

The traces that novel morphologically complex words leave in memory are abstract in nature

Laura de Vaan¹, Kobie Van Krieken¹, Winie Van den Bosch¹, Robert Schreuder[†] and Mirjam Ernestus^{1,2}

¹Centre for Language Studies, Radboud University Nijmegen / ²Max Planck Institute for Psycholinguistics, Nijmegen / [†]Deceased. Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen

Previous work has shown that novel morphologically complex words (henceforth neologisms) leave traces in memory after just one encounter. This study addressed the question whether these traces are abstract in nature or exemplars. In three experiments, neologisms were either primed by themselves or by their stems. The primes occurred in the visual modality whereas the targets were presented in the auditory modality (Experiment 1) or vice versa (Experiments 2 and 3). The primes were presented in sentences in a self-paced reading task (Experiment 1) or in stories in a listening comprehension task (Experiments 2 and 3). The targets were incorporated in lexical decision tasks, auditory or visual (Experiment 1 and Experiment 2, respectively), or in stories in a self-paced reading task (Experiment 3). The experimental part containing the targets immediately followed the familiarization phase with the primes (Experiment 1), or after a one week delay (Experiments 2 and 3). In all experiments, participants recognized neologisms faster if they had encountered them before (identity priming) than if the familiarization phase only contained the neologisms' stems (stem priming). These results show that the priming effects are robust despite substantial differences between the primes and the targets. This suggests that the traces novel morphologically complex words leave in memory after just one encounter are abstract in nature.

Keywords: morphologically complex neologisms, abstract lexical representations, exemplars, modality non-specificity, long-term priming

Introduction

Listeners and readers often encounter neologisms. Most of these neologisms are morphologically complex words (e.g., Baayen and Renouf, 1996; Bloom, 2000; Clark, 1993), whose meanings can be derived from the containing morphemes in combination with morphological rules (e.g., *fine+able*) or analogy (e.g., *protologism* from Greek *protos* 'first' plus Greek *logos* 'word' by analogy with *prototype* and *neologism*, Epstein, 2012). The question arises whether these neologisms are stored in the mental lexicon after just one encounter. Pinker (1991) argued that no regular complex words are stored in the mental lexicon. However, De Vaan, Schreuder, and Baayen (2007) and De Vaan, Ernestus, and Schreuder (2011) provided experimental evidence that readers may store morphologically complex neologisms in memory, like they store more frequent regular complex words (e.g., Alegre and Gordon, 1999; Pinker and Ullman, 2002a,b). This raises questions about the nature of the memory traces left by encounters with morphologically complex neologisms.

The present study investigates the specificity of the traces for morphologically complex neologisms. These traces can be abstract representations which only provide information that is necessary to distinguish the words from the other words in the language. Abstract representations do not provide detailed information about the properties of the tokens; they do not specify, for instance, whether the neologism was attested in written or in spoken language, the situational context (e.g., who uttered the word or what did the room where it was uttered look like) or the linguistic context (e.g., which were the preceding and following words). These abstract representations are stored in the mental lexicon.

The memory traces of neologisms may also contain more specific information of the attested tokens. They may be modality specific and contain information about the context in which the neologism occurred, including the situational and linguistic context. We will refer to these representations as exemplars.

Evidence that memory may contain detailed representations for word tokens comes predominantly from auditory priming experiments (e.g., Bradlow, Nygaard, & Pisoni, 1999; Craik & Kirsner, 1974; Goh, 2005; Goldinger, 1996; Janse, 2008; Mattys & Liss, 2008; McLennan, Luce, & Charles-Luce, 2003; McLennan & Luce, 2005; Palmari, Goldinger, & Pisoni, 1993). In these experiments, words are repeated and the second token of the word is either identical to the first token or it differs in, for instance, speech rate, time compression, the realization of a certain segment, or speaker's voice. Most experiments showed that participants react faster and more accurately to the second token if it is identical to the first one. These results strongly suggest that participants store the acoustic details of the first token of each word and that the recognition process for the second token involves this detailed lexical representation.

The detailed representations are assumed to be part of the mental lexicon in two types of models. In both types, representations of different tokens of the same word (exemplars) are organized in clouds. Exemplar models assume that the mental lexicon only consists of these clouds and does not contain any abstract lexical representations of the words' pronunciations.

Hybrid models, in contrast, combine abstract representations and exemplars. Support for hybrid models comes from priming experiments showing that the indexical properties (which may differ between different tokens of the same word) of the first token of a word (the prime) only affect processing of the second token (the target) when participants have adopted a certain level of processing speed (e.g., McLennan and Luce, 2005). For instance, a match in the indexical properties of the prime and target only speeds up responses to the targets if participants are forced to react after a certain time-interval (150ms after the offset of the spoken stimulus).

Detailed memory traces of word tokens need not be stored in the mental lexicon. They can be stored as episodes in episodic memory (e.g., Goldinger, 1998), which stores autobiographical events as time, place, associated emotions, and other contextual knowledge (e.g., Tulving, 1972; Yim, Dennis, & Sloutsky, 2013). In episodic models of language processing, episodic memory is combined with lexical memory, which contains abstract representations (e.g., Hanique, Aalders, & Ernestus, 2013). In the present study, we do not distinguish between exemplars and episodes, and call them both exemplars. We focus on the question whether memory representations for neologisms are abstract or contain token specific information, including modality and context (exemplars).

As mentioned above, most research on the role of exemplars in language processing has focused on speech comprehension. There has been little research on the role of exemplars in reading. The few studies that addressed this issue showed similar results as studies in the auditory domain: readers recognize the second token of a word faster if it appears in the same font as the first token (e.g., Goldinger, Azuma, Kleider, & Holmes, 2003).

We will investigate the nature of the memory traces for regular morphologically complex neologisms by directly building upon the work done by De Vaan and colleagues (2007, 2011), who examined the formation of memory representations for Dutch regular morphologically complex neologisms. These researchers addressed the question whether such neologisms leave traces in lexical memory with an experimental design in which participants first were presented with the stem of the neologism (stem priming) or the neologism itself (identity priming). Participants were then presented with the neologism (again) later on in the same experiment (De Vaan et al., 2007) or in a second experiment (De Vaan et al., 2011). Their results showed that participants reacted faster to the second token of the neologism in the identity priming condition than in the stem priming condition. This strongly

suggests that participants stored the neologisms that occurred as primes. If they would have only decomposed these neologisms and activated the stems, identity priming should have facilitated the processing of the targets as much as stem priming had done (contrary to fact).

De Vaan and colleagues (2007, 2011) presented all their stimuli (primes and targets) visually and in the same font. As a consequence, we do not know whether the participants in their experiments formed memory traces for the neologisms that contained token specific information, and thus whether these traces are abstract in nature. For instance, we do not know whether these traces were modality specific or modality general. If the prime and target had been presented in different modalities and the prime in the identity priming condition had still facilitated the processing of the target more than the prime in the stem priming condition, the lexical representation could have been classified as modality general. This would be evidence for abstract representations.

Further, the prime and target were always presented in the same situational and linguistic context. Participants saw the two tokens in the same type of experiment (lexical decision or self-paced reading), and if they were incorporated in different parts of an experiment, these were conducted by the same experiment leader in the same experiment room. The experiments could therefore not show whether the memory traces of the neologisms contained contextual information.

Importantly, the experiments conducted by De Vaan and colleagues (2007, 2011) tested various time intervals between the primes and targets. Participants either took part in an experiment where the time interval between prime and target was 1.5 minutes, or they took part in an experiment where it was 12 hours or even a week. Although the advantage of identity priming over stem priming diminished over time, it was still present after a week. This shows that the memory traces are long-lasting. Generally, traces in episodic memory slowly disappear over time (Goldinger, 1996), and exemplar effects typically arise in short experiments with little variation, and disappear if the time interval between prime and target is long or if the experiment is more complex (e.g., Goldinger, 1996; Hanique et al., 2013). These results therefore suggest that the memory traces formed for regular morphologically complex neologisms are abstract representations.

In the present study, we investigated whether the memory traces of regular morphologically complex neologisms are abstract representations or exemplars by conducting three experiments with a design that is similar to the one used by De Vaan and colleagues (2011). The experiments consist of two phases, namely a familiarization phase and a test phase. In the familiarization phase, participants were presented with half of the neologisms that we tested (identity priming) and with the stems of the other half of those neologisms (stem priming). In the test phase, participants were presented with all neologisms. Like De Vaan and colleagues, we

compared identity priming with stem priming, rather than with no priming at all, in order to rule out the possibility that priming from the neologism itself only results from facilitated decomposition as a consequence of base repetition.

We used this design to answer two subquestions. First, we investigated if the advantage of identity priming over stem priming remains if the prime and target are presented in different modalities (auditory vs. visual). The neologisms were spelled with the simple (shallow) phoneme-grapheme mapping rules for Dutch, so that their orthographic representations provide unambiguous information about their pronunciations and vice versa. In Experiment 1, the primes were presented in the visual modality and the targets in the auditory modality. In Experiments 2 and 3, this was reversed (auditory primes and visual targets).

Second, we studied whether the advantage of identity priming over stem priming is also found if the prime and the target occur in different contexts. In all three experiments, the experimental task in the familiarization phase differed from the experimental task in the test phase. In Experiment 1, the primes were presented in sentences in a self-paced reading experiment while the targets appeared in isolation in an auditory lexical decision experiment. In Experiments 2 and 3, participants heard the primes embedded in stories and read the targets in isolation in a visual lexical decision experiment or in stories in a self-paced reading experiment, respectively.

We increased the difference in situational context between the prime and the target by giving participants the impression that the primes and target belonged to different experiments. We asked them to take part in two experiments: our experiment and a colleague's experiment. Moreover, the two phases of the experiment were conducted by different experiment leaders.

In summary, we present three experiments in which the primes were presented in very different conditions than the targets were presented in. If we still observe an advantage of identity priming over stem priming, we can conclude that the formed representations are more likely to be abstract than to be token specific exemplars.

Experiment 1: Visual primes in context and auditory targets in isolation

Method

Participants

Thirty-six undergraduate students (nine male) of Radboud University were paid to take part in this experiment. They had a mean age of 22 years (range 18–32) and were native speakers of Dutch.

Materials

We tested 120 Dutch, morphologically well-formed and derived neologisms, which have been tested in De Vaan et al. (2011) as well. The neologisms end in the suffixes *-baar* (e.g., *betover+baar* 'bewitch' + 'able'), *-heid* (e.g., *antiek+heid* 'ancient' + 'ness'), and *-ing* (e.g., *ontzadel+ing* 'desaddle' + 'ing'; see Appendix A for all neologisms). There were 40 neologisms for each suffix. The suffix *-baar* transforms transitive verbs into adjectives (Booij, 2002), the derivational suffix *-heid* transforms adjectival stems into nouns, and the suffix *-ing* transforms verbs into action nouns. The suffix *-ing* is particularly productive for morphologically complex verbs. Overall, the suffixes *-baar* and *-heid* are more productive than the suffix *-ing* (e.g., Booij, 2002, Van Haeringen, 1971).

Suffixation of a stem with *-heid* does not lead to changes in the pronunciation of the stem. Suffixation with *-ing*, in contrast, leads to resyllabification of stem-final consonants to the following syllable formed by the suffix (e.g., *be-bo-ter* 'to butter' - > *be-bo-te-ring* 'buttering') and also to voicing alternation of the stem-final obstruents (e.g., /bəkłœyf/ *bekluij*'to endure' - > /bəkłœyvin/ *bekluiving* 'enduring') for four stems. Suffixation with *-baar* may lead to regressive voice assimilation (e.g., /vərpɔt/ *verpot* 'to report' - > /vərpɔdbar/ *verpotbaar* 'reportable').

All 120 neologisms have different stems. All stems are real words and are listed in the celex lexical database (Baayen, Piepenbrock, & Gullikers, 1995), whereas celex lists none of the neologisms. At the time of the experiments, each of the neologisms appeared in the Dutch section of the world wide web maximally 219 times, with a mean of 21.1 tokens and a median of 7 tokens. All neologisms are well interpretable, have a maximum of 13 characters (with the diphthong 'ij' counting as two characters), and have a maximum of 11 phonemes.

We conducted a rating study in order to investigate how natural the neologisms are to native speakers of Dutch and whether there are substantial differences among our neologisms. The rating study consisted of three experiments, one for every affix. The ratings for neologisms ending in *-baar* and *-ing* were provided by 71 participants. The rating experiment for neologisms ending in *-heid* was completed by 22 participants. This latter rating experiment was also reported in De Vaan et al. (2007). Participants were asked to rate how often they encounter each word on a scale of one to seven. In addition to the neologisms, the complete rating study also consisted of 183 real words ending in *-baar*, *-heid*, and *-ing*, ranging in celex base frequency from 0 to 86305. Most real words had higher base frequencies than the neologisms (range: 0–435). The average ratings per neologism, split for Suffix, and the average ratings for the real words are presented in Figure 1. The figure shows that most neologisms appeared rather natural to the participants which resulted in a grand mean average of 2.9. This average differs substantially from the grand mean average for the real words (4.7). Neologisms ending in *-heid* generally

received lower ratings (the average rating was 2.5) than the neologisms ending in the other suffixes (with average ratings of 3.1 for *-baar* and 3.1 for *-ing*; *-heid* vs. *-baar*: $t(1987.4) = -29.0, p < 0.0001$; *-heid* vs. *-ing*: $t(2071.1) = -27.2, p < 0.0001$). The neologisms ending in *-ing* and *-baar* also showed a significant but small difference ($t(2703.9) = 2.1, p < 0.05$). This small difference is no longer statistically significant when the effect of Base Frequency on the ratings is taken into account (with a linear model predicting Rating as a function of Base Frequency and Suffix). The relatively low ratings for the neologisms ending in *-heid* may be due to the fact that the base words ended in suffixes themselves (e.g., *pitloos* 'seedless'). We decided not to include the ratings as a predictor in the statistical analyses reported in this study (all three experiments), because the focus of this study is not on the differences among the neologisms.¹

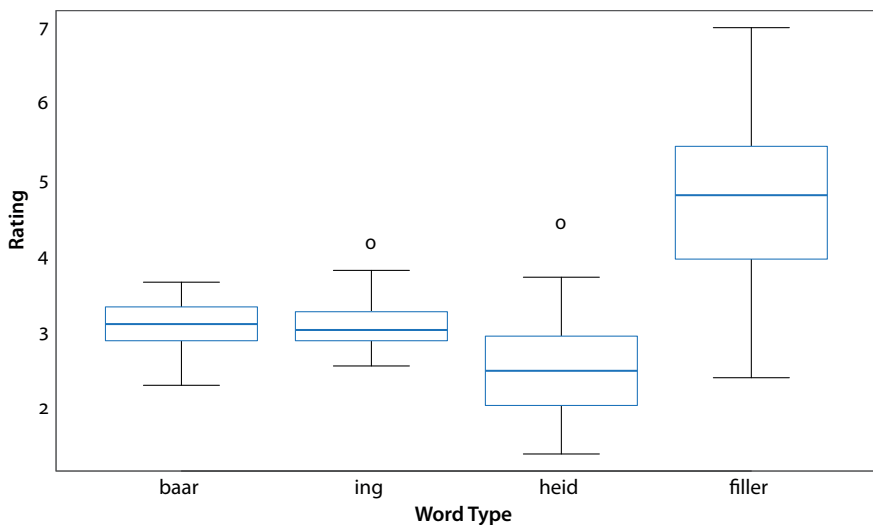


Figure 1. The average ratings per neologism, split for Suffix, and the average ratings for the fillers

For the familiarization phase of the experiment, a self-paced reading task, we created 120 sentences containing the stems of these neologisms and 120 parallel sentences containing the neologisms. An example of a sentence with a stem and its counterpart containing the corresponding neologism is:

1. We investigated whether including Rating as a predictor affects the conclusions that can be drawn from the results. This is not the case.

- (1) *Ik zou mijn zwarte haar graag willen ontkleuren, maar dat blijkt niet zo makkelijk te zijn.* 'I would like to decolorize my black hair but that appears to be not that easy.'
- (2) *Mijn zwarte haar blijkt niet zo makkelijk ontkleurbaar te zijn.* 'My black hair seems not to be decolorable that easily.'

The stem or the neologism was never among the last two words of the sentence. The sentences with the stems consisted of 6 to 22 words (mean: 13.2 words), while the sentences with the neologisms had a length of 5 to 23 words (mean: 12.9 words).

Special care was necessary for the construction of the parallel sentences with neologisms. It was impossible to simply replace the stem by the neologism since the two words belong to different grammatical classes. In the example given above, *ontkleuren* 'decolorize' is a verb while *ontkleurbaar* 'decolorable' is an adjective. Therefore, we had to adjust the grammatical structure of all parallel sentences.

In order to construct sentences that are as natural as possible, we inflected 85 stems that are adjectives (for gender and number; e.g., *summier* was presented as *summiere* 'minimal') or verbs (for person, number, and tense; e.g., the stem *frankeer* 'stamp' was presented as *frankeerde* 'stamped' or *frankeren* 'to stamp') by adding inflectional suffixes. For one irregular verb, we used the irregular past tense (e.g., *verkeek* 'misjudged' with the stem *ver kijk* 'misjudge'). We also inflected two adjectival neologisms ending in the suffix *-baar*, as a consequence of which the neologisms ended in *-bare* (e.g., *verkreukbare* 'wrinklable' and *gedenkbare* 'memorable').

To distract participants' attention from the neologisms, we added 240 filler sentences consisting of 7 to 24 words (mean: 13.4 words). These sentences were semantically and syntactically similar to the experimental sentences. We also added 18 practice sentences to familiarize participants with the experimental task. Both, filler sentences and practice sentences only contained real words.

We created one master list with the sentences of 60 randomly selected target neologisms, the sentences with the stems of the other 60 neologisms, and the 240 filler sentences. We randomized this master list four times without any restrictions. For each of the four resulting lists, we created a counterpart. If a sentence contained a neologism, the counterpart contained the corresponding parallel sentence with the neologism's stem, and vice versa. Each master list was preceded by the practice sentences.

In order to ensure that participants read the sentences carefully for content, one out of every six sentences (target and filler sentences) was followed by a YES/NO question about its contents. For example, sentence (1) was followed by the question

- (3) *Heb ik donker haar?* 'Do I have dark hair?'

Due to an error, only 59 questions were asked instead of 60.

The auditory lexical decision experiment in the test phase contained all 120 target neologisms as well as 120 real filler words and 240 filler pseudowords, all evenly divided over the three suffixes. The real filler words had an average length of 8.3 phonemes in a range of [4–13]. They had a celex lemma frequency in the range of [0–7,196], a mean of 437 and a median of 110. They appeared in the Dutch section of the world wide web with a frequency in a range of [3,580–112,000,000], with a mean of 4,298,613 and a median of 850,500.

We changed to a maximum of three phones of each target neologism and each real filler word to create the 240 filler pseudowords (e.g., *beweperbaar* from *bewap-enbaar* 'armable'), ensuring that the resulting words did not violate Dutch phonotactic constraints. The changes were always in the stem and never in the suffix. In order to test whether the pseudowords primed the target neologisms, we asked three participants, not participating in any of the other experiments reported in this paper, whether they could guess the words from which the pseudowords were derived. The participants were able to do so for only 4% of the pseudowords, which were all derived from real words.

All stimuli were carefully articulated by a female speaker in a noise-attenuated booth and recorded in mono sound (sampling rate: 44.1 kHz). The target neologisms had a mean word duration of 889ms, the real filler words of 895ms, and the pseudowords of 884ms, with standard deviations of 107ms, 174ms, and 125ms, respectively.

We created one master list containing the 120 neologisms, the 120 real filler words and the 240 filler pseudowords. This master list was randomized three times without any restrictions. Each participant was presented with one list. The lists were divided into four parts, which were presented to the participants in a random order.

Procedure

We presented the familiarization phase and test phase to our participants as two different experiments. That is, when approaching participants for our experiment, we gave them the impression that we were looking for participants who wanted to participate in our experiment and, later that day, in a colleague's experiment. Accordingly, the familiarization phase and the test phase were conducted by two different experiment leaders and participants were paid after the familiarization phase as well as after the test phase.

Participants were tested individually in a noise-attenuated experiment booth. In the familiarization phase, participants received self-paced reading instructions on the screen and were asked to read the sentences carefully and at a comfortable reading speed. The course of a trial was as follows. Participants saw a fixation point for 500 ms, indicating that a sentence was about to follow. The instruction 'press the button in the middle of the button box to go to the next sentence' appeared in Dutch on the screen. When participants pressed the button in the middle with

their dominant hand, the full length of the sentence was displayed in dashes, with each dash marking a letter position. By pressing the button in the middle again, they saw the letters of the first word. Words were presented in bold Courier New font 15 pt. With each button press, the following word appeared and the previous word was again replaced by dashes.

Fifty-nine sentences were followed by questions. The word *vraag* 'question' was presented in the center of the screen, followed by the yes/no question about the sentence. Participants were instructed to use their dominant hand to press the yes-button on the button box if the answer to the question was "yes" and otherwise to press the no-button with their non-dominant hand.

Participants first completed the eighteen practice sentences to get used to their task. Four of these sentences were followed by questions. After the practice sentences, participants could ask questions about the task. They then started with the actual experiment.

After this familiarization phase, participants were asked to return to the waiting room to wait for the second experiment. They had to wait approximately five minutes.

In the test phase, they were asked to decide as quickly and accurately as possible whether the item they heard through the headphones was a meaningful Dutch word. The experiment leader made clear that all meaningful words, whether they were real or not, required 'yes' responses, and that only meaningless words required 'no' responses. Participants were instructed to use their dominant hand to press the yes-button on the button box, and their non-dominant hand to press the no-button.

Each trial in the auditory lexical decision experiment began with a fixation mark positioned in the center of a monitor for 1000ms. The stimulus was played 50ms after the fixation mark disappeared. Participants had 2500ms from stimulus onset to decide whether the item was a meaningful Dutch word, before the fixation mark reappeared for the following trial.

Results and discussion

We calculated the percentage of errors for each neologism in the auditory lexical decision experiment. It ranged from 0% to 62%. As we are dealing with neologisms in lexical decision, the high error rates are not surprising. None of the participants had error rates over 15% for the whole experiment and over 40% for the target neologisms. We kept all participants and neologisms in the dataset. The percentages of errors were 15.4% for the neologisms primed by themselves and 16.3% for the neologisms primed by their stems.

We first analyzed whether participants' responses could be predicted by whether they had only seen the neologism's stem before or the complete neologism. We conducted a general linear mixed effect model with this predictor (Prime Type: stem

priming vs. identity priming) in addition to the covariates the logarithmically transformed lemma frequency of the stem of the neologism as listed in *celex* (henceforth Base Frequency), Suffix (*-baar*, *-heid*, and *-ing*), and Trial Number. We included Participant and Word as crossed random variables (e.g., Jaeger, 2008). Prime Type did not show a statistically significant effect, neither as a fixed effect nor as a random slope. This result is as expected because participants were asked to accept all words that they could easily interpret (e.g., Coolen, Van Jaarsveld, & Schreuder, 1991) and therefore their responses were mostly determined by the characteristics of the words such as the suffix, which showed a significant effect (words ending in *-ing* were more often rejected than words ending in *-baar* (22.0% vs. 9.2%; $\beta = 1.4357$, $z = 3.49$, $p < 0.001$) and *-heid* (8.5%; $\beta = 2.3604$, $z = 5.93$, $p < 0.0001$); words ending in *-baar* were also more often rejected than words ending in *-heid* ($\beta = 0.9248$, $z = 2.63$, $p < 0.01$)). We then analyzed participants' response times (measured from word onset) to the neologisms in the test phase, excluding all trials in which participants did not react. If these target neologisms were directly followed by other target neologisms, these directly following target neologisms were also excluded. In addition, we excluded response times longer than the mean response time plus 2.5 times the standard deviation ($> 2088\text{ms}$). We applied a logarithmic transformation to the remaining response times (94.17%, which equals to 4068 data points), to reduce the skewness of the distribution.

We pooled the trials in which participants had responded correctly and incorrectly. We did so because if we only analyzed the correct answers, the dataset would be 13% smaller.² Moreover, the analysis of the responses did not show an effect of our predictor of interest, being Prime Type (as reported above). However, because the two types of responses are given by the dominant versus the non-dominant hand and may also result from different underlying processes, we included the Correctness of the answer as a predictor.

We fitted a mixed-effect model of covariance using a stepwise variable selection procedure, with Participant and Word as crossed random variables (e.g., Baayen, Davidson, & Bates, 2008; Bates & Sarkar, 2005; Faraway, 2006; p-values were generated with the library *lmerTest*, Kuznetsova, Brockhoff, & Christensen, 2016). The dependent variable in our analysis is Response Time (RT), that is logarithmically transformed. Our main predictor is Prime Type. To reduce variance in the data, we added as covariates word specific features that previous studies (e.g., De Vaan et al., 2007, 2011) have shown to affect RTs in lexical decision experiments: the logarithmically transformed duration of the word (henceforth Word Duration), the logarithmically transformed Base Frequency, and Suffix. We also included as covariates

2. Analysis of only the correct answers shows that this data loss is indeed detrimental: this analysis does not show an effect of Prime Type.

the correctness of the participant's response (henceforth Correctness: correct vs. incorrect), whether the stems were inflected (yes or no), Trial Number, and the logarithmically transformed response time to the preceding stimulus (henceforth Previous RT).³ For all categorical variables, we used dummy coding. Finally, we checked for random slopes for all our fixed effects.

Here, as well as in the other experiments in the paper, we tested for simple effects of these variables as well as for significant interactions with our main predictor Prime Type. We tested for interactions because the literature indicates that the type of priming may differ as a function of processing speed (e.g., McLennan and Luce, 2005). Furthermore, we may expect that participants change their strategies (and therefore the effect of Prime Type) during the experiment as a function of the properties of the neologisms.

Inspection of the model's residuals revealed marked non-normality and we therefore removed outliers with standardized residuals outside the interval of $[-2.5, 2.5]$, and refitted the model (e.g., Crawley, 2002). The residuals of this trimmed model were approximately normally distributed, indicating that removal of the overly influential outliers resulted in a model with a better goodness of fit. This model is summarized in Table 1.

Table 1. Results of the Stepwise Multilevel Regression Model ($df = 3983$) fit to the Response Times to the target neologisms in the auditory lexical decision experiment of Experiment 1. Prime Type provides the effect of identity priming relative to stem priming. Suffix provides the effect of the suffix *-baar* or *-heid* relative to the suffix *-ing*. The standard deviation for the random effect of Word was estimated at 0.0611. The standard deviation for the by-subject adjustments to the intercept was 0.5034. In addition, there was a significant random effect for Participant by Previous RT (log-likelihood ratio = 6.86, $p < 0.05$; standard deviation = 0.0688, estimated coefficient for the correlation = -0.969). The residual standard deviation was 0.1594

	$\hat{\beta}$	t-value	p-value
(Intercept)	6.0735	48.83	< 0.0001
Prime Type (identity priming)	-0.0115	-2.27	< 0.05
Previous RT	0.1488	8.63	< 0.0001
Correctness	0.0173	2.01	< 0.05
Suffix <i>-baar</i>	-0.0420	-2.79	< 0.01
Suffix <i>-heid</i>	-0.0930	-6.18	< 0.0001

3. We included stem family size in our original analyses but found that family size was never significant because it was highly correlated with Suffix. Also suffix family size is highly correlated with Suffix and therefore cannot be tested.

We observed simple effects for the control predictors Previous RT, Suffix, and Correctness. As expected, RTs were longer if participants had also reacted slower in the previous trial. Furthermore, participants took longer to react to target neologisms ending in the suffix *-heid* than in the suffix *-baar* ($\hat{\beta} = -0.0510$, $t = -3.40$, $p < 0.001$) and longest to react to target neologisms ending in the suffix *-ing*. The effect for Correctness shows that participants took longer to reject a target neologism as a meaningful word than to accept it, which is in line with previous studies (e.g., Coolen et al., 1991), where native speakers also had difficulties rejecting items that were well-interpretable.

Most important for our study is that we observed a simple effect for Prime Type: target neologisms elicited faster responses in the identity priming condition (mean: 1208 ms; standard deviation: 290 ms) than in the stem priming condition (mean: 1221 ms; standard deviation: 292 ms). These results suggest that memory traces formed for the neologisms during the self-paced reading task facilitated subsequent auditory lexical decision for these same words. A trace of a neologism formed in the visual modality thus facilitated subsequent auditory processing of the word, which suggests that the memory trace is not modality specific and therefore must be abstract. Our finding that identity priming facilitated responses more than stem priming did is an extra indication that the traces are abstract, although there was a change in situational context between the prime and the target. Prime Type did not interact with Inflection, which suggests that the presence of affixes on primes did not decrease the priming effects. This is a final indication that the memory traces formed for the neologisms are abstract.

Experiment 1 strongly suggests that the visual presentations of neologisms facilitate subsequent auditory processing of these same neologisms more than the visual presentations of the neologisms' stems do. Experiment 2 investigated whether the opposite is also true, that is, whether an auditory presentation of neologisms facilitates subsequent visual processing of these words more than an auditory presentation of the stems. If the memory traces formed by neologisms are abstract, we expect that this is the case. In Experiment 2, participants were familiarized with the primes in auditorily presented stories, while in the test phase, the target neologisms were presented to the participants visually and in isolation. We extended the time interval between familiarization phase and test phase to one week, as opposed to a few minutes in Experiment 1. If we would still find that the effect of identity priming is greater than the effect of stem priming, it would be even more likely that the memory traces for the neologisms are abstract in nature.

Experiment 2: Auditory primes in context and visual targets in isolation

Method

Participants

Twenty-four undergraduate students (seven male) of Radboud University were paid to take part in this experiment. They had a mean age of 21 years (range 18–30), were native speakers of Dutch, and none had participated in Experiment 1.

Materials

We tested the same 120 neologisms as in Experiment 1.

For the familiarization phase, a listening comprehension task, we created three sets of four stories with each set containing 60 neologisms and the stems of the remaining 60 neologisms. The neologisms and the stems were equally divided over the stories. For an example of a story, see Appendix B. Each story consisted of 686 to 861 words (with a mean of 784.25 words). For each set of stories, we created a parallel set in which the sentences with neologisms were replaced by sentences with the corresponding stems and vice versa. For example the sentence,

- (4) *Op een **gedenkbaar** dag in juli besteed Willem in alle vroegte zijn paard.* 'On a **rememberable** day in July, early in the morning, Willem mounted his horse.'

had the following counterpart in the parallel story set

- (5) *Willem **gedenkt** nog vaak de dag in juli, toen hij in alle vroegte zijn paard besteed.* 'Willem often **remembers** the day in July, when he mounted his horse, early in the morning.'

This resulted in a total of six story sets. Each participant listened to one set.

As in the familiarization phase of Experiment 1, many of the stems were inflected: 262 of the 360 stems were followed by regular inflectional suffixes (e.g., /skanderə/ *scanderen* 'to chant' consisting of the stem /skander/ *scandeer* 'chant' and the suffix /ə/, indicating the infinitive) and seven were conjugated irregularly (e.g., /vərkek/ *verkeek* 'misjudged', which is the past tense of /vərkeik/ *verkiijk* 'misjudges'). Among the 360 neologisms, 18 ending in *-baar* were inflected (e.g., /ɣəɗɛŋbarə/ *gedenkbaar* 'rememorable' consisting of the neologism /ɣəɗɛŋbar/ *gedenkbaar* 'rememorable' and the suffix /ə/).

All stories presented to a given participant were read by the same speaker. Because we wished to exclude the possibility that our results would only be valid for one single speaker, we chose to have three speakers reading the stories. Each speaker read two story sets. The speakers were three female native speakers of Dutch in the age of 21, 27, and 29 years. Each story was recorded in a noise-attenuated booth

in mono sound in a natural way (sampling rate: 44.1 kHz). Their durations varied between 255s and 324s, with an average of 288s. On average, the speakers produced 4.36 syllables per second.

To force participants to listen carefully to the stories, we formulated five open questions about each story, for instance:

- (6) *In welk seizoen speelt het verhaal zich af?* 'In which season does the story take place?'

The test phase of the experiment was identical to the test phase of Experiment 1, except that we presented the words visually instead of auditorily. The target neologisms had a mean word length of 10.5 characters, the real filler words of 10.1 characters, and the filler pseudowords of 9.6 characters, with standard deviations of 1.05, 2.41, and 1.42 characters, respectively. The spelling of all words, including the filler pseudowords, obeyed the Dutch regular orthographic conventions.

Procedure

The procedure of Experiment 2 was identical to the procedure of Experiment 1 in many respects. As in Experiment 1, participants were approached for two different experiments, the two phases of the experiment were conducted by two different experiment leaders, participants were tested individually in a noise-attenuated booth, and participants received instructions for a lexical decision task in the test phase.

In addition, there were four changes in procedure. Firstly, the familiarization phase of Experiment 2 was an auditory comprehension task instead of a self-paced reading task. Participants were asked to listen to a set of four stories over headphones. We used Windows Media Player to play the sets. At the end of the set, participants were asked to answer the questions on the computer screen using the keyboard. We first referred to a story. For example:

- (7) *De volgende vragen gaan over het verhaal waarin Willem zijn zoektocht naar een uniek boek voltooit.* 'The following questions are about the story in which Willem finishes his search for a unique book'

Participants then pressed the ENTER key to see the first question. The question remained on the screen while the participants typed their answers. They then pressed the ENTER key for the next question. This procedure was repeated until they had answered all five questions about the given story. They then saw a white screen for 1000 ms, followed by a screen referring to the next story, after which they answered the five corresponding questions. This procedure of questioning was repeated for all four stories. After the last question, a word of thanks appeared on the screen. All text appeared in black, centered on a white screen, in bold Courier New font 18 pt.

Secondly, the procedure of Experiment 2 differed from that of Experiment 1 in that the time interval between the familiarization phase and test phase was extended from a few minutes to one week, which decreases the likelihood that the responses were only based on temporary memory traces.

Thirdly, the test phase consisted of a visual instead of an auditory lexical decision experiment. Each trial in this phase began with a fixation mark positioned at the center of the computer screen for 500ms. After 50ms of blank screen, the stimulus appeared, also centered on the screen. Stimuli were presented in black lowercase, in bold Courier New font 18 pt letters, on a white background and remained on the screen for 1750ms, which was the time interval in which the participants had to react.

Fourthly, we only offered participants financial compensation after the test phase to encourage them to participate in that phase as well. We presented that phase as a colleague's experiment for which participants were needed urgently.

Results and discussion

Participants widely varied in their performance in the familiarization phase: whereas some participants answered 90% of the questions correctly, others only answered 20% of the questions correctly. This variation possibly reflects differences among the participants in working memory because the questions only appeared after the participants had heard all the stories. Although some participants made many errors, we decided not to exclude participants on the basis of their performance in the familiarization phase. Instead, we included participants' performance in the familiarization phase as predictor for their performance in the test phase (see below).

In the test phase, the average percentage of errors for a neologism ranged from 0% to 85%, with twelve neologisms having error rates above 50%. None of the participants had error rates of over 30% for the entire experiment (target neologisms, real filler words and pseudowords). We decided not to exclude any participants or words on the basis of these error data. The percentages of errors were 18.7% for the neologisms primed by themselves and 18.3% for the neologisms primed by their stems.

We first analyzed whether participants' responses could be predicted by whether they had only heard the neologism's stem before or the complete neologism. As for Experiment 1, we conducted a general linear mixed effect model with this predictor (Prime Type: stem priming vs. identity priming) in addition to the covariates the logarithmically transformed lemma frequency of the stem of the neologism as listed in *celex* (Base Frequency), Suffix (-*baar*, -*heid*, and -*ing*), Rating, and Trial Number. We included Participant and Word as crossed random variables

(e.g., Jaeger, 2008). Prime Type did not show a significant effect, neither as a fixed effect nor as a random slope. Only Trial Number, Rating, and Suffix emerged as statistically significant effects. Participants rejected more neologisms when they were further in the experiment ($\beta = 0.2064$, $z = 3.75$, $p < 0.001$) and they rejected more neologisms ending in *-ing* than neologisms ending in *-baar* (33.1% vs. 15.7%; $\beta = 1.2472$, $z = 4.93$, $p < 0.0001$) and *-heid* (23.9%; $\beta = 1.3667$, $z = 4.97$, $p < 0.0001$). They rejected fewer neologisms when the rating scores were higher ($\beta = -0.6821$, $z = -5.48$, $p < 0.0001$).

For the analysis of the RTs in the test phase, we only excluded trials to which the participants did not react and the directly following neologisms. We also excluded RTs longer than 1643 ms, which is the mean response latency plus 2.5 times the standard deviation. This led to a dataset of 2695 response times (93.58%). The data fitting procedure was the same as in Experiment 1. We tested as fixed predictors Prime Type, Correctness of the answer, Previous RT, Suffix, Length (word length in number of characters), Base Frequency, Trial Number, and Familiarization Performance (the number of correct answers to the questions in the familiarization phase provided by the given participant). We applied a logarithmic transformation to the RTs, Previous RTs, and Base Frequency. We checked for random slopes for all our fixed effects. Finally, we removed outliers with standardized residuals outside the interval of $[-2.5, 2.5]$, and refitted the model. The final model is shown in Table 2.

Table 2. Results of the Stepwise Multilevel Regression Model ($df = 2657$) fit to the lexical decision times of Experiment 2. Suffix provides the effect of the suffix *-baar* or *-heid* relative to the suffix *-ing*. The standard deviation for the random effect of Word was estimated at 0.0688, and that for the Participant random effect at 0.1757. In addition, there was a significant random effect for Participant by Trial Number (log-likelihood ratio = 93.23, $p < 0.01$, standard deviation = 0.0002, estimated coefficient for the correlation = -0.190). The residual standard deviation was 0.2150

	$\hat{\beta}$	t-value	p-value
(Intercept)	5.909	31.54	< 0.0001
Prime Type (identity priming)	-0.3064	-1.79	< 0.1
Trial Number	-0.00111	-2.57	< 0.05
Length	-0.01087	-0.83	> 0.1
Previous RT	0.1570	8.69	< 0.0001
Suffix <i>-baar</i>	-0.0632	-3.25	< 0.01
Suffix <i>-heid</i>	-0.0772	4.17	< 0.0001
Prime Type (identity) by Trial Number	0.0015	2.38	< 0.05
Prime Type (identity) by Length	0.0272	1.67	< 0.1
Trial Number by Length	0.0001	2.42	< 0.05
Prime Type (identity) by Trial Number by Length	-0.0001	-2.31	< 0.05

As in Experiment 1, we observed simple effects for our control predictors Previous RT and Suffix. Response times were longer if the Previous RTs were longer and if the neologism ended in the suffix *-ing* (953 ms) instead of *-baar* (908 ms) or *-heid* (893 ms).

More important for our research question, we observed a three-way interaction of Prime Type, Length, and Trial Number. In order to interpret this interaction, we split the data in half, based on Trial Number (Trials in the first half of the experiment: Trial Number < 241; Trials in the second half of the experiment: Trial Number > 240). Analyses of the first half of the experiment showed a simple effect for Prime Type, regardless of the neologism's length ($\hat{\beta} = -0.0233$, $t = -1.96$, $p = 0.05$, mean RT and standard deviation for neologisms in the identity priming condition: 926 ms and 274 ms; for the neologisms in the stem priming condition: 943 ms and 289 ms). This suggests that in the first half of the experiment, participants reacted faster to a neologism if they had heard that neologism before than if they had only heard the neologism's stem.

For the second half of the experiment, we observed an interaction of Prime Type and Length ($\hat{\beta} = -0.0271$, $t = -2.408$, $p < 0.05$, mean RT and standard deviation for neologisms in the identity priming condition: 897 ms and 263 ms; for neologisms in the stem priming condition: 891 ms and 268 ms), which is shown in Figure 2. We split the trials based on Length (Short Words (SW): Length < 11 characters, 691 response times; Long Words (LW): Length > 10 characters, 662 response times) and investigated if Short Words and Long Words showed significant effects of Prime Type. The effect of Prime Type approached significance for the Short Words but not for the Long Words (SW: $\hat{\beta} = 0.0313$, $t = 1.87$, $p = 0.06$, mean RT and standard deviation for neologisms in the identity priming condition: 897 ms and 263 ms; for the neologisms in the stem priming condition: 891 ms and 268 ms; LW: $\hat{\beta} = -0.0224$, $t = -1.35$, $p > 0.1$, mean RT and standard deviation for neologisms in the identity priming condition: 909 ms and 286 ms; for the neologisms in the stem priming condition: 908 ms and 275 ms). Surprisingly, for Short Words, the effect goes in the opposite direction than for the neologisms in the first half of the experiment: participants seem to have reacted faster in the stem priming condition than in the identity priming condition.

In summary, we observed that identity priming (mean: 926 ms; standard deviation: 274 ms) was larger than stem priming (mean: 943 ms; standard deviation: 289 ms) in the first half of Experiment 2. These results strongly suggest that memory traces formed for the neologisms during the auditory comprehension task facilitated subsequent visual lexical decision for these same words. That is, a trace of a neologism formed in the auditory modality facilitated subsequent visual processing of the word, which strongly suggests that the trace is not modality specific and therefore must be abstract.

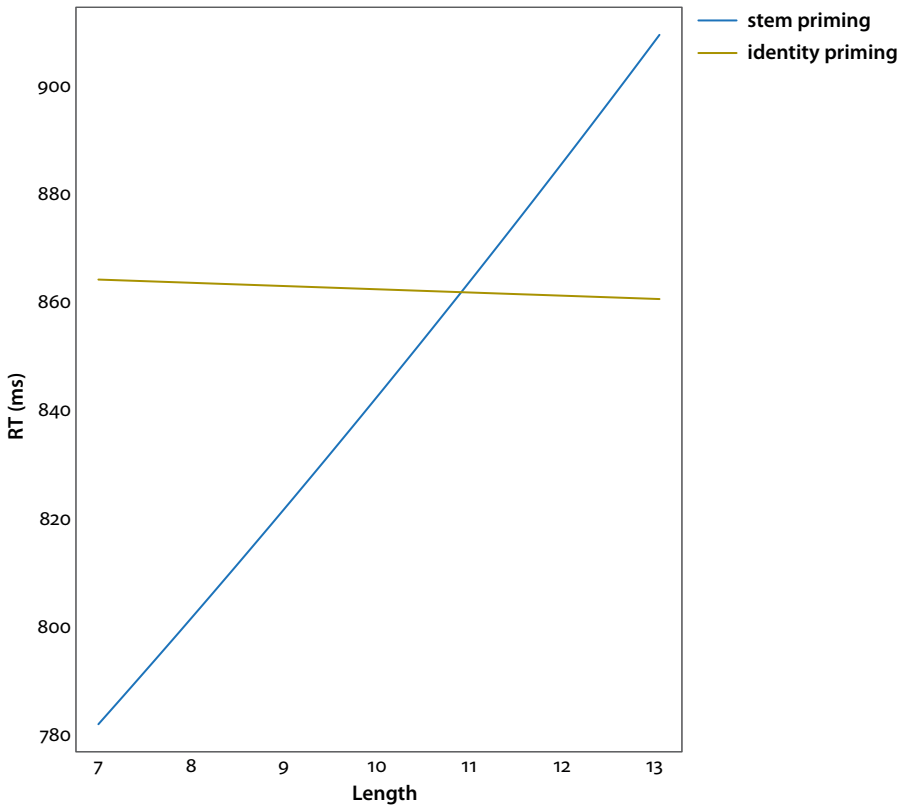


Figure 2. The effect of Word Length (in amount of characters) on Reaction Times (in ms) in the second half of Experiment 2, split by Prime Type

To obtain additional evidence for abstract representations, we varied the situational context between the familiarization phase with the primes and the test phase with the targets. Moreover, we extended the time interval between these phases from approximately a few minutes in Experiment 1 to one week in this experiment. Participants nevertheless reacted faster on the repeated neologisms than on the newly presented neologisms.

The results of the second half of Experiment 2 suggest that in the second part of the experiment participants reacted faster in the stem priming condition (mean: 891 ms; standard deviation: 268 ms) than in the identity priming condition (mean: 897 ms; standard deviation: 263 ms) for short words. These effects may reflect task specific behavior rather than natural everyday language processing. We postpone the interpretation of these results to the General Discussion.

In Experiment 3, we address the question whether a trace for a neologism formed during listening to sentences also facilitates the reading of that same

neologism if it occurs in a different linguistic context. A different linguistic context may hinder priming if the formed traces for the neologisms are context specific. On the other hand, if the formed traces are abstract, we do not expect priming to be affected by the difference in sentential context.

Experiment 3: Auditory primes and visual targets, both in context

Method

Participants

Twenty-four undergraduate students (seven male) studying at Radboud University were paid to take part in this experiment. They had a mean age of 21 years (range 18–24), were native speakers of Dutch, and none had participated in Experiment 1 or 2.

Materials

We tested the same 120 neologisms as in Experiments 1 and 2. The familiarization phase of Experiment 3 was identical to the familiarization phase of Experiment 2, with exactly the same materials.

For the test phase, a self-paced reading experiment, we created 24 short texts. Each text consisted of 8 to 16 sentences (mean: 12 sentences), 179 to 239 words (mean: 199 words), and contained five neologisms (see Appendix C for an example). The number of inflected neologisms was kept to a minimum (only six neologisms ending in *-baar* were inflected). The neologisms were always followed by at least two words in the same sentence. The reading times of these two words can show spill-over effects from the neologisms (e.g., for spill-over effects see Just, Carpenter, & Woolley, 1982; Ernestus & Mak, 2005). In addition, we created two practice texts. These texts did not contain any neologisms, and further followed the same structure as the 24 texts constructed for the test phase. Each text was followed by a question about its content, which stimulated participants to read the texts carefully.

Importantly, the semantic and syntactic context of a word was not the same in the familiarization phase as in the test phase. We illustrate this for the neologism *verrijding*, which occurred in context (8) in the familiarization phase and in context (9) in the test phase. As a consequence, the neologism could elicit slightly different interpretations in the two phases.

- (8) *De beuheid overviel hem daarnaast over de lengte van de routes die hij en zijn piano telkens moesten afleggen. Een **verrijding** maakte hij niet zo vaak mee, daarentegen was er wel vaak sprake van verkuiling van de oprijlaan.* In addition,

tiredness overcame him to the length of the routes that he and his piano had to travel. A **deviation** did not often happen, however there often were many holes in the driveway.’

- (9) *Hij voelde zich tijdens het rijden niet goed worden. Die onwelheid zorgde voor een **verrijding** op het hobbelige wegdek, waarna de man de macht over het stuur verloor.* ‘He felt unwell while driving. This unwellness resulted in **making movements** on the bumpy road after which the man lost control of the vehicle.’

We created one master list containing all 24 texts. This master list was randomized three times. Each participant was presented with one list. The lists were divided into four parts, which were presented to the participants in a random order.

Procedure

The procedure of Experiment 3 was identical to the procedure of Experiment 2 in many respects. As in Experiment 2, participants were approached for two different experiments, the two phases of the experiment were conducted by two different experiment leaders, participants were tested individually in a noise-attenuated booth, the time interval between the familiarization phase and the test phase was one week, and we only offered participants financial compensation after the test phase to encourage them to participate in that phase as well. We presented that phase as a colleague’s experiment for which participants were needed urgently.

In addition, there was one change in procedure. The test phase of Experiment 3 was a self-paced reading task. Participants received self-paced reading instructions, which were identical to the ones for the familiarization phase of Experiment 1.

Each text in the test phase was presented in three, more or less equally long parts (caused by the limitations posed by the E-prime software), with every part ending at the end of a sentence. To start a trial, participants pressed the button in the middle of a button box. The first part of the text, aligned to the left side of the screen, was displayed with dashes marking letter positions. By pushing the button in the middle again, participants could see the first word of the sentence in bold Courier New font 15pt. When they pressed the button in the middle again, the second word appeared and the first word was again replaced by dashes. This procedure was repeated until the participants had read the first part of the text. This first part of the text was then followed by the second and the third part, each preceded by a blank screen. Participants read these second and third parts in the way they had read the first part. When they had finished reading the third part, a yes/no question about the entire text appeared centered on the screen. Participants indicated their response by means of a button press (the ‘yes’-button with their dominant hand, the ‘no’-button with their other hand). To get used to the task, participants first completed two practice texts. They then started with the 24 texts of the test phase.

Results and discussion

Since there are often spill-over effects to the following words (e.g., Just et al., 1982; Ernestus & Mak, 2005) in self-paced reading experiments, we analyzed the reading times of the neologisms themselves and of the two following words.

We excluded one target neologism from analysis, namely *arceerbaar* 'shadable', because we accidentally made a new neologism out of the neologism we meant to incorporate (*arceerbaarheid* 'shadableness'). We excluded one participant because her standard deviation of the reading times was larger than her mean reading time. We also excluded two participants with error rates over 30% for the questions, which means eight or more errors in 26 questions (mean error rate was 15%; four errors in 26 questions). These participants have probably not seriously read the stories. We further excluded data points with reading times shorter than 200ms, as well as data points that were preceded by reading times shorter than 200ms, since these extremely short reading times may indicate that participants did not read these words carefully (Jegerski, 2014). Finally, we excluded data points with reading times exceeding 1105ms and data points that were preceded by reading times exceeding 1105ms, which is the mean response latency plus 2.5 times the standard deviation. The remaining dataset consisted of 6756 data points (which forms 78.2% of the complete original data set). The resulting RTs do not point to a processing difference between neologisms that are primed by themselves or by their stems (407ms vs. 404ms). We nevertheless decided to analyze the data because the effect may be obscured by the effects of other variables (as was the case in De Vaan et al., 2007).

The data fitting procedure was the same as in Experiments 1 and 2. We included three random intercepts in the mixed effect models of covariance: Participant, Word, and Neologism. The random variable Word refers to the word that is analyzed, which is the identity of the neologism, the identity of the first following word or of the second following word. The dependent variable in our analysis is Response Time (RT), which was logarithmically transformed. Our main predictors were Prime Type (stem priming vs. identity priming) and Position (with the levels: "the neologism", "the first following word" and "the second following word"). Our control predictors were again Length in characters, Suffix, the logarithmically transformed Base Frequency, and the logarithmically transformed Previous RT. We checked for random slopes for all our fixed effects. Finally, we removed outliers with standardized residuals outside the interval of $[-2.5, 2.5]$, and refitted the model.

We observed a four-way interaction between Prime Type, Position, Previous RT, and Length ($F(2,6402.7) = 6.69, p < 0.01$). To simplify the interpretation of this interaction, we split the data into three sets representing the levels of Position, and analyzed these subdatasets separately. The datasets for neologisms, first following words, and second following words contained 2301, 2226, and 2229 datapoints,

respectively (the mean RTs for stem priming were: for the reading of the neologism itself 362ms, for the reading of the first following word 440ms, and for the reading of the second following word 411ms; the mean RTs for identity priming were: for the reading of the neologism itself 367ms, for the reading of the first following word 443ms, and for the reading of the second following word 411ms). Except for Position, we incorporated the predictors and the interactions that were significant in the overall analysis and applied the same fitting procedure as for the overall dataset. For the analysis of the reading times of the neologisms, the random variables Word and Neologism are completely identical (i.e., the *Word* that the participant read was the *Neologism*). We left out the random variable Neologism. The final models for the three subdatasets are summarized in Tables 3, 4, and 5, respectively.

Table 3. Results of the Stepwise Multilevel Regression Model ($df = 2220$) fit to the Reaction Times for the neologisms in the self-paced reading part of Experiment 3. Prime Type provides the effect of identity priming relative to stem priming. Suffix provides the effect of the suffix *-heid* or *-ing* relative to the suffix *-baar*. The standard deviation for the random effect of Word was estimated at 0.0485 and that for the Participant random effect at 0.8405. In addition, there was a significant random effect for Participant by Previous RT (log-likelihood ratio = 22.22, $p < 0.01$, standard deviation = 0.1431, estimated coefficient for the correlation = -0.99). The residual standard deviation was 0.1765

	$\hat{\beta}$	t-value	p-value
(Intercept)	3.3779	15.42	< 0.0001
Prime Type (identity priming)	0.3590	2.41	< 0.05
Previous RT	0.4205	11.28	< 0.0001
Suffix <i>-heid</i>	-0.0269	-1.87	0.06
Suffix <i>-ing</i>	-0.0338	-2.36	< 0.05
Prime Type (identity priming) by Previous RT	-0.0596	-2.35	< 0.05

For the neologisms, we observed a simple effect for the control variable Suffix. In contrast to what we found in Experiments 1 and 2, reading times were slightly longer for neologisms ending in *-baar* than for neologisms ending in *-heid* (marginally) and *-ing* (mean RTs were 361ms, 360ms, and 358ms, respectively).

We also found an interaction between Prime Type and Previous RT, as illustrated in Figure 3, which is important for our main question. To interpret this interaction, we split the data in half, based on Previous RT ($\leq 340ms$ and $> 340ms$) and analyzed these datasets separately. These analyses showed that RTs were shorter in the identity priming condition than in the stem priming condition, but only when participants had read the preceding word relatively slowly ($\hat{\beta} = -0.1312$, $t = -2.77$, $p < 0.01$). When participants had read the preceding word relatively fast, the identity priming condition and the stem priming condition did not show a statistically significant difference.

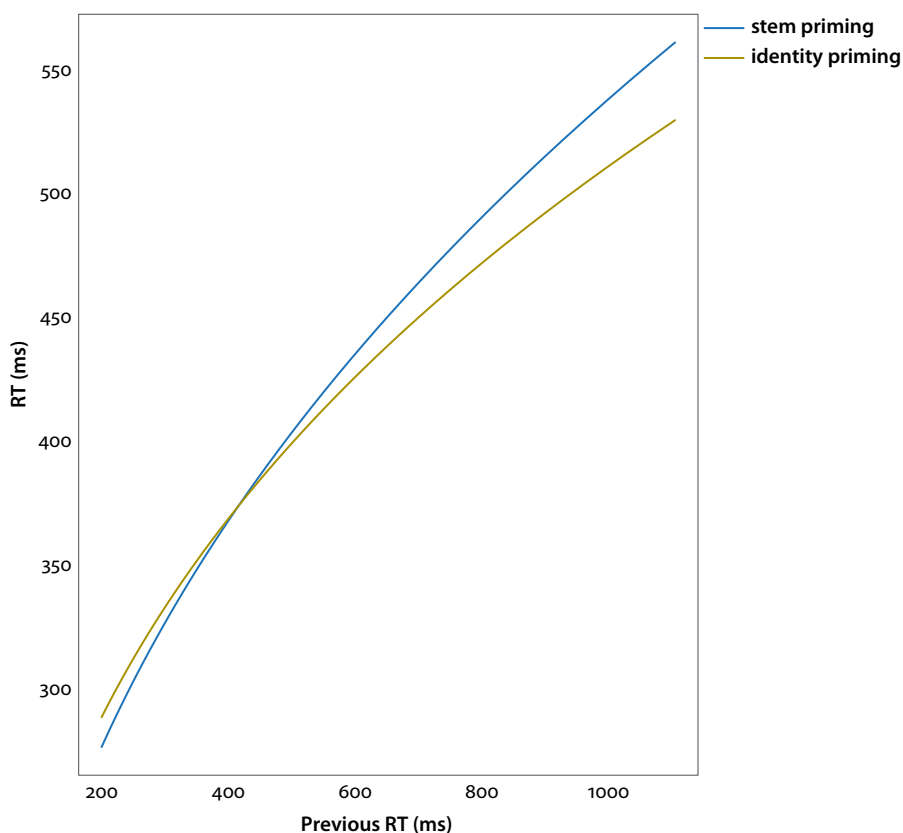


Figure 3. The effect of the RTs for the previous trial (in ms) on the RTs for the neologisms (in ms) in the self-paced reading task of Experiment 3, split by Prime Type

The analysis of the RTs to the word immediately following the neologism showed an interaction between Prime Type, Previous RT, and Length (see Table 4). We split the dataset in two equal parts based on Previous RT ($\leq 334ms$ and $> 334ms$) in order to interpret this interaction. For RTs preceded by short RTs on the neologisms, we observed a simple effect for Prime Type ($\hat{\beta} = -1.6936$, $t = -2.57$, $p < 0.05$) and an interaction between Prime Type and Previous RT ($\hat{\beta} = 0.3012$, $t = 2.58$, $p < 0.05$): Reading times increase relatively more in the identity priming condition than in the stem priming condition with increasing previous RTs, as shown in Figure 4, left panel. As a consequence, identity priming was only larger than stem priming after short previous RTs. For the other half of the data, the RTs preceded by long RTs on the neologisms, we observed an interaction between Prime Type and Length ($\hat{\beta} = -0.0180$, $t = -2.26$, $p < 0.05$): as Length increases, the benefit of identity priming compared to stem priming increases as well, as shown in Figure 4, right panel.

Table 4. Results of the Stepwise Multilevel Regression Model ($df = 2163$) fit to the Reaction Times for the first following words in the self-paced reading part of Experiment 3. Prime Type provides the effect of identity priming relative to stem priming. The standard deviation for the random effect of Neologism was estimated at 0.0656 and that for the Participant random effect at 1.7317. In addition, there was a significant random effect for Participant by Previous RT (log-likelihood ratio = 71.49, $p < 0.01$, standard deviation = 0.2767, estimated coefficient for the correlation = -0.99). The residual standard deviation was 0.2543

	$\hat{\beta}$	t-value	p-value
(Intercept)	3.9081	7.38	< 0.0001
Prime Type (identity priming)	-1.3422	-2.77	< 0.01
Length	-0.1818	-2.13	< 0.05
Previous RT	0.3569	4.07	< 0.001
Prime Type (identity) by Previous RT	0.2350	2.83	< 0.01
Prime Type (identity) by Length	0.2795	2.55	< 0.05
Previous RT by Length	0.0316	2.15	< 0.05
Prime Type (identity) by Previous RT by Length	-0.0489	-2.59	< 0.01

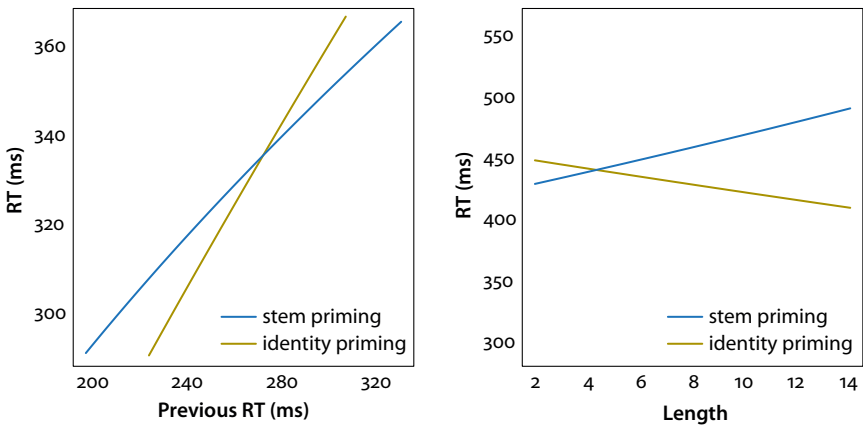


Figure 4. The effect of Prime Type in the self-paced reading task of Experiment 3. Left panel: The effect of half of the RTs on the neologisms (in ms), the shorter half, on the words following the neologisms, split by Prime Type. Right panel: The effect of the number of characters on the RTs (in ms) for the words following the neologisms that were read relatively slow, split by Prime Type

The analysis of the second following word (see Table 5) showed simple effects for the control variables Suffix and Previous RT. Reading times for the second following words after neologisms ending in *-baar* were longer than for the ones after neologisms ending in *-heid* and *-ing* (428ms, 409ms, and 395ms, respectively). Reading times increased along with increasing Previous RTs.

Table 5. Results of the Stepwise Multilevel Regression Model ($df = 2169$) fit to the Reaction Times for the second following words in the self-paced reading part of Experiment 3. Prime Type provides the effect of identity priming relative to stem priming. The standard deviation for the random effect of Neologism was estimated at 0.0597, that for the Word random effect at 0.0228 and that for the Participant random effect at 1.1983. In addition, there was a significant random effect for Participant by Previous RT (log-likelihood ratio = 54.25, $p < 0.01$, standard deviation = 0.1893, estimated coefficient for the correlation = -1.00). The residual standard deviation was 0.2301

	$\hat{\beta}$	t-value	p-value
(Intercept)	5.076	18.07	< 0.0001
Prime Type (identity priming)	0.0717	3.51	< 0.001
Previous RT	0.1481	3.33	< 0.01
Length	0.0127	2.92	< 0.01
Suffix <i>-heid</i>	-0.0495	-2.50	< 0.05
Suffix <i>-ing</i>	-0.0724	-3.63	< 0.01
Prime Type (identity priming) by Length	-0.0181	-4.08	< 0.0001

Importantly, we also observed an interaction between Prime Type and Length, which is illustrated in Figure 5. To investigate how to interpret this interaction, we split the data in half, based on Length (≤ 3 characters and > 3 characters) and analyzed these datasets separately. The analysis of the RTs for the relatively long words showed a simple effect for Prime Type ($\hat{\beta} = 0.1010$, $t = 2.15$, $p < 0.05$) and an interaction between Prime Type and Length ($\hat{\beta} = -0.0216$, $t = -2.95$, $p < 0.01$): RTs for the long second following words were shorter in the identity priming condition than in the stem priming condition, but only when these words were relatively long. The analysis of the RTs for short words showed no effect of Prime Type.

In summary, Experiment 3 was constructed to find out whether a neologism also leaves a detectable trace in the mental lexicon when the prime and target are presented in different modalities as well as in different situational and linguistic contexts. The reading times for the neologism itself, the first following word, and the second following word showed effects of Prime Type.

For the neologisms and the first following words, the effect of Prime Type was modulated by the reading time for the previous word. Identity priming facilitated the reading of the neologism but only if the preceding word was read relatively slow. The word following the neologism showed this facilitatory effect of identity priming only if the preceding word was read fast. For the following two words, the effect of Prime Type was (also) modulated by word Length. Participants read the two words following neologisms faster in the identity priming condition than in the stem priming condition but only if these following words were long (and for the directly following word, this only holds for relatively slow readers).

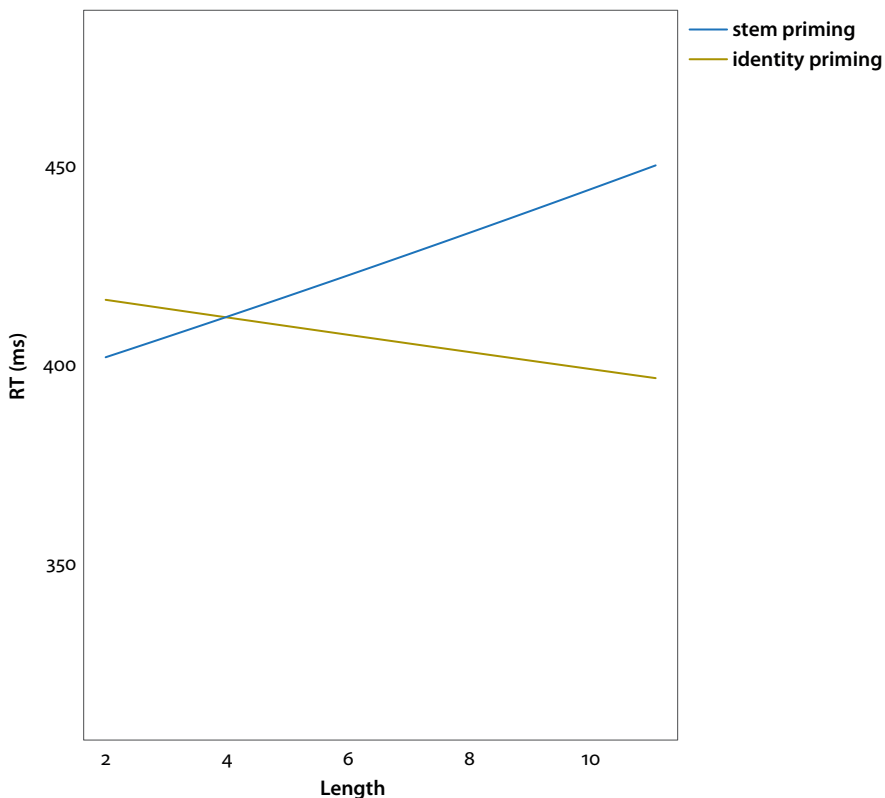


Figure 5. The effect of Word Length (in amount of characters) of the second following word on its reading times (in ms), split by Prime Type, in the self-paced reading task of Experiment 3

General discussion

This article addressed the question whether regular morphologically complex neologisms form abstract representations or exemplars in memory. We investigated this question by examining the processing of Dutch neologisms consisting of real stems and real productive suffixes. We first presented participants with either the stem of a neologism (stem priming) or with the complete neologism (identity priming) and then presented the target neologism. We included the stem priming condition as a baseline in order to investigate whether priming in the identity priming condition resulted from the full neologism or just its stem. If it was just the stem of the neologism that induced priming, the stem and the neologism should show

similar priming effects. If, in contrast, participants stored the full neologisms, the neologisms should show more priming than just their stems.

We addressed two subquestions. First, does identity priming facilitate the processing of neologisms more than stem priming does, also if the prime and the target are presented in different modalities (visual vs. auditory and vice versa)? Second, is identity priming also more beneficial than stem priming if the prime is presented in one experimental task and the target in another? If change in modality and context eliminate the benefit of identity priming over stem priming, the memory traces for neologisms are probably token specific (exemplars). If not, these traces are more likely to be abstract.

We conducted three experiments. All primes (stems and neologisms) were presented in a familiarization phase and the targets in a later phase, namely the test phase. Furthermore, participants were made to believe that these phases were independent experiments. The task in the test phase was different from the one in the familiarization phase.

In Experiment 1, the primes were presented in sentences in a self-paced reading task and the targets in isolation in an auditory lexical decision task. Target neologisms elicited faster responses in the identity priming condition than in the stem priming condition. A memory trace formed for a neologism during a self-paced reading task can thus facilitate subsequent processing of that neologism in an auditory lexical decision task. These results indicate that the formed traces are modality general and task independent, which strongly suggests that they are abstract lexical representations.

In Experiment 2, the primes were presented in sentences in an auditory story comprehension task and the targets in isolation in a visual lexical decision task. The time interval between both parts of the experiment was one week as opposed to a few minutes in Experiment 1. In the first half of the lexical decision experiment, participants reacted faster to a neologism if they had heard that neologism before than if the neologism was entirely new. These results indicate that memory traces formed for neologisms during a listening task can facilitate subsequent visual lexical decisions, even after one week. Priming effects after a week, induced by neologisms, are thus not restricted to experiments in which the primes and targets are presented in the same modality. This outcome enforces the results of Experiment 1, showing that memory traces for neologisms are modality general and task independent, which strongly suggests that they are abstract lexical representations.

Noteworthy, the second half of Experiment 2 showed a reversed pattern of results. Participants reacted faster in the stem priming condition than in the identity priming condition but only if the neologisms were short. A possible explanation for this result is that participants caught on to the morphological structures of the words after a while. These all consisted of a (pseudo)stem and a real suffix.

Participants may have adopted a strategy to strip the suffixes from the stimuli and then base their lexical decisions just on the stems. Suffix stripping may be easier for shorter words because they are generally read at a glance. This would explain why stem priming had an advantage over identity priming in the second half of the experiment and especially for short words. In the test phase of Experiment 1, we did not observe a similar interaction. This, being an auditory lexical decision experiment, is less conceivable among others because of resyllabification, which makes affix stripping more difficult.

In Experiment 3, the primes were presented in the same auditory story comprehension task as in Experiment 2, and the targets were presented in different stories in a self-paced reading task. The time interval between both parts of the experiment was again one week. The reading times for the neologism itself and for the two following words all showed effects of Prime Type, which by itself suggests that neologisms leave different traces than their stems do in the comprehension system. Neologisms were read faster if they were primed by themselves than by their stems, but only if the participants read the directly preceding words relatively slowly (and thus probably had a slow local reading speed). The reading times for the following word showed the same facilitatory effect of identity priming, but only if the participants had read the preceding neologism relatively fast. These two facilitatory effects of identity priming probably represent one single effect, which arises some time after the presentation of the neologism. When slow readers (i.e., participants who were slow for the given sequence of trials) are still reading the neologism, fast readers (i.e., participants who were fast for the given sequence of trials) have already turned to the following words.

In addition, we found interactions between the effect of Prime Type and the number of the word's characters for both the reading of the directly following word and the second word: identity priming facilitated the reading of that word but only if it was long. We assume that these interactions are also related to the speed in which priming affects processing, since longer words take more time to read.

In Experiment 3, the effect of identity priming does not occur immediately when readers see the neologism. This does not come as a surprise since many effects arise rather late in this experimental paradigm, for instance, effects from biases for the grammatical subjects of subclauses following implicit causality verbs (Koornneef & Van Berkum, 2006), effects from intra- and interparadigmatic effects on verb forms (Ernestus & Mak, 2005), and effects from syntactic ambiguity (Swets, Desmet, Clifton, & Ferreira, 2008).

The results of Experiment 3 confirm the results of Experiments 1 and 2. The three experiments together clearly show that a neologism encountered in one modality facilitates the processing of the second token of that word in another modality. This is even true when the two tokens occur in very different experimental

tasks (Experiments 1, 2, and 3), when participants believe that the two tasks form independent experiments (Experiments 1, 2, and 3), and when the time interval between the prime and the target is one week (Experiments 2 and 3). Experiment 3 contributes to the findings from Experiments 1 and 2 by showing that the priming effect also arises when the two tokens of the word occur in very different sentences and may therefore have slightly different meanings. These data strongly suggest that a neologism leaves a memory trace that is abstract. In addition, these traces may contain token specific information. Further research has to investigate whether this is the case.

The results also show that the difference between stem priming and identity priming is small (the maximal difference observed in the study is 17ms). Furthermore, the presence and size of the difference varies with experimental task and for some tasks also with trial number. We interpret this variation as a task effect. Important is that all experiments (or at least the first parts of all experiments) showed the same pattern: neologisms are processed more quickly when primed with themselves than when only primed with their stems.

Due to the consistent phoneme-to-grapheme and grapheme-to-phoneme mappings in Dutch, there was hardly any ambiguity about the orthographic representations of the auditorily presented neologisms and about the pronunciation of the orthographically presented neologisms. Research by Bakker, Takashima, Van Hell, Janzen, & McQueen (2014) showed priming of Dutch morphologically simple neologisms between different modalities. They used lexical competition between these neologisms and their real phonological/ orthographic neighbors to test lexical integration. Auditorily acquired neologisms contributed to the lexical competition in the written modality after 24 hours. On the other hand, visually acquired neologisms required extra training and a consolidation period of a week before they contributed to the lexical competition during auditory word comprehension. When tested in the same modality, both auditorily and visually acquired neologisms showed competition effects after 24 hours. This study thus also shows that words acquired in one modality can affect the processing of that word in another modality.

Bakker et al. (2014) found that the conversion from visual representations to auditory representations took more effort than the conversion from auditory representations to visual representations (participants needed more training tokens and a longer consolidation period). Yet, we observed cross-modal priming almost immediately and after only one training token. A possible explanation for this difference in results is that our neologisms consisted of novel combinations of real morphemes, which already have their own representations in the mental lexicon. Participants only have to create a new link between both real morphemes (stem and suffix). In contrast, Bakker et al. (2014) investigated the acquisition of completely novel morphemes.

The question arises how priming between different modalities may take place. One possibility is that the presentation of a neologism in any modality leads to a simultaneous creation of both a visual and an auditory lexical representation. When presenting a second token of a neologism, either the visual or the auditory representation would be activated in lexical memory, depending on the modality of the second token. This activation would lead to a faster recognition of the neologism. The assumption of specific representations, however, cannot account for our findings that a change in linguistic context or a slight change in meaning also do not obstruct the priming effect. We therefore prefer the interpretation that the neologisms form abstract representations, which match to the neologisms in both modalities because of the simple (shallow) phoneme-grapheme mapping rules for Dutch.

Because we created linguistic contexts for the neologisms and their stems that were as natural as possible, we had to inflect many stems and some of the neologisms. Over 70% of the stem primes were inflected in the familiarization phases of all three experiments, while no more than five percent of the neologism primes were inflected. Inflection comprised of the adding of inflectional prefixes, interfixes, and suffixes (e.g., *visrijk* vs. *visrijke* 'plentiful fish', *verkreuk* 'wrinkle' vs. *verkreukte* 'wrinkled', *omduikel* 'tumble over' vs. *omgeduikeld* 'tumbled over'). Furthermore, there were some phonological differences between the stems in the stem primes and in their corresponding target neologisms, due to final devoicing (e.g., /bæklœyʃ/ 'endure' as stem for /bæklœyʋɪŋ/ 'enduring') and vowel substitution in irregular verbs (e.g., /vərkek/ 'misjudged' as stem for /vərkeikɪŋ/ 'the misjudgement').

The inflected words functioning as the stem primes can themselves be stored in the mental lexicon and processed via these lexical representations (e.g., Butterworth, 1983; Rumelhart & McClelland, 1986; Seidenberg & Gonnerman, 2000). One might therefore argue that due to our operation of the stem priming condition, this condition is not fit as a baseline in order to investigate whether priming in the identity priming condition results from the full neologism or just its stem. There is, however, evidence that even when morphologically complex words are processed via their full lexical representations, the corresponding stems are activated as well. For instance, Stemberger and MacWhinney (1986) showed that the processing speed of inflections is not only influenced by the frequencies of the inflections themselves, but also by the frequencies of their stems. Furthermore, Crepaldi, Rastle, Coltheart, and Nickels (2010) showed that irregularly inflected words prime their base words more than orthographically matched and unrelated control words. So, *fell* primes *fall* more than *fill* or *hope* do. This strongly suggests that although many stems in the stem priming condition were inflected, the stem priming condition was fit as a baseline.

We observed differences in the processing of neologisms ending in the suffixes *-baar*, *-heid*, and *-ing*. The suffixes *-baar* and *-heid* are more productive than the suffix *-ing* (e.g., Booij, 2002; Van Haeringen, 1971), and one may therefore expect

that neologisms ending in the suffixes *-baar* and *-heid* are processed faster than neologisms ending in the suffix *-ing*, which we observed in Experiment 1. Also in Experiment 2, neologisms ending in the suffix *-baar* were processed faster than neologisms ending in the suffix *-ing*, but we did not find a difference in processing speed between the neologisms ending in the suffix *-ing* and those ending in the suffix *-heid*. In Experiment 3, we observed a different pattern: Neologisms ending in the suffix *-baar* were processed more slowly than the neologisms ending in the suffixes *-heid* and *-ing*. A possible explanation for this unexpected pattern is that the targets in Experiment 3 were presented in sentences, whereas in Experiments 1 and 2, they were presented in isolation. In the sentences of Experiment 3, 37 out of 40 neologisms ending in the suffix *-heid* and 32 out of 40 neologisms ending in the suffix *-ing* were preceded by determiners and possessive pronouns, whereas 31 out of 39 neologisms ending in the suffix *-baar* were preceded by other types of words, including adverbs, nouns, pronouns, adjectives, and conjunctions. The neologisms ending in the suffixes *-heid* and *-ing* were therefore possibly more predictable than the ones ending in the suffix *-baar*, which may explain the difference in reading times. Moreover, 6 out of the 40 target neologisms ending in *-baar* were inflected in Experiment 3, whereas none of the neologisms ending in the suffixes *-heid* and *-ing* were inflected (none of the neologisms were inflected in Experiments 1 and 2 since they were presented in isolation). This may have caused an increase in reading time too.

So far, we have argued that our results can best be explained with the assumption that neologisms are lexically stored in the form of abstract representations. The model developed by Baayen, Milin, Filipovic Durdjevic, Hendrix, and Marelli (2011) offers a different account. This computational model, referred to as the naive discriminative learning model, can account for all data that have so far been interpreted as evidence for the storage of morphologically complex words without actually assuming lexical form representations for these words. The model assigns a central role to the connections between the different units in its semantic and input (orthographic) layers, which have higher weights if the orthographic unit provides more important evidence for the semantic units. These weights are essential in the model's account for all types of frequency effects. The question arises whether this model can account for our observation that just one exposure of a neologism is sufficient to observe an advantage in the subsequent processing of that neologism. One single presentation of a neologism then has a substantial influence on the weights in the naive discriminative learning model. Only simulations can show whether this model can indeed account for the data on neologisms that were presented in this study, in De Vaan et al. (2007), and in De Vaan et al. (2011). Before these simulations can be conducted, however, the model has to be extended such that it can automatically learn new words (outcomes).

For now, we conclude that we presented experimental support for the formation of abstract memory traces for morphologically complex neologisms. The

prior processing of a neologism facilitates the processing of its second occurrence even if the two tokens are presented in different modalities, in different linguistic contexts, in different types of experiments, and if these experiments are presented as two separate experiments.

References

- Alegre, M. and Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language*, 40:41–61. doi:10.1006/jmla.1998.2607
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59:390–412. doi:10.1016/j.jml.2007.12.005
- Baayen, R. H., Milin, P., Filipovic Durdevic, D., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118:438–482. doi:10.1037/a0023851
- Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX lexical database (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.
- Baayen, R. H. and Renouf, A. (1996). Chronicling The Times: Productive Lexical Innovations in an English Newspaper. *Language*, 72:69–96. doi:10.2307/416794
- Bakker, I., Takashima, A., Van Hell, J. G., Janzen, G., and McQueen, J. M. (2014). Competition from unseen or unheard novel words: Lexical consolidation across modalities. *Journal of Memory and Language*, 73:116–130. doi:10.1016/j.jml.2014.03.002
- Bates, D. M. and Sarkar, D. (2005). The lme4 library. [On-line], Available: <http://lib.stat.cmu.edu/R/CRAN/>.
- Bloom, P. (2000). *How Children Learn the Meanings of Words*. MIT Press, Cambridge, MA.
- Booij, G. (2002). *The Morphology of Dutch*. Oxford University Press, Oxford.
- Bradlow, A., Nygaard, L., and Pisoni, D. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception and Psychophysics*, 61:206–219. doi:10.3758/BF03206883
- Butterworth, B. (1983). Lexical representation. In Butterworth, B., editor, *Language production (Vol.II): Development, Writing and other Language Processes*, pages 257–294. Academic Press, London.
- Clark, E. V. (1993). *The Lexicon in Acquisition*. Cambridge University Press, Cambridge. doi:10.1017/CBO9780511554377
- Coolen, R., Van Jaarsveld, H. J., and Schreuder, R. (1991). The interpretation of isolated novel nominal compounds. *Memory and Cognition*, 19:341–352. doi:10.3758/BF03197138
- Craik, F. and Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, 26(2):274–284. doi:10.1080/14640747408400413
- Crawley, M. J. (2002). *Statistical computing. An introduction to data analysis using S-plus*. Wiley, Chichester.
- Crepaldi, D., Rastle, K., Coltheart, M., and Nickels, L. (2010). 'Fell' primes 'fall', but does 'bell' prime 'ball'? Masked priming with irregularly inflected primes. *Journal of Memory and Language*, 63:83–99. doi:10.1016/j.jml.2010.03.002
- De Vaan, L., Ernestus, M., and Schreuder, R. (2011). The lifespan of lexical traces for novel morphologically complex words. *The Mental Lexicon*, 6(3):374–392. doi:10.1075/ml.6.3.02dev

- De Vaan, L., Schreuder, R., and Baayen, R. H. (2007). Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *The Mental Lexicon*, 2(1):1–24. doi:10.1075/ml.2.1.02vaa
- Epstein, M. (2012). *The Transformative Humanities: A Manifesto*. Continuum Publishing Corporation, New York.
- Ernestus, M. and Mak, W. M. (2005). Analogical effects in reading Dutch verb forms. *Memory and Cognition*, 33(7):1160–1173. doi:10.3758/BF03193220
- Faraway, J. J. (2006). *Extending Linear Models with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman & Hall/CRC, Boca Raton, FL.
- Goh, W. (2005). Talker variability and recognition memory: instance-specific and voice specific effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(1):40–53.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5):1166–1183.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105:251–279. doi:10.1037/0033-295X.105.2.251
- Goldinger, S. D., Azuma, T., Kleider, H. M., and Holmes, V. M. (2003). Font-specific memory: more than meets the eye? In Bowers, J. and Marsolek, C., editors, *Rethinking implicit memory*, pages 157–196. Oxford University Press, Oxford.
- Hanique, I., Aalders, E., and Ernestus, M. (2013). How robust are exemplar effects? *The Mental Lexicon*, 8:269–294. doi:10.1075/ml.8.3.01han
- Jaeger, T. F. (2008). Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59:434–446. doi:10.1016/j.jml.2007.11.007
- Janse, E. (2008). Spoken-word processing in aphasia: Effects of item overlap and item repetition. *Brain and Language*, 105:185–198. doi:10.1016/j.bandl.2007.10.002
- Jegerski, J. (2014). Self-paced reading. In Jegerski, J. and VanPatten, B., editors, *Research methods in second language psycholinguistics*, pages 20–49. Routledge, New York.
- Just, M. A., Carpenter, P. A., and Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111:228–238. doi:10.1037/0096-3445.111.2.228
- Koornneef, A. W. and Van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, 54(4):445–465. doi:10.1016/j.jml.2005.12.003
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2016). *Tests in Linear Mixed Effects Models*. <ftp://centos.ustc.edu.cn/CRAN/web/packages/lmerTest/lmerTest.pdf>.
- Mattys, S. and Liss, J. (2008). On building models of spoken-word recognition: When there is as much to learn from natural 'oddities' and artificial normality. *Perception and Psychophysics*, 70(7):1235–1242. doi:10.3758/PP.70.7.1235
- McLennan, C. and Luce, P. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(2):306–321.
- McLennan, C., Luce, P., and Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(4):539–553.
- Palmari, T., Goldinger, S., and Pisoni, D. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(2):309–328.

- Pinker, S. (1991). Rules of language. *Science*, 153:530–535. doi:10.1126/science.1857983
- Pinker, S. and Ullman, M. (2002a). Combination and structure, not gradedness, is the issue: Reply to McClelland and Patterson. *Trends in the Cognitive Sciences*, 6(11):472–474. doi:10.1016/S1364-6613(02)02013-2
- Pinker, S. and Ullman, M. (2002b). The past and future of the past tense. *Trends in the Cognitive Sciences*, 6(11):456–462. doi:10.1016/S1364-6613(02)01990-3
- Rumelhart, D. E. and McClelland, J. L. (1986). On learning the past tenses of English verbs. In McClelland, J. L., Rumelhart, D. E., and the PDP research group, editors, *Parallel distributed processing: Explorations in the microstructure of cognition*. Volume II, pages 216–271. MIT Press, Cambridge, MA.
- Seidenberg, M. S. and Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, 4(9):353–361. doi:10.1016/S1364-6613(00)01515-1
- Stemberger, J. P. and MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory and Cognition*, 14(1):17–26. doi:10.3758/BF03209225
- Swets, B., Desmet, T., Clifton, C., and Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory and Cognition*, 36(1):201–216. doi:10.3758/MC.36.1.201
- Tulving, E. (1972). Episodic and semantic memory. In Tulving, E. and Donaldson, W., editors, *Organization of memory*, pages 381–403. Academic Press, New York.
- Van Haeringen, C. B. (1971). Het achtervoegsel -ing: Mogelijkheden en beperkingen [The suffix -ing: Possibilities and restrictions]. *De Nieuwe Taalgids*, 64:449–468.
- Yim, H., Dennis, S. J., and Sloutsky, V. M. (2013). The development of episodic memory: Items, contexts, and relations. *Psychological Science*, 24:2163–2172. doi:10.1177/0956797613487385

Appendix A

Target materials of Experiments 1, 2, and 3:

annoteerbaar; arceerbaar; balsembaar; bekeurbaar; bekogelbaar; bekrasbaar; bespuwbaar; bestijgbaar; bestormbaar; bestraatbaar; bestuifbaar; betoverbaar; bevuilbaar; bewapenbaar; blondeerbaar; braadbaar; fohnbaar; frankeerbaar; gedenkbaar; omspanbaar; ontkalkbaar; ontkleurbaar; ontruimbaar; ontrukbaar; onttronbaar; ontwapenbaar; overreedbaar; paneerbaar; pocheerbaar; scandeerbaar; shockeerbaar; tutoyeerbaar; verfraaibaar; verknoeibaar; verkreukbaar; vernauwbaar; verpootbaar; verpotbaar; verspilbaar; verversbaar

kortafheid; lobbigheid; labielheid; pitloosheid; saprijkheid; summierheid; tactvolheid; tembaarheid; tilbaarheid; visrijkheid; pipsheid; enormheid; jofelheid; koketheid; onwelheid; ovaalheid; riantheid; antiekheid; banaalheid; basaalheid; dementheid; erkendheid; gammelheid; ludiekheid; ondiepheid; royaalheid; aftandsheid; beschutheid; bezweetheid; contentheid; coulantheid; bebrildheid; markantheid; beuheid; blusbaarheid; geurloosheid; kalkrijkheid; onattentheid; onbelastheid; ontroerdheid

afkluiving; afzouting; beboting; bedamming; bedrinking; bedrupping; begiering; behakking; bekonkeling; bekruising; bekwijling; benageling; beolieing; bepareling; bepoeiering; beriddering; bespijking; betralieing; bewalming; bijeenlegging; omduikeling; omkeiling; ontadeling; ontharsing; ontmugging; onttuiging; ontvloeiing; ontzadeling; oppieping; toezinging; verboersing; vergokking; verkijking; verkletsing; verkuiling; vernaaiing; verrijding; versmeding; verspelding; wegstoting

Appendix B

Example of a story used in the auditory comprehension task in Experiments 2 and 3. Primes (stems and neologisms) are underlined in bold.

Op een **gedenkbaar** dag in juli besteeg Willem in alle vroegte zijn paard. Zijn eindbestemming lag nog zeker een uur rijden bij hem vandaan. In rustige draf reed hij voort, zich bewust van de **verboersing** van het platteland. Willems gedachten richtten zich onvermijdelijk op de vondst die hij zou gaan doen, de vondst die hem beroemd zou maken. Sinds hij geruchten had gehoord over het bestaan van een onbekend werk uit de **antiekheid**, had hij grondig onderzoek verricht naar de mogelijke verblijfplaats van dit werk. Vele plaatsen had hij bezocht, maar door de **summier** informatie die hij had, zat hij telkens op een dood spoor. Maar nu wist Willem het zeker: zijn schat zou hij vinden in de kelder van een oude geneesheer. Het was een norske man, wist Willem, die bekendstond om zijn **bebrild**, vierkant hoofd en haakneus. Het zou Willem moeite kosten de man te overtuigen dat hij zijn kelder moest inspecteren. Na een lange **bekonkeling** had hij besloten zich voor te doen als een ongediertebestrijder. Met zijn redeneerkunst zou hij de oude man vast kunnen overhalen, zijn welbespraaktheid was tenslotte een **erkendheid** in zijn geboortedorp. Verzonken in gedachten en inwendig trillend van spanning, reed Willem gestaag door. Wat zou er gebeuren als hij zijn vondst kenbaar zou maken aan de wereld? Hij stelde zich een grootse huldiging voor en zag hoe hij **beridderd** en bewierookt werd. Deze droom duurde niet lang, want Willem zag dat hij zich op nog slechts enkele meters van zijn bestemming bevond. Hij stapte van zijn paard. De **ontzadeling** kon hem niet snel genoeg gaan. Zijn bovenlip was **bepareld** met zweet, hij veegde het weg en stapte op de deur af, de deur naar eeuwige roem. Hij klopte kort driemaal achter elkaar. De deur ging open en Willem staarde naar een **markante** man. Wat volgde was een lange stilte waarin de twee mannen elkaar taxeerden. Willem stelde zich voor als dienstverlener, hij zou de man een dienst verlenen en zijn kelder vrijmaken van ongedierte. Hoe **basaal**, dacht hij bij zichzelf. Tot zijn grote verbazing toonde de oude man een allesomvattende **pitloosheid** en werd hij direct de woning binnen gelaten. De kamer omringde hem met kruiden, **versmede** handwerktuigen en een arsenaal aan potjes met een ondefinieerbare inhoud. De kraan was groen uitgeslagen terwijl die toch makkelijk te **ontkalken** zou moeten zijn. De muren en het plafond vertoonden duidelijke tekenen van achterstallig onderhoud. De geneesheer zette zich neer en **frankeerde** enkele enveloppen, schijnbaar zonder enige acht te slaan op zijn bezoeker. Willem twijfelde even en verontschuldigde zich toen voor zijn onaangename bezoek. De man antwoordde dat hij **onbelast** was en wees met een handgebaar aan waar de toegang tot de kelder zich bevond. Willem liep naar de deur toe en opende deze. Voorzichtig liep hij de houten trap af, inwendig boos op zichzelf voor de **bekwijling** van zijn boord. Hij depte die droog en inspecteerde de ruimte. Zijn oog viel meteen op een ogenschijnlijk normale doos die dienst deed als tafelsteun. Hij sjoerde even aan het tafelblad en constateerde dat dit goed **tilbaar** was. Hij pakte het blad op en legde het op de grond. Met verhoogde hartslag maakte hij de doos open. Hij verwijderde een paar vellen **verkreukt** papier totdat een groot, gerafeld boek tevoorschijn kwam. Willem haalde het boek uit de doos en een hevige **ontroerdheid** maakte zich van hem meester. Zijn zoektocht was eindelijk voorbij, zijn schat lag in zijn handen. Vanaf nu zou hij **onttroonbaar** zijn en de grootste ontdekking in de geschiedenis gedaan hebben. Hij opende het boek en las de beginzinnen. Het waren de mooiste zinnen die hij ooit zou lezen, perfect **scandeerbaar** en van een grote verhevenheid doordrenkt. Willem stelde zich voor hoe hij de zinnen zou overschrijven, ze **arceren** en ze eindeloos herhalen. Voor een boek uit het verre verleden verkeerde het in goede staat. Het was hier en daar verouderd en gehavend, maar met de nieuwste technieken zou het zeker **verfraaibaar** zijn. Hij zou het boek bij thuiskomst eerst **balsemen** zodat het niet verder zou beschadigen. Hij las verder en ontdekte een verhaal over een oude bekende familie, te gronde gebracht door hun **ontdeling** na

*een familieschandaal. Een lastig verhaal dat om uitleg voor de lezer vroeg. Maar het verhaal was goed **anoteerbaar** en Willem zag voor zich hoe hij iedereen zou verbazen met zijn unieke editie van dit boek. Zo zou hij het gaan doen, hij zou het boek eerst editeren en daarna pas bekendmaken. Zelfs als rijksten van het land zouden besluiten tot de **bijeenlegging** van al hun goudstukken, zouden zij de waarde van deze vondst nog niet behalen. Vergenoegd stopte Willem het boek onder zijn kleren, liep naar boven, mompelde wat tegen de oude man en verliet het huis. Buiten stapte hij op zijn paard, dat hij direct kon **bestijgen**, en reed in galop huiswaarts.*

'On a **rememberable** day in July, early in the morning, Willem mounted his horse. His final destination was at least an hour's ride away. He rode at a calm trot, while he was aware of the countryside **becoming rustic**. Willem's thoughts inevitably focused on the discovery that he was about to make, the discovery that would make him famous. Since he had heard rumors about the existence of an unknown work of **antiquity**, he had thoroughly investigated the possible whereabouts of this work. He visited many places, but he always hit a dead end because of the **scant** information he had. But now Willem knew for sure that he would find his treasure in the basement of an old physician's house. Willem knew the physician was a surly man, who was known for his **bespectacled**, square head and hooked nose. Willem would find it hard to convince the man of the need to inspect his basement. After a long **intrigue**, he decided to impersonate a pest control technician. Using this disguise and rhetoric, he would persuade the old man. After all, his eloquence was a **recognized nature** in his native village. Willem rode steadily while deep in thought, and trembled inside with suspense. What would happen if he made his discovery known to the world? He imagined a great homage and saw himself being **knighthed** and incensed. This dream did not last long, as Willem saw that he was only a few meters from his destination. He got off his horse. The **desaddling** could not go fast enough. His upper lip was **pearled** with sweat; he wiped it away and walked to the door, the door to eternal fame. He knocked three times in short succession. The door opened and Willem stared at a **remarkable** man. What followed was a long silence in which the two men rated each other. Willem introduced himself as a service provider, he would provide a service to the man and make the cellar vermin-free. How **basic**, he thought. To his surprise the old man showed an overall **spiritlessness** and he was admitted directly into the man's home. He was surrounded by herbs, **welded** hand tools, including an array of pots with indefinable contents. The tap had turned green even though it should have been easy to **descal** the tap. The walls and ceiling showed clear signs of lack of maintenance. The doctor sat down and **stamped** some envelopes, seemingly without any regard for his visitor. Willem hesitated a moment and then apologized for his unannounced visit. The man replied that he was **unencumbered** and pointed with a hand gesture to where the access to the basement was. Willem went to the door and opened it. Carefully he walked down the wooden stairs, internally angry at himself for the **slobber** on his collar. He dabbed it dry and inspected the room. His eye immediately fell on a seemingly ordinary box that served as a table support. He tried to move the tabletop and noticed that it was easy to **lift**. He lifted the tabletop and put it on the ground. With increased heart rate, he opened the box. He removed a few **crumpled** sheets of paper until a large, frayed book appeared. Willem took the book out of the box and intense **emotions** overcame him. His search was finally over, his treasure was in his hands. From now on he would be **unthronable**, and he did the greatest discovery in history. He opened the book and read its first sentences. They were the most beautiful sentences he had ever read, perfectly **parsed** and a high level of cohesion. Willem imagined how he would copy the sentences, **shade** them and repeat them endlessly. For an ancient book, it was in good condition. In some places, it was old and battered, but with the newest techniques it would definitely be **trimable**. When he arrived home, first he would **preserve** the book so that it could not be

damaged any further. He read further and discovered a story about an old family, now ruined and **denobled** because of a family scandal. A difficult story that had to be explained to the reader. But the story was well **annotatable**, and Willem imagined how everyone would be surprised by this unique edition of the book. This is the way he would do it: first he would edit the book and then publish it. Even if the richest people in the country decided to **put together** all their gold, they wouldnt come close to the value of this discovery. Willem, content put the book under his clothes, went up, muttered something to the old man and left the house. Outside he mounted his horse, which he could immediately **ascend**, and rode home at a gallop.’

Appendix C

Example of target material from Experiment 3. Neologisms are underlined in bold.

*Sinds de invoering van het rookverbod zijn eigenaren van cafés **bekeurbaar** als zij hun klanten toch toestaan binnen te roken. Ondanks alle acties en protesten van de eigenaren toont de minister van Volksgezondheid zich nog altijd moeilijk **overreedbaar** om het verbod af te schaffen. Alleen voor kleine cafés zonder personeel geldt het verbod niet. De **beuheid** van de café-eigenaren over het rookverbod is inmiddels tot grote hoogte gestegen. Wat hun klanten in hun café doen, of dat nu gaat om het opsteken van een sigaret of om **bedrinking** met wodka, is volgens hen geen zaak voor de politiek. Bovendien verliezen ze door deze maatregel veel klanten en dus veel omzet. Zij beweren dat mensen nu liever thuis blijven om iets te drinken, omdat ze daar niet naar buiten hoeven voor een sigaret. Veel eigenaren laten hun klanten nog altijd in hun café roken, hoewel ze hiermee flinke boetes riskeren. De eigenaren blijven intussen strijden tegen het verbod en bereiden een nieuwe actie voor, die in zijn **ludiekheid** ongeëvenaard zal zijn. Wat de actie precies inhoudt en wanneer deze zal plaatsvinden, willen de eigenaren niet zeggen.*

Overtreden veel café-eigenaren de wet?

’Since the introduction of the smoking ban, owners of cafes are **fineable** if they still allow their customers to smoke inside. Despite all the owners’ actions and protests, the Minister of Health is still not easily **persuadable** about abolishing the ban. Only for small cafes without staff does the prohibition not apply. In the meantime, the **stiredness** of the cafe owners about the smoking ban has risen to great heights. What their customers do in their cafe, whether lighting up a cigarette or **drinking too much** vodka, according to them, has nothing to do with politics. Moreover, this law has lost them many customers and the associated turnover. They claim that people prefer to stay at home to drink, because then they don’t have to go outside for a cigarette. Many owners still let their customers smoke in their cafes, even though they risk high fines. The owners continue to fight the ban and prepare for new action, which will be unmatched in its **absurdity**. The owners won’t say what the action is and when it will occur.’

’Do many cafe owners violate the law?’

Corresponding address

Laura de Vaan
 Radboud University Nijmegen
 Centre for Language Studies
 P.O. Box 9103, 6500 AD, Nijmegen
 The Netherlands
 laura@lauradevaan.nl