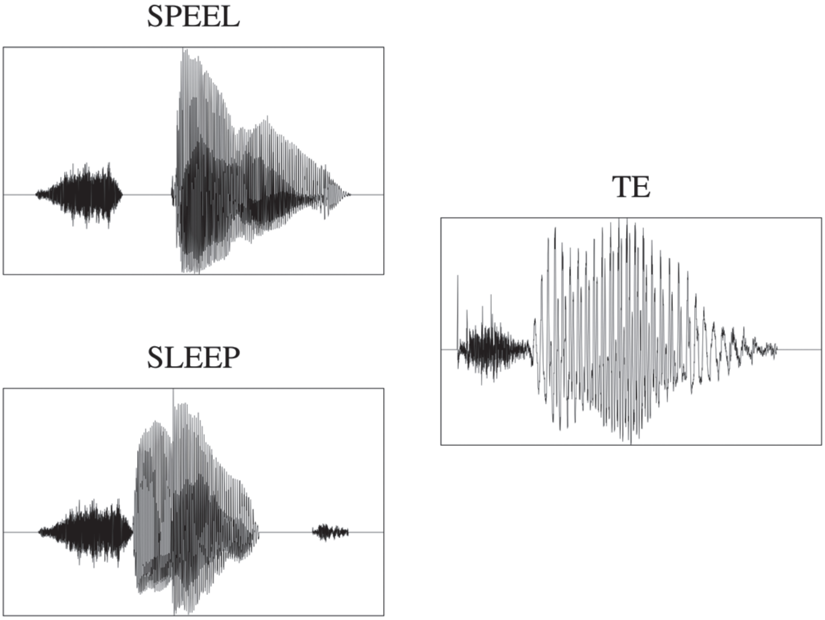

Figure 2. Construction of the past tense verb *sleepte* and its ungrammatical control **speelte*. Both stimuli carry the exact same acoustic ending (-*te*)

initial -*s* consonant and the -*ool* sequence in *spool* and *stool*). Figure 1 shows how the singular noun *seconde* and its pseudoword control *spolde* were built. Figure 2 shows how the past tense verb *sleepte* and its ungrammatical control *speelte* were built.

5.3 Procedure

Participants were seated in an acoustically and electrically shielded room, looking at a monitor in a dimly lit room, at the ACLC lab for EEG studies of the University of Amsterdam. Sounds were presented through Sennheiser HD419 headphones. Volume was set the same for all participants and they were instructed to watch a silent movie while paying attention to the subtitles of the movie and to ignore the sounds coming through the headphones. The stimuli were presented through software written by the university's lab engineer. The software received the acoustic stimuli and sent it through one channel and then duplicated to be heard on both sides of the headphone, while using another channel to send an acoustic cue to the Biosemi amplifier that synced the cue to either the start of the standard stimuli or the start of the deviant (-*de*, -*te*) with a high temporal precision.

Participants were given the option to choose among three subtitled Dutch movies namely *Lek*, *Zwartboek* and *Alles is liefde.* The experiment lasted approximately 80 minutes.

Each session consisted of either 4 or 3 blocks (one block for each condition), with each block lasting approximately 20 minutes. Participants had to take part in two sessions that took place in different days, totaling 7 blocks per participant in total. There was an 83% of standard stimuli per block and deviants occurred at a total of 17%. To habituate participants to the presentation of the sounds through the odd-ball paradigm, each block started with 10 standard stimuli played consecutively; these standards were not included in the analysis.

5.4 EEG recording and processing

Electroencephalogram data was recorded with a 64-channel Biosemi Active Two system (Biosemi Instrumentation BV, Amsterdam, The Netherlands). Scalp electrodes were arranged following the 10-20 convention. In addition to the 64 electrodes of the cap, three additional electrodes were placed on the mastoids and the nose. Offline processing of the data was done in Praat (Boersma & Weenink 2001): the reference electrodes (right and left mastoids) were subtracted from the EEG data, linear detrending and filtering with a high-pass filter of 1 Hz and a low-pass filter of 25 Hz. Automatic artifact rejection removed all epochs that had a value exceeding +75 or -75 µV. 48 electrodes (from FPz to Cz, including all corresponding lateral electrodes, plus CPz) were selected on which pre-processing was applied. A peak-to-peak eye-blink correction was applied to remove epochs with eye-blinks. This was done by calculating the voltage difference between the most positive and most negative peak in any epoch, which was compared to a threshold of 3.5 Z-scores, based on the mean voltage of all epochs. Epochs containing artifacts were removed, as were the first 10 standard epochs at the start of each block and any standard epoch directly following a deviant stimulus. Data of participants with fewer than 75 deviants left after artifact rejection were not included in further analysis. The sampling rate was at 8 kHz which was downsampled to 512 Hz after recording (Biosemi Decimator 86). Epochs started 110 ms before the onset of the deviant stimulus, and lasted 700 ms after onset of the deviant stimulus. The onset of the deviant sound was detected by including a sinusoid on the second track of the audio signal that contained the stimuli. The sinusoid was automatically detected by an external hardware, which then inserted a marker into the recording stream of the EEG recording device. This method provides an accuracy better than 30 µs.

In order to select the MMN component from each epoch, a Praat script was used that searched for the most negative value of the ERP component in a time window between 80 milliseconds after the deviant onset and until 210 milliseconds. Visual inspection of the epochs revealed that there was a second, later negative peak. However, because we had no predictions regarding such a second, late MMN peak, we did not include it in the analysis. The second peak is, however, visible in the grand average plots (Figure 4).

The procedure to compute the average voltage of the MMN component was as follows: First the most negative voltage value within the specified time window was found. Once the peak was found, the area of 20 milliseconds around the peak (10 milliseconds to the left and 10 to the right) was calculated. To measure latency, we used the time at which the most negative value of the MMN component (its negative peak) was present. As mentioned earlier, we are interested in assessing the strength of the neural memory trace of monomorphemic, past tense and a derived Dutch word. We will do so by comparing the difference in the MMN amplitude of these words to the MMN amplitude of control pseudowords and ungrammatical forms. We therefore assume that the average amplitude of the MMN peak reflects the strength of each memory trace. The more negative the amplitude is, the more consolidated the memory trace is assumed to be. We interpret the latency of the MMN peak as a proxy for how fast the memory trace of a word was activated.

6. Analysis

We built linear mixed-effects models in R (R Core Team 2013) using the lmerTest package (Kuznetsova et al. 2017). We used grammaticality (grammatical versus ungrammatical) or lexicality (word versus pseudoword) as a fixed effect, depending on whether we compared real words to ungrammatical or to pseudowords. We used a maximal structure for random effects (Barr et al. 2013). This means that each fixed effect we included in a model was also included as a random slope with the variable of subject as a random intercept. We established orthogonal contrasts for each fixed effect (Baguley 2012) which resulted in one contrast for the lexicality/grammaticality variable, two contrasts for the hemisphere variable and four contrasts for the region variable (see Table 2). We then proceeded to compare nested models through analysis of variance tests, such that the simpler model was always included in the more complex model to which it was compared. If the more complex model provided a significantly better fit, as indicated by a significant p-value, then we kept that model and compared it to a more complex, nested model until we found the best model.

Analysis of the MMN is usually circumscribed to midline and frontal electrodes. However, visual inspection of our data revealed that the MMN was also present on the electrodes at both the left and right sides of the midline, and extended well into centroparietal areas. We therefore divided the electrodes into 3 hemispheres (left, midline and right) and 5 regions spanning frontopolar electrodes until the parietal electrode of the midline. We did this so as to be able to take the variation of the MMN across hemispheres and regions into account when comparing the MMN responses.

Although the midline is not considered a different hemisphere, we decided to separate it from the electrodes in the left and right hemisphere, so as to compare how the MMN response propagates at the left and the right sides of the midline. We specified two hemispheres contrasts in the model. First, we compared the average MMN response in the left and right hemispheres to the average MMN response at the midline electrodes. Second, we compared the MMN responses of the right and left hemisphere while ignoring those of the midline.

Our regions included, from anterior to posterior: 1) frontopolar and anterofrontal, 2) frontal, 3) frontocentral, 4) central and 5) centroparietal. We decided to include the midline electrode of the parietal region, but not the electrodes to its sides, since at this region the MMN did not propagate to the sides as clearly as in the more anterior regions. We specified 4 contrasts for the 5 levels of region. We first compared the peripheral regions (frontopolar-aterofrontal and centroparietal) to the central regions (frontal, fronto-central and central) in order to confirm that our MMN responses were stronger at central regions, as previous literature has shown. Our second region contrast compared peripheral regions to each other. That is, the frontopolar and anterofrontal electrodes against the centroparietals. Our third contrast compared central and frontocentral MMN responses against those of the frontal regions, while the fourth contrast compared central versus frontocentral MMN responses. Table 2 shows the contrasts that were used for each fixed effect.

Visual inspection of the data revealed that the MMN did not always peak at the central midline electrodes, as is usually the case in the MMN literature. Indeed, experiments that elicit the MMN with meaningful linguistic stimuli have produced MMN responses that propagate across hemispheres and regions. This has caused researchers to take different decisions on what electrodes to include in their analyses. In an early study, Pulvermüller & Shtyrov (2003) used either electrodes Fz and FCz plus two electrodes to each side, or Cz and Pz plus

Table 2. Contrasts for fixed effects

Contrasts	Lexicality/Grammaticality	Hemisphere	Region
Contrast 1	Word/Grammatical vs Pseudoword/Ungrammatical	Left & Right vs Middle	Peripherals versus Centrals
Contrast 2	–	Left vs Right	Peripheral anterior vs peripheral posterior
Contrast 3	–	–	Central & Frontocentral vs Frontal
Contrast 4		–	Central vs Frontocentral

two electrodes to each side, finding different effects depending on the choice of electrodes. Alexandrov et al. (2011) used only electrodes Fz, Cz and Pz, plus one electrode to each side, for their analysis. Leminen et al. (2013) used 48 electrodes spread across frontal, central and parietal regions, while Hanna & Pulvermüller (2014) used the global root mean square of all 64 electrodes. These studies serve as an example of the great variability there is in the field of syntactic MMN studies when it comes to the choice of electrodes to use.

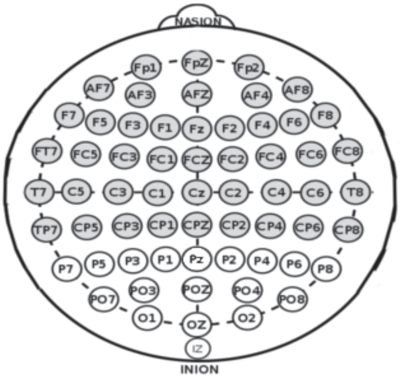

Figure 3 Electrodes included in the analysis of the MMN component in yellow

For our choice of electrodes for analysis, which included 7 midline electrodes and 19 electrodes that expanded to each hemisphere (see Figure 3). These 19 electrodes at each side can be ordered in 5 regions. The frontal, frontocentral, central and centroparietal regions all have 4 electrodes set out in order of closeness to the midline, with the exception of the Anteriofrontal electrodes, which have only two, and the frontopolar section, with only one electrode to the side of the midline.

Whenever a statistical model resulted in significant interactions of lexicality/grammaticality and one of the other two possible fixed effects (hemisphere and region), linear mixed effects models were constructed on subsets of the data to examine the interaction. In order to compare the latency of the MMN peaks, we built linear mixed effects models on the peak amplitude of the first peak of each electrode.

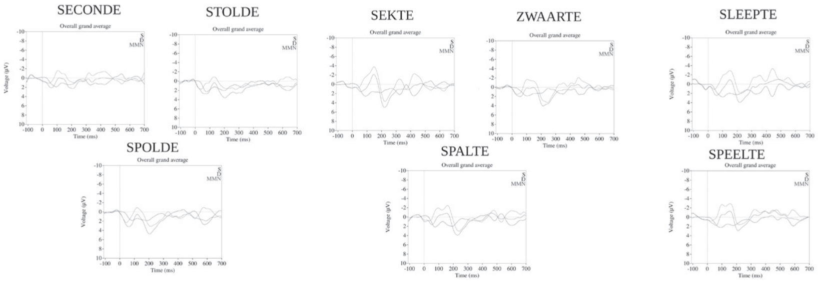

Figure 4. ERP plots of all stimuli used. Top: Simple and Complex Dutch words. Bottom: Pseudowords and ungrammatical verb used as controls. *Seconde* and *stolde* were compared to *spolde*; *sekte* and *zwaarte* were compared to *spalte*; *sleepte* was compared to *speelte*

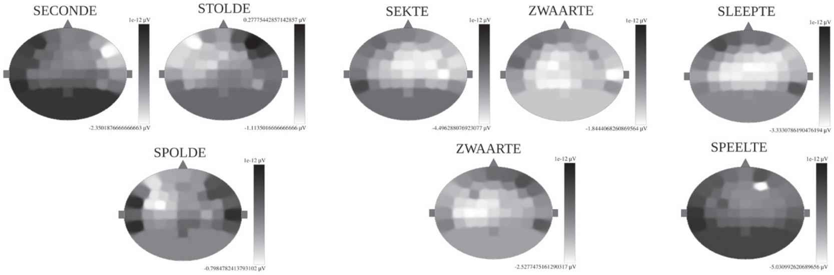

Figure 5. Scalp maps of all stimuli used. Top: Simple and Complex Dutch words. Bottom: Pseudowords and ungrammatical verb used as controls. *Seconde* and *stolde* were compared to *spolde*; *sekte* and *zwaarte* were compared to *spalte*; *sleepte* was compared to *speelte*

7. Results

Figure 4 shows the ERP plot for every word (top row) and their corresponding pseudoword or ungrammatical control (bottom row). Figure 5 shows the scalp maps for each word and its control stimulus.

7.1 MMN on monomorphemic Dutch nouns

The best model to account for the difference in MMN amplitude of the earliest peak between the monomorphemic noun ***seconde*** and the pseudoword ***spolde*** included lexicality, hemisphere, region and a lexicality-hemisphere interaction as fixed effects (Table 3).

Although *seconde* was estimated to be more negative than *spolde*, their estimated difference was not significant (0.32 microvolts, t [13.23]= 0.69, p =

Table 3. Statistical model for the difference between the noun *seconde* and the pseudoword *spolde*

	Estimate	Std. Error	df	t value	p value
(Intercept)	-1.83	0.14	12.50	-13.23	< 0.001
Word vs Pseudoword	0.32	0.47	13.23	0.69	0.50
Right vs Left Hemisphere	0.19	0.13	12.41	1.42	0.18
Right & left Hemisphere vs Midline	-0.05	0.09	20.56	-0.55	0.59
Centrals versus Peripherals	0.03	0.09	15.22	0.35	0.73
Peripheral anterior vs Peripheral posterior	0.50	0.29	14.75	1.71	0.11
Central & FrontoCentral vs Frontal	-0.15	0.12	16.81	-1.27	0.22
Central vs Frontal	-0.14	0.09	25.76	-1.53	0.14
Word vs Pseudoword:Right vs left	-0.54	0.21	14.61	-2.54	0.02
Word vs Pseudoword:Right & left Hemisphere vs Midline	-0.02	0.19	16.92	-0.13	0.90

0.50 with a 95% confidence interval from -0.56 to 1.22 microvolts). Although hemisphere and region improved the model, the coefficient of their effects did not reach significance.

The interaction of lexicality and the right versus left hemisphere contrast was significant (*t*=2.54, p=0.02). Linear mixed effect models were run on the left and right hemisphere data separately to inspect the interaction. We found that on the left hemisphere the difference between the singular noun and its pseudoword control was very small (0.06 microvolts with a 95% confidence interval from -0.78 to 0.90 microvolts) and not significant (*t* [13.28] = 0.14, *p* = 0.89). For the right hemisphere, the model estimated that the singular noun *seconde* produced a more negative MMN response than its pseudoword control by 0.56 microvolts, with a 95% confidence interval from -0.50 to 1.63 microvolts. The difference, although larger than in the left hemisphere, also failed to reach significance (*t* [14.73] = 1.04, *p* = 0.36). This directionality of the effect is in line with the predictions of the lexical MMN response.

The average amplitude of the MMN response across hemispheres for *seconde* and *spolde* can be seen in Figure 6. It is clear that the response to the singular noun was more negative than the response to the pseudoword in the right hemisphere.

In terms of latency, a linear mixed effects model with lexicality as a fixed effect and as a random slope for the random intercept of subject estimated that both stimuli peaked at almost the same time, with the noun *seconde* peaking only 1.73 milliseconds later than the control pseudoword. This

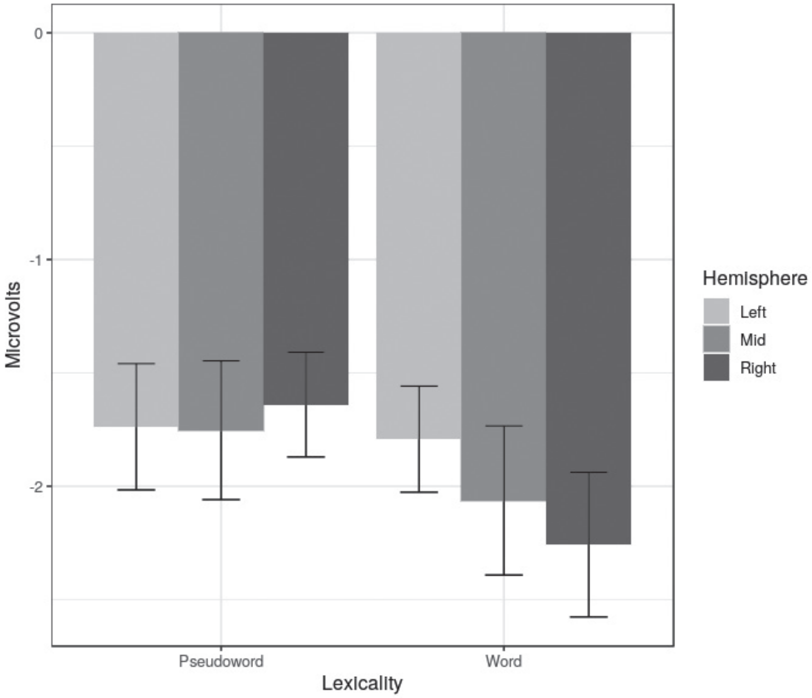

Figure 6. Average amplitude of the MMN response for word *seconde* and its control pseudoword across hemispheres

difference was, unsurprisingly, not statistically significant (t [7.42] = 0.24, $p = 0.82$ with a 95% confidence interval from -12.22 to 15.67 milliseconds).

Sekte. The best model to account for the difference in MMN amplitude between the MMN of the monomorphemic noun *sekte* and its pseudoword control *spalte* included lexicality, hemisphere, region and an interaction between lexicality and hemisphere as fixed effects (see Table 4). The model estimated *sekte* to have produced a more negative MMN amplitude than *spalte* by 0.64 microvolts. This difference was not significant (t [11.48] = 1.20, $p = 0.26$ with a 95% confidence interval from -0.41 to 1.68 microvolts).

None of the coefficients for hemisphere reached significance. They all, however, showed the same trend: Small differences in between left and right hemispheres, and the midline having the most negative MMN responses. The effect of region was due to the MMN being more negative at central than peripheral electrodes by 0.67 microvolts. This difference was significant (t [16.34] = 5.00, $p < 0.001$ with a 95% confidence interval from 0.41 to 0.93 microvolts).

Table 4. Statistical model for the difference between the noun *sekte* and the pseudoword *spalte*

	Estimate	Std. Error	df	t value	p value
(Intercept)	-3.72	0.46	14.63	-8.07	< 0.001
Word vs Pseudoword	0.64	0.53	11.48	1.20	0.26
Right vs Left Hemisphere	0.03	0.19	12.61	0.15	0.88
Right & left Hemisphere vs Midline	-0.27	0.14	16.39	-1.94	0.07
Centrals versus Peripherals	0.67	0.13	16.34	5.00	< 0.001
Peripheral anterior vs Peripheral posterior	0.38	0.27	16.61	1.40	0.18
Central & FrontoCentral vs Frontal	-0.03	0.14	16.75	-0.22	0.83
Central vs Frontal	-0.19	0.13	19.28	-1.39	0.18
Word vs Pseudoword:Right vs Left Hemisphere	-1.06	0.31	12.68	-3.35	0.01
Word vs Pseudoword:Sides vs Midline	0.09	0.20	25.92	0.42	0.68

As with the previous singular noun, here we also observed an interaction between the right versus left hemisphere contrast and lexicality (*t* [12.68] = 3.35, *p* = 0.01). Linear mixed effect models run on the right and left hemisphere separately revealed in both hemispheres that the word *sekte* produced a more negative response than its pseudoword control. However, at the left hemisphere this difference was small and not significant (0.24 microvolts. *t*[9.81] = 0.39, *p* = 0.71 with a 95% confidence interval from -0.98 to 1.47 microvolts). At the right hemisphere, the difference was in the same direction (word more negative than pseudoword), larger (1.40 microvolts) and reached statistical significance (*t* [9.88] = 2.75, *p* = 0.02 with a 95% confidence interval from 0.40 to 2.40 microvolts). Figure 7 shows that the word *sekte* produced a more negative MMN than the pseudoword *spalte* at the midline and the right hemisphere, while the MMN responses were quite similar in the left hemisphere.

In terms of latency, a linear mixed effects model with lexicality as a fixed effect and as a random slope for the random intercept of subject estimated that the the noun peaked later than the pseudoword by 7.62 milliseconds. This difference did not reach significance (*t* [10.68] = 0.92, *p* = 0.38).

7.2 MMN on Past tense verbs

The best model to account for the difference in MMN amplitude between the past tense ***stolde*** and the pseudoword ***spolde*** was a model with lexicality, hemisphere, region and a lexicality-region interaction as fixed effects. We predicted that the grammatical past tense ***stolde*** should produce a smaller

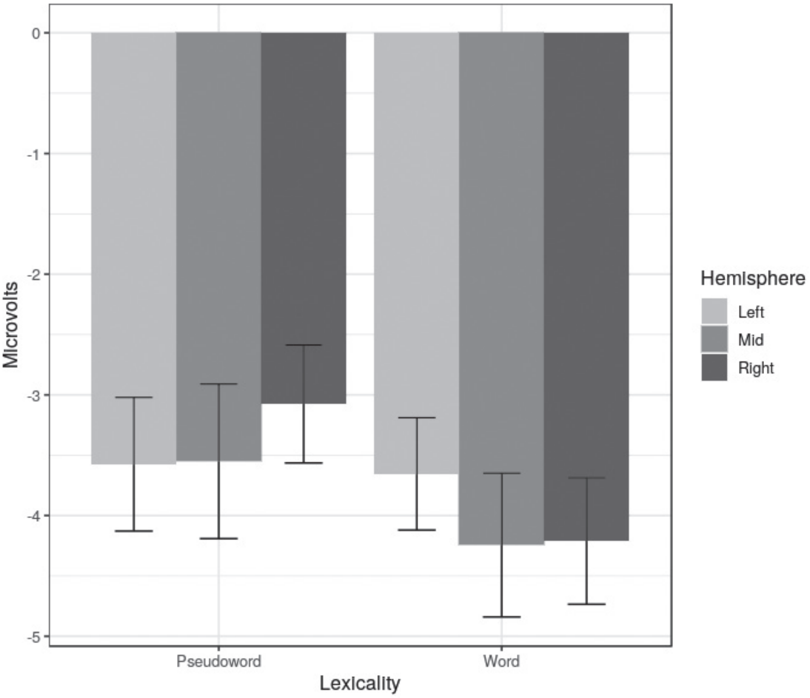

Figure 7. Average amplitude of the MMN response for word sekte and its control pseudoword across hemispheres

MMN than the pseudoword control. As it turns out, the response to ***stolde*** was slightly more negative than its pseudoword control by 0.21 microvolts. This difference, however, was small and did not reach significance (***t*** [9.37] = 0.64, ***p*** = 0.54 with a 95% confidence interval from -0.44 to 0.86 microvolts). The past tense ***stolde*** did not produce MMN responses that reliably differed from the pseudoword control. None of the other coefficients of the model, including the coefficients for the region contrasts, were significant (see Table 5).

We examined the interaction between lexicality and the central versus peripheral contrast, since it was the coefficient closest to a significant effect (*t* [13.58] = 1.21 , *p* = 0.25). We carried out linear mixed effect models on central (frontal, frontocentral and central) and peripherals (anterofrontal and centroparietal) electrodes separately. The models estimated that on central electrodes, the past tense *stolde* produced a slightly more negative MMN response than its pseudoword control by 0.07 microvolts. The difference was not significant (*t* [12.16] = 0.18, *p* = 0.86 with a 95% confidence interval from -0.67 to 0.81 microvolts). On peripheral electrodes, we found the opposite effect.

Table 5. Statistical model for the difference between the past tense *stolde* and the pseudoword *spolde*

	Estimate	Std. Error	df	t value	p value
(Intercept)	-1.64	0.29	13.48	-5.57	< 0.001
Word vs Pseudoword	0.19	0.33	9.05	0.57	0.58
Centrals versus Peripherals	0.06	0.08	15.58	0.83	0.42
Peripheral anterior vs Peripheral posterior	0.33	0.19	14.16	1.69	0.11
Central & FrontoCentral vs Frontal	0.02	0.12	12.21	0.18	0.86
Central vs Frontal	-0.06	0.09	11.66	-0.73	0.48
Right vs Left Hemisphere	-0.09	0.14	15.96	-0.67	0.51
Right & left Hemisphere vs Midline	-0.03	0.11	16.70	-0.26	0.80
Word vs Pseudoword:Centrals versus Peripherals	-0.16	0.13	13.58	-1.21	0.25
Word vs Pseudoword:Peripheral anterior vs Peripheral posterior	0.38	0.36	15.97	1.05	0.31
Word vs Pseudoword:Central & FrontoCentral vs Frontal	-0.10	0.20	18.40	-0.48	0.64
Word vs Pseudoword:Central vs Frontal	-0.03	0.18	17.92	-0.17	0.87

The model estimated that the past tense verb yielded a less negative MMN than its pseudoword control by 0.09 microvolts. The directionality of the effect is in accordance with the predictions of the syntactic MMN, but the effect size was small and not significant (*t* [11.73] = 0.29, *p* = 0.78 with a 95% confidence interval from -0.72 to 0.53 microvolts). Figure 8 shows the average amplitude for each region for the past tense *stolde* and its pseudoword control *spolde*.

In terms of latency, a linear mixed effects model with lexicality as a fixed effect and as a random slope for the random intercept of subject estimated that the past tense *stolde* peaked 5.52 milliseconds later than the control pseudoword *spolde*. This difference did not reach significance (*t* [11.02] = 1.08, *p* = 0.31 with a 95% confidence interval from -15.56 to 4.53 milliseconds).

The best model to account for the difference in MMN amplitude between the grammatical past tense ***sleepte*** and the ungrammatical ****speelte*** has lexicality and hemisphere as fixed effects (see Table 6). Region could not be included in this model because of convergence errors that could not be solved with the available tools in the lmerTest package. The model estimated that the grammatical past tense ***sleepte*** produced a more negative MMN amplitude than the ungrammatical ****speelte*** by 0.42 microvolts. This difference was not significant (***t*** [15.58] = 0.58, ***p*** = 0.57 with a 95% confidence interval from

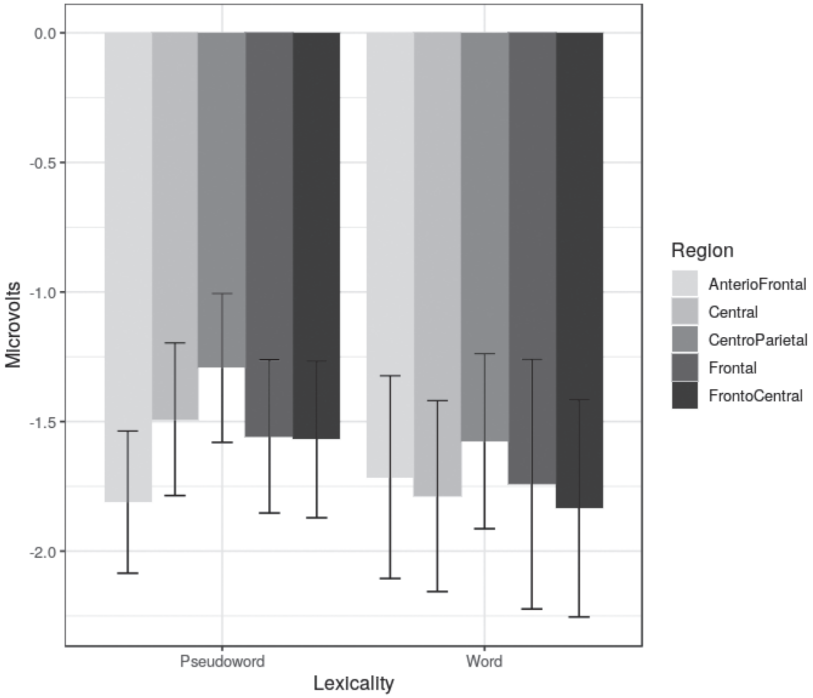

Figure 8. Average amplitude of the MMN response for word *stolde* and its control pseudoword across regions

-1.01 to 1.86 microvolts). The effect of hemisphere was due to a significant difference between the right versus the left hemisphere, and between the average of the right and left hemisphere versus the midline. Specifically, the MMN response was 0.46 microvolts more negative in the right than in the left hemisphere. The difference was significant (***t*** [15.12] = 2.51, ***p*** = 0.02, with a 95% confidence interval from 0.10 to 0.83 microvolts). The MMN was also more negative in the midline than in the average of the right and left hemisphere by 0.50 microvolts. The difference was significant (***t*** [18.66] = 3.18, ***p*** = 0.005 with a 95 % confidence interval from -0.80 to -0.19 microvolts). Figure 9 shows the estimated average amplitude of each MMN response for ***sleepte*** and its ungrammatical control ***speelte**** across hemispheres. Regarding the latency of the first peak, a linear mixed effects model with grammaticality as a fixed effect and as a random slope for the random intercept of subject estimated that the grammatical ***sleepte*** peaked 3.64 milliseconds after the ungrammatical ****speelte.*** This difference was not significant (***t*** [10.45] = 0.49, ***p*** = 0.64 with a 95% confidence interval from -18.27 to 11.00 milliseconds).

Table 6. Statistical model for the difference between the past tense *sleepte* and the ungrammatical *speelte*

	Estimate	Std. Error	df	t value	p value
(Intercept)	-3.82	0.32	11.00	-11.80	< 0.001
Grammatical vs Ungrammatical	0.42	0.73	15.58	0.58	0.57
Right vs Left Hemisphere	0.46	0.18	15.21	2.51	0.02
Right & Left Hemispheres vs Midline	-0.50	0.16	18.66	-3.18	< 0.001

Table 7. Statistical model for the difference between the derived noun *zwaarte* and the pseudoword *spalte*

	Estimate	Std. Error	df	t value	p value
(Intercept)	-3.34	0.33	13.60	-9.99	0.00
Word vs Pseudoword	-0.47	0.57	9.08	-0.81	0.44
Right vs Left Hemisphere	-0.19	0.18	11.73	-1.06	0.31
Right & left Hemisphere vs Midline	-0.38	0.11	14.05	-3.51	0.003
Centrals versus Peripherals	0.38	0.10	13.46	3.63	0.003
Peripheral anterior vs Peripheral posterior	0.09	0.21	13.33	0.44	0.66
Central & FrontoCentral vs Frontal	-0.06	0.12	13.61	-0.49	0.63
Central vs Frontal	-0.04	0.12	15.98	-0.37	0.71
Word vs Pseudoword:Right vs Left Hemisphere	-0.54	0.16	9.31	-3.33	0.008
Word vs Pseudoword:Sides vs Midline	0.11	0.21	12.45	0.50	0.63

7.3 MMN on the derived noun

The best model to account for the difference in MMN amplitude between the derived noun *zwaarte* and the pseudoword *spalte* included lexicality, hemisphere, region and a lexicality-hemisphere interaction as fixed effects (see table 7). The derived noun was estimated to be less negative than the pseudoword control by 0.47 microvolts. This difference, however, was not significant (t [9.08] = 0.81, p = 0.44 with a 95% confidence interval from -1.59 to 0.66 microvolts). The effect of hemisphere was due to responses being more negative in the midline than in the average of the left and right hemisphere by 0.38 microvolts. This difference was significant (t [14.05] = 3.51, p = 0.003 with a 95% confidence interval from -0.59 to -0.17 microvolts). The effect of region was due to a significant difference between central electrodes (frontal, frontocentral and central regions) and peripheral electrodes (anterofrontal and centroparietal electordes), with central electrodes having a more negative MMN by 0.38 microvolts (t [13.46] = 3.63, p = 0.003, with a 95% confidence interval from 0.17 to 0.58 microvolts).

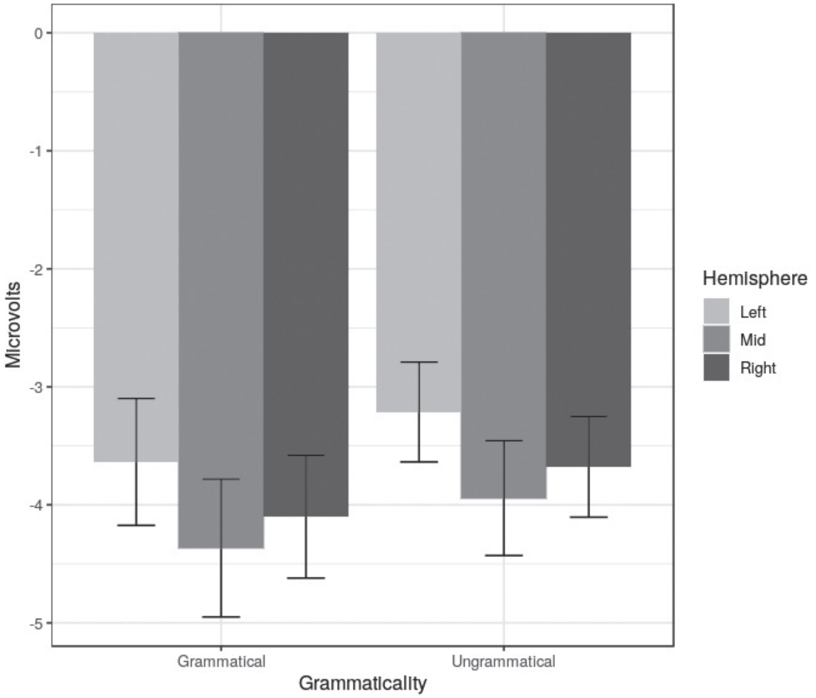

Figure 9. Average amplitude of the MMN response for word *sleepte* and its ungrammatical control across hemispheres

The interaction between lexicality and hemisphere was due to the right versus left hemisphere contrast (*t* [9.31] = 3.33, *p* = 0.008). Linear mixed effects models were carried out on each hemisphere separately to inspect the interaction. The models showed that on the left hemisphere the derived word *zwaarte* was less negative than its pseudoword control by 0.71 microvolts. This difference was not significant (*t* [11.32] = 1.40, *p* = 0.19 with a 95% confidence interval from -1.70 to 0.28 microvolts). On the right hemisphere, the difference between the derived word and the pseudoword followed the same direction (derived word producing a less negative response than the control pseudoword), but the difference was smaller (0.20 microvolts), and also not significant (*t* [10.83] =0.39 , *p* = 0.70 with a 95% confidence interval from -1.20 to 0.80 microvolts). Figure 10 shows the estimated amplitude of each stimulus across hemispheres.

In terms of latency, a linear mixed effects model with lexicality as a fixed effect and as a random slope for the random intercept of subject estimated that the derived word *zwaarte* peaked 7.49 milliseconds earlier than its

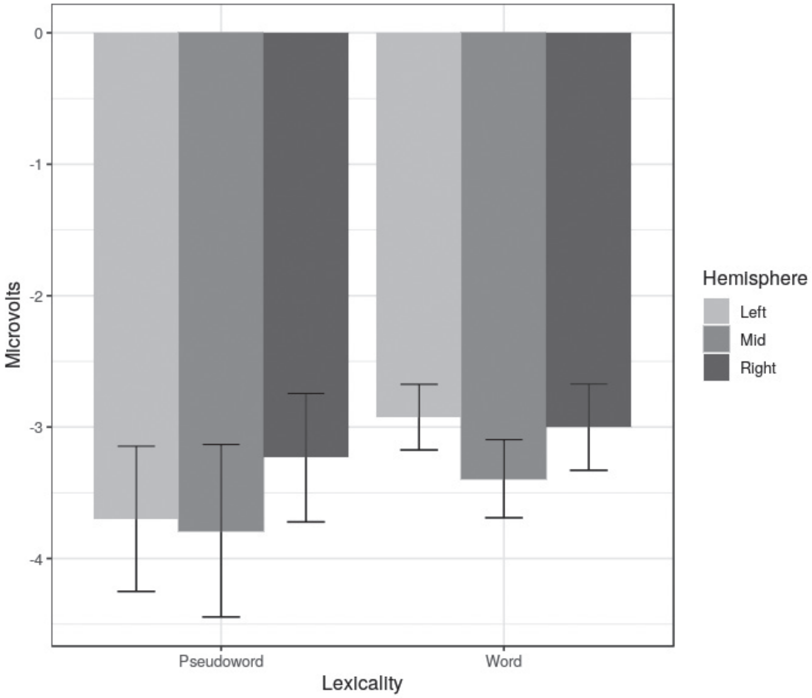

Figure 10. Average amplitude of the MMN response for the derived word *zwaarte* and its ungrammatical control across hemispheres

pseudoword control. The difference was not significant (t [10.71] = 0.98 , p = 0.34 with a 95% confidence interval from -7.2 to 22.21 milliseconds).

8. Conclusions

One of the goals of our research was to test the use of the MMN response to assess the strength of memory traces of words in Dutch, as previous research had shown for other languages (Pulvermüller & Shtyrov 2006). The purpose of such a test was to confirm that MMN response can distinguish sound strings that are present in Dutch speakers' lexical memory from those sound strings that do not have memory traces in lexical memory. Our interest in testing this research paradigm rested on the assumption that words which are decomposed in the earliest stages of word processing should have weaker memory traces than non-existing sound strings like pseudowords and ungrammatical words, as previous research had shown

Table 8. Predictions and results of first MMN peak

Condition	MMN Prediction	MMN Result
seconde vs *spolde*	Lexical MMN	Non-significant right hemisphere lexical MMN
sekte vs *spalte*	Lexical MMN	Significant right hemisphere lexical MMN
stolde vs *spolde*	Syntactic MMN	Negligible, non-significant syntactic MMN at peripehral anterior
sleepte vs *speelte*	Syntactic MMN	Non-significant lexical MMN
zwaarte vs *spalte*	Lexical MMN	Non-significant syntactic MMN at left and right hemipsheres

(Bakker et al. 2013; Leminen et al. 2013). Therefore, by showing that the MMN paradigm can distinguish sound strings that are stored in the lexicon from sound strings that are not, we can be sure the MMN response is tapping the memory trace of stored lexical items. With this in mind, we set out to compare the amplitude of MMN responses elicited by two monomorphemic nouns (*seconde* and *sekte*) and three morphologically complex Dutch words: two past-tense verbs (*stolde* and *sleepte*) and one derived noun (*zwaarte*).

The MMN responses elicited by our experimental stimuli were distributed across a wide range of electrodes, spanning 5 regions and the left and right hemispheres plus the midline. Our analyses showed that the MMN response varies considerably across regions and hemispheres, and that taking this variation into account can influence how words differ from their controls at the early stages of processing that are tapped by the MMN response.

Table 8 shows our predictions and our results for the first MMN peak. The table is a confirmation that when using the lexical and syntactic MMN responses to predict experimental outcomes, they should be studied in detail across the brain.

Regarding our monomorphemic words, we expected existing words should produce, according to the literature on lexical MMN (Pulvermüller & Shtyrov 2006), more negative MMN amplitudes than their pseudoword controls. Although both *seconde* and *sekte* were estimated to have more negative MMN than their controls, the difference was not significant. However, when taking into account the variation of the response across the brain, we found evidence that suggests the existence of lexical MMN responses on the right hemisphere for both words (although the effect was significant only for *sekte*). We believe the consistency in the location of both lexical MMN supports the hypothesis that this hemisphere-specific effect was due to the processing of monomorphemic words, and not because of other non-lexical factors, such as acoustic features of the stimuli.

The results of our past tense verbs do not allow for a straightforward interpretation. Based on previous results and discussions in the MMN literature, we defined the syntactic MMN as a difference between grammatical and ungrammatical/pseudoword stimuli, such that the former yield smaller (less negative) MMN amplitudes than the latter. This was not always the case in our analysis. However, a more in-depth analysis of the MMN elicited by each verb does show interesting results.

The regular past tense *stolde* did produce a smaller MMN than its pseudoword (i.e. a syntactic MMN) only at the peripheral anterior region. This difference, however, was small and not significant. This alone is, of course, not enough evidence to claim this past tense has no surface form representation. The other regular past tense, *sleepte*, showed less evidence of being decomposed. In fact, the general analysis of this condition revealed no clear pattern, with both stimuli, the past tense and its ungrammatical control, showing similar responses.

Our derived noun, *zwaarte* also showed unexpected results. While there are theories that assume all morphologically complex words, including derivations, are decomposed (Taft 2004, Stockall & Marantz 2006), most experimental evidence, especially within the MMN field, suggests derived words have surface form lexical representations (Bozic & Marslen-Wilson 2010; Leminen et al. 2013; Hanna & Pulvermüller 2014). The same is the case for previous psycholinguistic evidence from Dutch, which showed that derived words have surface form representations (Bertramet al. 2000). Consequently, we expected our derived noun to produce a lexical MMN. Our statistical models, however, estimated *zwaarte* to have produced smaller, less negative MMN responses than its pseudoword control, especially on the left hemisphere. The estimated voltage of the MMN response to *zwaarte* suggests, following the premise of the syntactic MMN, that the surface form of the word might not be part of Dutch speakers' lexical memory, and that, therefore, the derived noun is decomposed into its constituent elements when being processed. Because the estimated voltage difference between *zwaarte* and its pseudoword control was not significant, we cannot be certain there really was a syntactic MMN response for this derived word.

We were surprised at the small effect sizes of latency. All our tests showed considerably small differences between the peaks of each MMN. Since the literature on decomposition assumes this process takes additional processing time, we expected to find later latencies for words that produced syntactic MMN responses. However, from the words that could have produced a syntactic MMN, only *stolde* peaked after its control pseudoword. What is more, the complex word that showed the most evidence of decomposition

– *zwaarte* – peaked earlier than its pseudoword control. Our results of the latency of the peak of complex words is therefore incompatible with the view that decomposed words take longer to process than monomorphemic words. It is worth noting, however, that none of the latency differences were significant. We cannot, therefore, take the latency differences as true differences between our stimuli. Another possibility is that that the nature of the MMN paradigm, which requires participants to listen to the critical stimuli several times, might cancel out the extra processing time required to processes newly seen pseudowords.

Overall, this experiment shows that existing words produce more negative responses than control pseudowords. In this respect, the MMN can be considered as a useful tool for examining lexical memory trace representations. However, it seems to us that the use of linguistically complex stimuli in this paradigm requires new approaches to the analysis of MMN differences between experimental and control stimuli. Indeed, most of the results that were of interest to us were found thanks to the inclusion of hemisphere and region variation in our statistical models. Had we followed conventional choices of analysis, such as using the midline electrode(s) only, we would have missed a great deal of information.

Regarding morphologically complex words, we believe that while our evidence does not fully support the predictions based on a decompositional account, it does give enough evidence to assume these words are processed differently from monomorphemic ones. The results of the past tense verbs could be interpreted as being in line with the very general claim of dual-mechanism models that words belonging to the same inflectional paradigm can be processed by different mechanisms (Pinker 1999). This is because *stolde* showed partial evidence of being decomposed (i.e. having a weaker memory trace than its pseudoword control), but *sleepte* showed evidence of being stored in its surface form. The verb *stolde* has a relative low-frequency and is certainly less known by the average Dutch speaker than the verb *sleepte*. Following the assumptions of, for example, the Words & Rules model, it could be the case that only low-frequency, uncommon verbs are decomposed, while more commonly used ones can have surface form lexical representations. However, this is speculation, and more research needs to be carried out to study the differences, if there are any, between frequency and infrequent regular past tense verbs.

Following the same reasoning, we could try to account for the unexpected syntactic MMN elicited by the derived noun *zwaarte*. The rule that turns adjectives into nouns with the *-te* is fairly uncommon. Dutch uses the suffixes *-heid*, *-ist* and -(*i*)*teit*, to name a few, more commonly than *-te* to create nouns.

A low frequency of occurrence of this derivational paradigm could favor the decompositional route for such words. However, this is mere speculation, and more research needs to be carried out with derivational morphology in Dutch to know how these words are stored in the mental lexicon of Dutch speakers.

Our results also suggest that previous findings in Dutch were influenced by the lack of an appropriate temporal resolution used in behavioral tasks. Indeed, the lack of evidence for decomposition in Dutch was puzzling, compared to that of other languages. However, that lack of evidence also correlated with the lack of masked priming studies, and of paradigms that were better suited to tap into early stages of word processing. The results presented in this study suggest that there might be decomposition in Dutch after all. Certainly, confirming evidence from other experimental paradigms that can also tap into the earliest stages of word processing would strengthen the hypothesis that Dutch makes use of decompositional mechanism to process morphologically complex words.

This study also has its own limitations that could account for the unexpected results presented here. The critical words and their pseudoword controls were made to be as acoustically similar as possible. However, because of the constrains of the paradigm, and the need to use meaningful words, a complete control of acoustic variables between test items and their controls was not possible. This could have had an effect on the results that we have presented.

Overall, the MMN paradigm stands as a promising tool to probe the earliest stages of lexical processing. However, more studies are needed across languages to know how the MMN component reacts to morphologically simple and complex words across languages. Using detailed analysis of the MMN responses could contribute to a better understanding of the mechanisms underlying the lexical and syntactic MMN, and avoid losing studies because of null results.

References

Alegre, M. & P. Gordon (1999). Frequency effects and the representational status of regular inflections. *Journal of memory and language* 40(1), 41–61.

Alexandrov, A.A., D.O. Boricheva, F. Pulvermüller & Y. Shtyrov (2011). Strength of word-specific neural memory traces assessed electrophysiologically. *PLoS one* 6(8), e22999.

Baayen, R.H., T. Dijkstra & R. Schreuder (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language* 37(1), 94–117.

Baayen, R.H., P. Milin, D.F. Đurđević, P. Hendrix & M. Marelli (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118(3), 438-481.

Badecker, W. & A. Caramazza (1989). A lexical distinction between inflection and derivation. *Linguistic Inquiry* 20(1), 108–116.

Baguley, T (2012). *Serious stats: A guide to advanced statistics for the behavioral sciences*. Hampshire, NJ: Palgrave Macmillan.

Bakker, I., L.J. MacGregor, F. Pulvermüller & Y. Shtyrov (2013). Past tense in the brain's time: Neurophysiological evidence for dual-route processing of past-tense verbs. *Neuroimage* 71, 187–195.

Barr, D.J., R. Levy, C. Scheepers & H.J. Tily (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3), 255–278.

Bertram, R., R. Schreuder & R.H. Baayen (2000). The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy, and productivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26(2): 489–511.

Bod, R., J. Hay & S. Jannedy (2003). *Probabilistic Linguistics*. Cambridge, MA: MIT Press.

Boersma P. & D. Weenink (2001). *Praat: Doing Phonetics by Computer.* <http://www.fon.hum.uva.nl/praat/>

Bozic, M. & W. Marslen-Wilson (2010). Neurocognitive contexts for morphological complexity: Dissociating inflection and derivation. *Language and Linguistics Compass* 4(11), 1063–1073.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* 10(5), 425–455.

Cappelle, B., Y. Shtyrov & F. Pulvermüller (2010). Heating up or cooling up the brain? MEG evidence that phrasal verbs are lexical units. *Brain and Language* 115(3), 189-201.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and brain sciences* 22 (6), 991–1013.

Clahsen, H. & K. Neubauer (2010). Morphology, frequency, and the processing of derived words in native and non-native speakers, *Lingua* 120 (11), 2627–2637.

Diependaele, K., J.A. Duñabeitia, J. Morris & E. Keuleers (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language* 64(4), 344–358.

Embick, D. & R. Noyer (2007). Distributed morphology and the syntax/morphology interface. In: G. Ramchand & C. Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*. Oxford: Oxford University Press, 289–324.

Ernestus, M. & H. Baayen (2001). Choosing between the Dutch past-tense suffixes *-te* and *-de*. In: T. van der Wouden & H. de Hoop (eds.), *Linguistics in the Netherlands 2001*. Amsterdam: John Benjamins, 77–88.

Ernestus, M. & R.H. Baayen (2004). Analogical effects in regular past tense production in Dutch. *Linguistics*, 873–904.

Forster, K.I. (1999). The microgenesis of priming effects in lexical access. *Brain and language* 68(1), 5–15.

Fruchter, J., & A. Marantz (2015). Decomposition, lookup, and recombination: MEG evidence for the full decomposition model of complex visual word recognition. *Brain and Language* 143, 81–96.

Hanna, J., S. Mejias, M.-A. Schelstraete, F. Pulvermüller, Y. Shtyrov & H.K. van der Lely (2014). Early activation of Broca's area in grammar processing as revealed by the syntactic mismatch negativity and distributed source analysis. *Cognitive Neuroscience* 5(2), 66–76.

Hanna, J. & F. Pulvermüller (2014). Neurophysiological evidence for whole form retrieval of complex derived words: a mismatch negativity study. *Frontiers in Human Neuroscience* 8, 53–65.

Hanna, J., B. Cappelle & F. Pulvermüller (2017). Spread the word: MMN brain response reveals whole-form access of discontinuous particle verbs. *Brain and Language* 175, 86–98.

Jong, N.H. de, R. Schreuder & R.H. Baayen (2000). The morphological family size effect and morphology. *Language and Cognitive Processes* 15(4-5), 329–365.

Kuznetsova A., P. B. Brockhoff & R.H.B. Christensen (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13), 1–26.

Langacker, R. (2017). Entrenchment in Cognitive Grammar. In: H.-J. Schmid (ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge*. Berlin: Mouton de Gruyter, 39–56.

Leminen, A., M. Leminen, T. Kujala & Y. Shtyrov (2013). Neural dynamics of inflectional and derivational morphology processing in the human brain. *Cortex* 49(10), 2758–2771.

Lignos, C. & K. Gorman (2012). Revisiting frequency and storage in morphological processing. *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 48, 447–461.

MacGregor, L.J. & Y. Shtyrov (2013). Multiple routes for compound word processing in the brain: evidence from EEG. *Brain and language* 126 (2), 217–229.

Marantz, A. (2013). No escape from morphemes in morphological processing. *Language and Cognitive Processes* 28(7), 905–916.

Marslen-Wilson, W.D., M. Ford, L. Older & X. Zhou (1996). The combinatorial lexicon: Priming derivational affixes. In: G.W. Cottrell (ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*. New York: Routledge, 223-227.

Marslen-Wilson, W.D. & L.K. Tyler (2007). Morphology, language and the brain: the decompositional substrate for language comprehension. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362(1481), 823–836.

Marslen-Wilson, W.D. & L.K. Tyler (2008). Morphology, language and the brain: The decompositional substrate for language comprehension. In: J. Driver, P.

Haggard & T. Shallice (eds.), *Mental processes in the human brain*. Oxford: Oxford University Press, 99–121.

Morris, J. & L. Stockall (2012). Early, equivalent ERP masked priming effects for regular and irregular morphology. *Brain and language* 123 (2), 81–93.

Pinker, S. (1999). *Words and rules.* New York: Basic Books.

Pinker, S. & M. T. Ullman (2002). The past and future of the past tense. *Trends in cognitive sciences* 6(11), 456–463.

Pulvermüller, F. & Y. Shtyrov (2003). Automatic processing of grammar in the human brain as revealed by the mismatch negativity. *Neuroimage* 20(1), 159–172.

Pulvermüller, F. & Y. Shtyrov (2006). Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Progress in neurobiology* 79(1), 49–71.

R Core Team (2013). *R: A language and environment for statistical computing.* <http://www.R-project.org/>

Rastle, K. & M. H. Davis (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes* 23 (7-8), 942–971.

Rumelhart, D.E. & J.L. McClelland (1988). *Parallel distributed processing*. Cambridge, MA: MIT Press.

Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In: R.H. Baayen & R. Schreuder (eds.), *Morphological structure in language processing*. Berlin: De Gruyter Mouton, 355–390.

Schreuder, R. & R.H. Baayen (1997). How complex simplex words can be. *Journal of Memory and Language* 37(1), 118–139.

Schriefers, H., A. Friederici & P. Graetz (1992). Inflectional and derivational morphology in the mental lexicon: Symmetries and asymmetries in repetition priming. *The Quarterly Journal of Experimental Psychology* 44(2), 373–390.

Shtyrov, Y. & F. Pulvermüller (2002). Memory traces for inflectional affixes as shown by mismatch negativity. *European Journal of Neuroscience* 15(6), 1085–1091.

Stockall, L. & A. Marantz (2006). A single route, full decomposition model of morphological complexity: MEG evidence. *The Mental Lexicon* 1(1), 85–123.

Tabak, W., R. Schreuder & R.H. Baayen (2005). Lexical statistics and lexical processing: semantic density, information complexity, sex, and irregularity in Dutch. In: S. Kepser & M. Reis (eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives*. Berlin: De Gruyter Mouton, 529–555.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology, Section A* 57(4), 745–765.

Taft, M. & K.I. Forster (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior* 14(6), 638–647.