Reading homophonous verb forms:

An eye-tracking experiment

Tijn Schmitz
4034570
August 31, 2018
Master Thesis
Word count: 24,204
Supervisor: Mirjam Ernestus
Second reader: Robert Chamalaun

Acknowledgements

First and foremost, I owe Mirjam Ernestus an immense *thank you.* One year ago, we did not know each other. I sent her an e-mail telling that I wanted to do something with spelling errors, and despite her extremely busy schedule, she agreed to supervise me. In the time I worked on this thesis, I acquired an incredible amount of new academic skills, varying from writing to presenting, and from methodological to statistical skills. Under the supervision of Mirjam, I learned more than I ever learned in any course during my study program. I have always been a perfectionist, but Mirjam succeeded in bringing my perfectionism to a different level.

The second person I want to thank is Robert Chamalaun. One of the tweets used as stimulus in this experiment was *Ik stel af en toe echt onmogelijk domme vragen aan mijn docent, maar hij beantwoordt ze altijd lief en rustig* ‘Sometimes I ask my teacher really stupid questions, but he always answers them nicely and quietly’. Robert could well be this teacher. He was always prepared to help, and his comments were always formulated in a positive way. Furthermore, it would not have been possible to test so many participants in such a short time without the financial support from his project.

I also want to thank Helen de Hoop, for always supporting me, also when I decided to change my thesis topic. Without all the opportunities Helen provided me with during my study, I would not have come as far in the academic world as I am now.

Two other persons that deserve to be thanked, are Margret van Beuningen and Bob Rosbag. When I started testing in the CLS Lab, they moved the eye-tracker to a separate room especially for me, as the lab was so busy at that time. Without this brilliant solution, I would have needed at least a month of extra time for testing.

I also want to thank Theresa Redl, for being my eye-tracking oracle, and Thijs Trompenaars, for his *Alziend Oog* (‘all-seeying eye’).

Another person that has been very important to me in writing this thesis, is my brother Luuk. We were thesis-buddies for a long time, and had many fruitful discussions about both the content of our theses and the writing process itself. It was very nice to have someone that understood so well what I was going through, and gave me exactly the right amount of distraction at the right times. Furthermore, he helped me out in formatting all my tables to LaTeX, which was not the most exciting part of writing this thesis.

Then, I want to thank my dear parents, for always being proud of me, believing in me, being interested in what I do, supporting me, and taking care of me during the many ups and downs I experienced while writing this thesis. It was thanks to you (and the paddling pool in the garden) that I survived writing my thesis during the record-heat wave of this summer.

The final person I want to thank is Jana. For always believing in me, even when I was not able to do so myself. For always supporting me, for taking care of me, and for always exactly knowing what I needed, even when this was in conflict with your own needs. Thank you for letting me confiscate our study room right after we moved into our new apartment and for accepting my sometimes inhuman bedtimes. I will forgive you for occasionally interrupting me with random facts about Pomsky's, the Sims, or Mario. It were hard times, and not only for me, but without your love and support it would even have been harder to finish this thesis. I love you.

This thesis has been a tough mountain to climb, and I am extremely proud and grateful that I made it to the top.

Table of contents

Abstract

Although the Dutch verb spelling system seems to be very straightforward, many spelling errors are made, both by children and adults (e.g., Sandra, Frisson, & Daems, 2004). These errors mainly occur with homophonous verb forms, which are common in the inflectional paradigm of Dutch verbs. While many studies investigated factors important in the production of these homophones, less is known about the processes underlying their perception. By means of an eye-tracking experiment with spontaneously produced sentences containing correctly and incorrectly spelled homophones, I investigated whether two factors found to be important in spelling homophones, namely whole-word frequency and verb suffix (<d>/<dt>), also affect the online perception process of these homophones. In production, homophones that are relatively frequent are more easily produced, compared to their homophone counterparts with a relatively lower frequency. Similarly, forms ending in <d> are more easily produced than forms ending in <dt>. The results show that these factors are also important in perception, and that errors that are made more often, are initially overlooked more often during reading, but lead to a processing delay in a later stage. The fact that the factors I investigated have different effects at different stages of the reading process, supports the assumption that a frequency-based retrieval procedure and a rule-based computational procedure simultaneously try to determine the correct spelling and are constantly in competition with each other. This can be explained in terms of Parallel Dual Route Models of spelling. In contrast to spelling production, however, in perception the competition between the two routes does not necessarily result in a single form, but can be seen as more dynamic and may vary over time during the perception process.

1 Introduction

Spelling errors are an informative phenomenon for theories of language production. Studying spelling errors can be helpful in making inferences regarding principles underlying lexical representation and morphological and phonological processing. Do people solely rely on the conscious application of explicitly taught rules, or is the spelling process also affected by other factors, such as word frequency, and, more generally, any form of statistical regularities of the language? While many studies so far investigated the production processes underlying spelling, the current study investigates which factors play a role in the perception of both correct and incorrect spelling in everyday language behavior, and what this can tell us about the cognitive processes underlying the spelling process in general.

In the process of spelling, the acoustic form of a word has to be transformed to an orthographic representation. When a word is non-homophonic, this process is relatively simple, as there is only one orthographic representation that matches the acoustic form of the word. However, for homophonous words, the spelling process is much more complex. When a single acoustic form has multiple, different orthographic representations, a choice between these orthographic representations has to be made. This choice has to be made in a very limited period of time, while the remainder of the sentence has to be planned simultaneously. It is therefore not surprising that it is especially this group of words – homophones – that represents a large part of the spelling difficulties in many languages (see e.g., Assink, 1985; Bertram, Hyönä, & Laine, 2000; Bosman, 2005; Largy, Fayol, & Lemaire, 1996; Sandra & Fayol, 2003; Verhaert, 2016). Thus, homophones provide us with an especially interesting situation in the study of spelling and the cognitive processes behind it.

In spelling homophonous words, multiple strategies can be used to obtain an orthographic representation. Obviously, correctly applying the relevant spelling rules always leads to the correct spelling. This procedure is also referred to as the *computational procedure* (Sandra, Frisson, & Daems, 1999; Verhaert, Danckaert, & Sandra, 2016). For this procedure, the relevant spelling rule has to be determined, and information relevant for the application of this rule has to be stored in working memory until the rule has been applied. As working memory has only limited capacity, it becomes harder for the computational procedure to determine the correct spelling when the cognitive load increases.

In addition, the speed and ease of the computational procedure are dependent on the lemma frequency of the word form. As the lemma frequency of a word form increases, the lemma can be accessed earlier (see e.g. McCormick, Brysbaert, & Rastle, 2009; Taft, 1979) and hence, the process of applying the relevant spelling rule can take place in an earlier stage. In addition, fewer errors are made when the lemma frequency is high (e.g., Verhaert, 2016), which implies that application of the relevant spelling rules is easier when the lemma frequency is high. Vice versa, the processing costs for the computational procedure will increase when the lemma frequency is lower, as it takes more time to access the stem of the word form and to apply the spelling rule to it. In general, this means that, the higher the processing costs are for application of the computational procedure, the less likely it is that this procedure will be decisive in selecting the orthographic representation of the word form that has to be spelled.

Instead of applying the computational procedure and determining the correct spelling using the relevant spelling rules, previous studies suggest that writers often directly retrieve the spelling of a certain form from their mental lexicons (Kapatsinki, 2010; Sandra & van Abbenyen, 2009). By accessing full word forms in long term memory, the process of rule-application can be skipped. When a word form is more frequent, it is more strongly represented in the mental lexicon, which means that accessing the form is easier and faster (e.g., Rubenstein & Pollack, 1963; Whaley, 1978). This retrieval procedure works well for non-homophonous forms: The more frequent a word form is, the faster it can be retrieved from long term memory. When the form has been retrieved, there is no need to wait until the computational procedure has finished determining how the word form should be spelled according to the spelling rules.

Although the retrieval procedure is helpful and can speed up the production of non-homophonous words, it leads to problems in spelling homophonous forms. The multiple different spelling patterns of a homophone pair are all represented in the mental lexicon and how strong each representation is, depends on its whole-word frequency. When one of the homophones is more frequent than the other form, the highly frequent form will be accessed faster, even when the lower frequent form would have been the correct form. The bias towards the highly frequent form is stronger when the difference in frequency between the two forms increases. This effect, sometimes referred to as the *Homophone Dominance Effect*, has been demonstrated in various languages, including Dutch (Assink, 1985; Frisson & Sandra, 2002, Sandra et al., 1999), French (Bonin & Fayol, 2002; Largy et al., 1996), English (White, Abrams, Zoller, & Gibson, 2008), Finnish (Bertram, Laine, Baayen, Schreuder, & Hyönä, 2000), and Mandarin Chinese (Caramazza, Costa, & Miozzo, 2001). As a result of the homophone dominance effect, the retrieval procedure is prone to errors in case of homophones, as it is impossible to determine the correct spelling of a homophonous word form by only relying on frequency information, without taking the word's grammatical function into account and applying the relevant spelling rules.

In the retrieval of words from the mental lexicon, not only frequency information of single word forms is of importance: The mental lexicon also contains frequently-occurring fixed combinations of words, or *multi-word units*. Sprenger, Levelt, and Kempen (2006) for instance showed that multi-word units in the form of idioms (e.g., *kick the bucket*) have their own representation in the mental lexicon. Similarly, Arnon and Snider (2010) showed that non-idiomatic four-word phrases (e.g., *don't have to worry*) are processed faster when they are more frequent, which implies that they are more strongly represented in the mental lexicon. When a frequent multi-word unit contains a homophonous form, this form may therefore be more often spelled correctly in the context of the multi-word unit.

Although the storage of multi-word units in the mental lexicon can facilitate the production of these word combinations, it can sometimes cause problems as well. Any word combination that is perceived often, automatically gets a stronger representation in the mental lexicon, also when the combination itself would be illegal without grammatical context. In a sentence context, for instance, the Dutch words *het* 'it' and *gebeurd* 'happened' may follow each other in certain grammatical structures (such as *dat het gebeurd is* 'that it has happened'), while the combination *het *gebeurd* is ungrammatical in itself. Still, it is likely to be represented in the mental lexicon when perceived often in grammatically correct contexts. The fact that it is homophonous with the correct *het gebeurt* 'it happens' may explain why many spelling errors are found in situations like these. In addition, the mental representation of these ungrammatical combinations automatically becomes stronger every time people erroneously use them, ultimately leading to even more errors.

In summary, both the retrieval and the computational procedure have advantages and disadvantages in the spelling process: the retrieval procedure is often fast, but not always correct; the computational procedure is always correct, but involves more processing and is therefore slower. According to Parallel Dual Route Models of morphology (e.g., Baayen, Dijkstra, & Schreuder, 1997; Bertram, Schreuder, & Baayen, 2000; Laudanna & Burani, 1985), the two procedures form two separate but simultaneous routes which are constantly in competition during spelling, and the route that comes with an output the fastest determines the outcome. When a form is much more frequent than its homophone counterpart, the bias towards the highly frequent form makes it likely for that form to become the winning form, in which case the retrieval procedure wins from the computational procedure. Simultaneously, the computational procedure tries to apply the relevant spelling rule to the stem of the word. The smaller the bias towards one of the forms is, the more likely it is that the computational procedure will be determinant in what the output form should be.

As was already briefly mentioned above, cognitive load is an important factor influencing the spelling process. In the competition between the computational and the retrieval procedure, a higher cognitive load makes it less likely that the computational procedure is able to successfully apply the relevant spelling rules, and as a result, it becomes more likely that the form that is retrieved the first will be the winning form. An increased cognitive load can have various reasons. In terms of language-external factors, time pressure is a well-known example to increase cognitive load (e.g., Paas, Tuovinen, Tabbers, & van Gerven, 2003). It has been shown that more spelling errors are made when there is only limited time available, as the spelling process via the computational procedure is impeded (see e.g. Sandra et al., 2004). The same holds for other cognitively demanding situations, such as performing an additional task (e.g., recalling word sequences or click counting) during the spelling process. In an experiment on French, Fayol, Largy, and Lemaire (1994) demonstrated that almost no subject-verb agreement errors were made when sentences had to be recalled in isolation, but that the number of errors significantly increased when the participants simultaneously had to perform an additional task.

Also language-internal factors can result in an increased cognitive load during the spelling process. When the spelling of a certain word form depends on the grammatical properties of another word (as is the case in, for instance, subject-verb agreement), it is easiest to determine the spelling when these two forms are adjacent. In that case, the relevant information is still salient in working memory when the word form has to be spelled, and the information is not in competition with other information relevant for the production of the rest of the sentence. In an experiment on French subject-verb agreement, Largy et al. (1996) showed that, in sentences like (1), participants in some cases tend to write the third person plural *arrivent*, which is a homophone of the (correct) third person singular *arrive*. The fact that the intervening *des voisins* 'of the neighbours' is plural provides misleading information in determining the verb form, leading to an error in subject-verb congruence. This shows that people cannot always inhibit intervening information during the spelling process and instead tend to rely on information that is most accessible in working memory.

(1)	Le	chien	des	voisins	arrive.
	The	dog	POSS-PL	neighbor-PL	arrive-3-SG

'The neighbours' dog is arriving.'

A comparable situation exists in Dutch, which contains many homophones in the inflectional paradigm of verb forms. An example is the homophone pair *betaalt*/*betaald* 'pays/ payed', where the former is a second/third person singular present tense form and the latter is a past participle. When the subject and the verb form are adjacent (as in *hij betaalt* 'he pays'), the information from the subject, needed to select the verb suffix, is still salient in working memory at the moment when the verb form has to be determined, facilitating a correct spelling. However, when the subject and the verb form are not adjacent, as in *dat hij mij morgen eindelijk betaalt* 'that he will finally pay me tomorrow', more errors on subject-verb agreement are found, as the subject *hij* 'he' and the verb form *betaalt* 'pays' are separated by three intervening words (see e.g. Assink, 1985; Sandra et al., 1999, 2004). Furthermore, as in the French example in (1), sometimes conflicting information can impede the correct spelling in Dutch as well, as is illustrated in (2). Here, the subject *hij* 'he' is adjacent to the past participle *verbaasd* 'surprised', which might cause confusion and encourage to write the present tense *verbaast* 'surprises' instead. Additionally, the information from the auxiliary verb *is* 'is', which points out that the verb form is a past participle, only becomes available after processing the verb form itself. The combination of these two factors makes that the past participle verbaasd is often misspelled as the third person singular verbaast in contexts like these (Assink, 1985).

(2)	Dat	hij	verbaasd	is,	had	ik	wel	verwacht.
	That	he	surprised-PP	is-AUX	had	I	DM	expected

'I expected that he is surprised.'

Spelling errors in homophonous verb forms, such as the examples described above, are a common problem in Dutch written language. Even experienced writers produce an unexpectedly large number of errors on regularly inflected homophonous verb forms. Sandra (2010) for instance showed that Dutch 18-year-olds make up to 25% errors in spelling homophonous verb forms. This large number of errors shows that, despite the relative simplicity of the spelling rules, homophones lead to problems in the spelling process.

In Dutch, regularly inflected verb forms are formed by morphological rules which concatenate the stem and one or multiple suffixes. Two important spelling rules underlie a large part of the verb spelling errors. The first rule, illustrated in example (3), marks second and third person singular by adding <t> to the stem.

(3) Hij werk+t   /ɦɛi ʋɛrk+t/   'He works'

In verbs with a stem-final <d>, the application of this rule results in a homophone pair. As illustrated in (4) for the verbal stem *vind* 'find', these verbs have a form ending in <d> for the first person singular and a form ending in <dt> for the second and third person singular. As the Dutch phonological system contains rules for final devoicing and degemination, both verb forms are pronounced identically, namely as ending in /t/.

(4) a. Ik vind   /ɪk vɪn**t**/   'I find'
b. Hij vind+t   /ɦɛi vɪn**t**/   'He finds'

Combined with the spelling rule explained in (3), a second rule responsible for many verb spelling errors marks the past participle and adds the prefix <ge> and either the suffix <d> or <t>, depending on the last sound of the stem. When this sound is voiceless, as is the case for the /k/ in *werk* in (5a), the suffix is <t>; when it is voiced, as the /m/ in *noem* in (5b), the suffix is <d>.

(5) a. ge+werk+t   /xə+ʋɛrk+t/   'worked'
b. ge+noem+d   /xə+num+t/   'mentioned'

In so-called *weak-prefix verbs*, application of this rule again results in a homophone pair. Weak-prefix verbs are verbs starting with an unstressed (semi-)prefix (e.g., *geloof* 'believe', *betaal* 'pay'). In forming the past participle of these verbs, no additional prefix <ge> is used, and this causes a homophone pair of the third person singular and the past participle, as is illustrated in (6). Again, due to final devoicing, both forms are pronounced as ending in /t/.

(6) a. Het gebeur+t   /ɦɛt xəbør**t**/   'It happens'
b. Het is gebeur+d   /ɦɛt is xəbør**t**/   'It has happened'

As a result of the two rules demonstrated above, the Dutch inflectional verb paradigm contains verb forms ending in <d>, <dt>, and <t>. Due to the word-final position of these consonants or consonant clusters, they are all pronounced identically, namely as /t/. This explains why so many errors are made in Dutch verb spelling.

A common finding is that there exists a preference to write <d>, rather than <t> or <dt>. Bosman (2005) for instance found a tendency to write <d> instead of <t> in weak-prefix verbs. Frisson and Sandra (2002) also found a <d>-bias: Participants tended to write <d> instead of <dt> more often than the other way around. Schmitz, Chamalaun, and Ernestus (2018-in press) also found a preference for <d> over both <dt> and <t> in a corpus study on verb spelling errors in spontaneously produced language.

The <d>-preference can be explained in several ways. First, it is again a matter of frequency. Ernestus and Mak (2005) explain that <d> is much more frequent in the inflectional paradigm of Dutch verbs, compared to other word-final segments. For the verbal stem *leg* 'lay', for instance, <t> is only written in the second/third person singular *legt*, while <d> is present in many other forms (e.g., the singular simple past *legde*, the past participle *gelegd*, the present participle *leggend*, etc.). Ernestus and Mak (2005) showed that people prefer analogy in the inflectional paradigm, and as a result, they have a preference for word-final <d> in the entire paradigm. Another explanation for the <d>-preference is hypercorrection (Neijt & Schreuder, 2007). As a result of final devoicing, the sound /t/ is sometimes spelled as <d> in Dutch (namely, in word-final position). The reverse, however, /d/ spelled as <t>, is systematically absent. This causes a tendency to write <d> instead of <t> (also in situations where it is inappropriate) and explains why the reverse, writing <t> when <d> would have been correct, is found less often. Hanssen, Schreuder, and Neijt (2015) found support for this explanation. They showed that Dutch first-graders initially have a bias to write <t>, as they spell what they hear. Later on, they learn that there are words that sound as ending in /t/ but are spelled with a <d>. As a result of overgeneralization, then, the tendency to write <t> turns into a tendency to write <d>, also in situations in which it is inappropriate and <t> would have been correct. This form of hypercorrection appears to be very persistent and the preference to write <d> only seems to diminish when people get older and become more educated (i.e., advanced high-school students, Frisson & Sandra, 2002; and university students, Bosman, 2005).

The production of spelling errors in Dutch homophonous verb forms has been extensively investigated in experimental settings (e.g., Assink, 1985; Bosman, 2005; Frisson & Sandra, 2002; Sandra et al., 1999, 2004; Verhaert et al., 2016). In these experiments the participants' tasks were, for instance, inserting a verb form in a sentence or metalinguistic tasks such as indicating which strategy they used to determine the spelling of the verb form. Additionally, in many experiments the tasks had to be performed under time pressure, which (as I already discussed above) has been shown to increase the cognitive load, leading to a higher number of errors (Fayol et al., 1994; Sandra et al., 2004). Furthermore, in many experiments participants were aware of the fact that the study was about spelling errors. As a result, they might have used different spelling strategies than they would normally do during spontaneous writing. For instance, it is possible that participants rely more on the computational procedure in an experimental setting than they would do in a natural setting.

In studying the cognitive processes underlying spelling behavior, and in studying all kinds of cognitive processes in general, it is important to maximally approach the natural situation in which these processes normally take place. Determining the correct spelling of a verb form in a provided sentence is different from the spontaneous production of entire sentences, taking into account the correct spelling of each word in the sentence whilst integrating all information relevant for the production of the sentence. For this reason, Schmitz et al. (2018-in press) performed a corpus study on verb spelling errors in spontaneously produced language in the form of tweets. By using tweets, this study gives a better reflection of how people genuinely write in an informal setting and which factors play a role in the spelling errors they produce.

Schmitz et al. (2018-in press) showed that most factors found in experimental studies on the production of spelling errors in homophonous verb forms largely apply to spontaneously produced language as well. They showed that when the intended verb form is less frequent than its homophone counterpart, more errors are made. This is in accordance with many previous studies starting from Assink (1985), and implies that, when a homophonous verb form is more frequent, it is more easily retrieved from the mental lexicon compared to the form that is less frequent. This means that the frequency effect found in experimental studies on production also exists in spontaneous writing. In addition, in accordance with many previous studies, Schmitz et al. (2018-in press) found a preference to write word-final <d>, rather than <dt> or <t>, which means that people prefer to write the form ending in <d> also in spontaneously written language. Schmitz et al. (2018-in press) showed that the suffix preference effect can be overruled by the frequency effect only when the other form is much more frequent, as is the case for the verb *worden*[1]. This suggests that when the frequencies of the two homophones are close to each other, people prefer to write the <d>-form, but when the <dt>-form is much more frequent, the preference shifts to the latter form.

Schmitz et al. (2018-in press) did not find the effect of adjacency that was found in experimental studies. Based on earlier research starting from Assink (1985), it could be expected that fewer errors occur when the verb form and the word determining its suffix are adjacent, as the information from the subject, needed to determine the verb suffix, is still salient in working memory when the verb form has to be spelled. However, Schmitz et al. (2018-in press) found that, in tweets, this effect was reversed, which means that they found *more* errors when the verb form and the word determining its suffix were adjacent. This can be explained by a correlation between adjacency and relative frequency in their dataset: When the relative frequency of the written verb form compared to its homophone counterpart increased (which is associated with fewer errors), the verb form and the word determining its suffix were more often separated. Apparently, the frequency effect played a much more important role than the effect of adjacency and this explains why the effect of adjacency was inhibited to arise in the expected direction. This shows that results found in perfectly controlled experimental settings do not always directly translate to a real language situation, and that this should be kept in mind when interpreting the results of experiments.

In research on the cognitive processes behind spelling and spelling errors, it is important to not only focus on production, but also take perception into account. Neither of them can exist without the other in communication and they share important characteristics, but they also differ on important points and are by no means always exactly each others mirror. Meyer, Huettig, and Levelt (2016) and references therein give an extensive overview of current developments in the comparison between production and perception. The contributors to this special issue largely share the consensus that language production and perception involve skills and representations that are distinct, but tightly linked to each other. However, others, including Gollan et al. (2011), Roelofs (2003), and Zwitserlood (2003), argue that the two processes might be more distinct.

The effect of word frequency is a common example of a factor to which both production

[1]The form *wordt* is much more frequent (41101) than *word* (1209) (frequencies taken from the Dutch Morphology Wordforms from CELEX, Baayen, Piepenbrock, & Van Rijn, 1995).

and perception are sensitive. As we have already seen, in production fewer errors are made on highly frequent forms, compared to lower frequent forms. This is the case for both written production (e.g., Assink, 1985; Largy et al., 1996; Pacton & Fayol, 2003) and speech production (Stemberger & MacWhinney, 1986). In perception, the effect of frequency is comparable: Processing times for highly frequent forms are shorter than for lower frequent forms, both in visual lexical decision (Burani, Salmaso, & Caramazza, 1984; Colé, Segui, & Taft, 1989; Katz, Rexer, & Lukatela, 1990; Sereno & Jongman, 1997; Taft, 1979) and in auditory lexical decision (Baayen, McQueen, Dijkstra, & Schreuder, 2003). These similarities suggest that, although production and perception are different modalities, they make use of the same lexical representations.

Another factor which has been found to be important in both production and perception, is the influence of analogy. In forming the Dutch past tense, <te> or <de> is added to the verbal stem, depending on the voicedness of the last sound of this stem. Despite this clear rule, Ernestus and Baayen (2004) showed that people rely on information from phonological neighbors of the verb to a large extent, rather than applying the rule. For instance, the verbal stem *krab* 'scratch' is pronounced as ending in /p/ through final devoicing. In forming the past tense, this often made participants write *krapte* instead of the correct past tense form *krabde*, analogous to other verbs with the cluster *pte* in their past tense form (e.g., *stapte* 'stepped', *hapte* 'bit', *klopte* 'knocked'). Similarly, Ernestus and Mak (2005) showed that this so-called *sublexical homophony* also influences perception. When the incorrect suffix was supported by the phonological neighbors of the verb, the incorrect form was processed more quickly in a self-paced reading task, compared to forms where the incorrect suffix lacked this support. This again shows that both production and perception make use of the same representations, also from a phonological point of view.

However, it is not always the case that results found in production directly map onto perception and the other way around. Although production and perception share important characteristics, the nature of these two processes differs in some respects. In production, on the one hand, the process begins with meaning and ends with access to lexical forms. This process thus has to select the correct word from a set of semantically related concepts (Levelt, Roelofs, & Meyer, 1999). In perception, on the other hand, the process begins with accessing lexical forms and ends with meaning. In this case, lexical representations are activated when they (partially) match the (orthographic or acoustic) input signal, and lexical access is assumed to be achieved when form-related but irrelevant words are inhibited. Because of the different nature of these two processes, they are also sensitive to different factors, of which I will give several examples below.

Gollan et al. (2011) argue that lexical access is a fundamentally different process in language comprehension and in language production. Although both modalities have been shown to be sensitive to frequency effects, Gollan et al. (2011) show that these effects differ for production and perception. They contrasted the results of a picture naming experiment with the results of a visual lexical decision task and an eye-tracking experiment. They showed that, in a semantically constraining context, frequency effects were larger in perception, but without a constraining context, frequency effects were larger in production. Based on these outcomes, they suggest that production is primarily driven by semantic context, whereas comprehension is primarily frequency-driven.

Other support from studies that argue that comprehension and production are two distinct processes concerns effects of neighborhood density. When a word differs from many other words by only one phoneme, it has a high neighborhood density (e.g., *cat*, which has many neighbors including *bat/cut/car/chat/at*, et cetera). It has been shown that a high neighborhood frequency facilitates the production process (e.g., Slattery, 2009). However, in comprehension, a high neighborhood density sometimes causes a processing delay (e.g., Dell & Gordon, 2003; Goldinger, Luce, & Pisoni, 1989). This discrepancy can be explained by the following. During perception, the target words' neighbors act as competitors, which inhibit recognition of the target word by creating a temporal distraction. During production, however, the neighbors act as primes for the target word. As a result, the target word is activated faster, which facilitates production of the word (Dell & Gordon, 2003).

When the abovementioned studies are compared, it is noteworthy that production and perception show similar effects of analogy in the case of phonological analogy (Ernestus & Baayen, 2004; Ernestus & Mak, 2005), but that there exists a discrepancy between production and perception in the case of orthographical analogy (Dell & Gordon, 2003; Slattery, 2009). Overall, these and other studies discussed above thus show that perception and production are closely related processes which are sensitive to many shared factors, but that they also differ at other points. In order to fully understand the cognitive principles behind the spelling process of homophonous verb forms, it is therefore important to study both modalities and avoid blindly drawing the comparison between them.

A study that systematically compares the production and the perception of spelling errors in Dutch homophonous verb forms and the cognitive processes behind them, is Verhaert (2016). In the first part of her dissertation, she investigates which factors play a role in the production of spelling errors by performing both offline and online production experiments. In line with previous studies, she found more errors involving the use of the highly frequent form instead of the lower frequent form, than the other way around. She explains this by the fact that word forms with high frequencies are activated faster than word forms with lower frequencies, and that participants are sometimes unable to suppress the highly frequent forms. Although the highly frequent forms were processed faster in general, the online task suggested that it was these highly frequent forms that sometimes also caused a delay. Probably, participants sometimes initially rejected the highly frequent form and labelled it as 'suspicious'. Only when the slower computational procedure confirmed or rejected this highly frequent form as the correct form, participants were able to give a response.

After these production studies, the remaining question was whether the cognitive infrastructure underlying spelling also underlies perception. In the perceptional part of her dissertation, Verhaert (2016) therefore examined whether highly frequent homophonous verb forms are also processed faster during reading, and whether errors involving a highly frequent form are overlooked more often than errors involving a lower frequent form. To this purpose, Verhaert (2016) investigated the perception of homophonous verb forms, in isolation, in a minimal context, and in a sentence context. In isolation, both homophonous verb forms are obviously correctly spelled, as there is no grammatical context. Therefore, working memory does not have to identify the grammatical function of the verb form and use this information to determine the correct suffix. In a lexical decision task, Verhaert (2016) found a clear preference for the most frequent homophone, indicating that whole-word representations indeed play an important role in accessing the lexical representation of a homophonous verb form presented in isolation.

When the verb forms were presented in a minimal context (i.e., preceded by the grammatical subject), they could either be correctly or incorrectly spelled. When the verb form was correctly spelled, reaction times in a spelling decision task were shorter and fewer errors were made when the verb form was more frequent. When the verb form was incorrectly spelled, however, reaction times increased and more errors were made when the verb form was more frequent, especially when it was written with <d>. It appears that initially, the highly frequent form causes a tendency to accept the verb form, even if it is incorrect. Only after the grammatical analysis integrating information from the subject and the verb form, it becomes clear that the verb form is incorrectly spelled. In sum, these results again point towards the involvement of whole-word representations in the perception of spelling errors in homophonous verb forms, and indicate that the time-consuming computational procedure can (at least in some cases) be overruled by the quicker process of whole-word retrieval.

In the most natural reading task – an eye-tracking experiment with homophonous verb forms embedded in entire sentences – Verhaert (2016) did not find the same effect of whole word frequency (although she did find a strong trend towards it, especially on the spillover region). A possible explanation to this could be the high skipping rates of the verb form. This problem was solved by performing a self-paced reading task, where word skipping is not possible. Indeed, the frequency effect was found in this task, and again on the spillover region. When the verb form was more frequent than its homophone counterpart, it was processed faster. As a consequence of this faster processing, the risk of missing an error increased as the frequency of the written form increased. Furthermore, the results indicate that the whole-word representations of homophonous verb forms are not only accessed when the verb form is presented in isolation or in a minimal context, but also during sentence reading.

Although the work of Verhaert (2016) forms an important base for both the research on the perception of spelling errors and the comparison with production, a limitation is that it does not reflect perception of spelling errors in spontaneously produced language. The materials used in the eye-tracking experiment consisted of made-up sentences not only containing spelling errors with homophonous verb forms, but also additional, unnatural spelling errors making the sentences less ‘realistic’. An example of such a spelling error is **antiebiotika* instead of *antibiotica* ‘antibiotics’, which is highly unlikely to be written in this way. The high number of superficial errors in the sentences might therefore have distracted the attention from the verb spelling errors, which could explain the lack of a more robust effect of frequency. Furthermore, Verhaert (2016) already points out that the sampling rate she used, 300Hz, might have been too low to measure fast saccades between the different words and that the effect of frequency could have been missed in this way.

The unnaturalness of the stimuli in Verhaert (2016) and possibly also the insufficient sampling rate make it difficult to draw conclusions towards factors important in the perception of genuine spelling errors in real language. This leaves us with a gap in our knowledge of how people read (both correctly and incorrectly spelled) homophones in spontaneously produced language, which I aim to fill with the current study. To this purpose, I performed an eye-tracking experiment with a higher sampling rate than Verhaert (2016) did, and I used tweets containing verb spelling errors as material. Tweets are spontaneously produced utterances, produced outside an experimental testing situation, and thus reflect how people genuinely write in an informal setting. Usually, twitter users are busy with their everyday life while writing a tweet and do not think too long about what they write. This contributes to the spontaneous character of tweets. By using tweets containing correctly and incorrectly spelled homophonous verb forms, it is possible to investigate real errors in real language, and which factors are important in the perception of homophonous verb forms in general.

Eye-tracking closely resembles natural reading. The online perception process of reading can be measured with eye-tracking, and in this way it is possible to shed light on the cognitive processes underlying the reading process, both in terms of the recognition of words and of their integration into a sentence context (Rayner, 1998). The eyes can freely fixate across all words. In contrast to, for instance, a self-paced reading task, in eye-tracking it is possible to skip words and to make regressions to earlier words. This increased freedom in possible reading strategies is conducive to the naturalness of eye-tracking and, although the results are more complex to analyze and interpret than, for instance, those of a self-paced reading task, the naturalness of eye-tracking is able to reveal a complex picture of how cognitive processes in reading unfold over time (Witzel, Witzel, & Forster, 2012).

By using eye-tracking on tweets containing homophonous verb forms, the current study provides an excellent opportunity to compare the production of homophonous verb forms in spontaneous language (Schmitz et al., 2018-in press) with the perceptional side of this topic. Two important factors found to influence the spelling process of homophonous verb forms in production studies – relative frequency and suffix – will be assessed in this study to investigate whether they also play a role in the perception of homophonous verb forms. This leads us to the main question of the current study: What are the effects of relative frequency and suffix on how people read (in)correctly spelled homophonous verb forms occurring in everyday language?

The homophonous verb forms I use in this experiment are associated with the Dutch spelling rule that was illustrated in examples (3) and (4) – the rule that marks second and third person singular by adding <t> to the stem. This leads to a homophone pair with the first person singular in verbs with stem-final <d> (e.g., *vind/vindt*, which are both pronounced as ending in /t/ due to final devoicing and degemination).

I expect that homophonous verb forms that are more easily produced, are processed faster during reading as well. This means that I expect fewer and shorter fixations as the written verb form is more frequent compared to its homophone counterpart, and fewer and shorter fixations on forms written with <d>, compared to forms written with <dt>. Participants will not always become aware of a mismatch between the grammatical subject and the verb form, and there will be no certainty of whether they did or did not register the mismatch. If they do so, I expect to find a delay, but even if participants do not (consciously) register the mismatch, it is possible that I will find a delay as the combination of subject and verb form cannot be retrieved as a unit due to its low frequency. Generally, I expect that errors that are made more often during production, will be overlooked more often during reading. More specifically, this would mean that errors will be overlooked more often when the written form becomes more frequent than its homophone counterpart, and that <d>-substitutions will receive less attention than <dt>-substitutions.

It has been demonstrated that many processes associated with word retrieval take place while a word is within fixation (e.g., Ehrlich & Rayner, 1983). Therefore, I especially expect effects on the verb form itself. However, there is a limit to the range of processes that are carried out immediately during the first fixation on a word. Ehrlich and Rayner (1983) suggest that only processes directly relevant to lexical access and, to a smaller extent, syntactic parsing are carried out before the eyes move on to the next fixation. Thus, more complex processes such as integrating a word into the rest of the sentence are not necessarily completed during the fixation on which the process was initiated (see also Carpenter & Just, 1983, and Just & Carpenter, 1980). Therefore, it is not unlikely to find spillover effects in an eye-tracking task, as more complex processes are sometimes only completed when other, new fixations have already occurred (see also Bertram, Hyönä, et al., 2000; Witzel et al., 2012). In the current experiment, this would mean that the effects I expect on the verb form will perhaps be (partially) delayed and become visible on the words immediately following the verb form.

In addition, readers do not only extract information from the word they are fixating on. During a fixation, information within 2° of the visual angle (or approximately eight characters) is within the foveal vision (e.g., Rayner, 1998). In addition, the processing of information from the parafoveal area, which can be up to 5° of the visual angle, is already initiated as well. As a result, short words are often skipped during reading (e.g., Engbert, Longtin, & Kliegl, 2002; Veldre & Andrews, 2018). It has been shown that the presence of an orthographic illegal sequence or a low-frequent word in the parafoveal area disrupts processing of the word in the fovea (i.e., the word that is fixated on at that moment) (Angele, Slattery, & Rayner, 2016; Drieghe, Rayner, & Pollatsek, 2008; Hutzler et al., 2013; Kliegl, Hohenstein, Yan, & McDonald, 2013). Similarly, when the word in the parafovea is highly frequent, processing of the word in the fovea is facilitated. In the current experiment, this so-called *parafoveal preview benefit* would mean that it is possible that I already find effects of the spelling of the verb form before it actually has been fixated on.

It is difficult to say how the effects I expect will unfold over time. The factors influencing the ease of processing of the homophonous verb form can be in conflict with each other, which leads to a rather complex situation. First, a high frequency is likely to make processing of the verb form easier, and is likely to facilitate the processing of the following word or words as well, while I expect the reverse for forms with a lower frequency (e.g., Demberg & Keller, 2008; Drieghe et al., 2008; Witzel et al., 2012). Second, correctly spelled forms will facilitate processing, compared to incorrectly spelled forms, which are likely to impede the processing of the form itself and the following words (Angele et al., 2016; Drieghe et al., 2008; Hutzler et al., 2013; Kliegl et al., 2013). Third, forms written with <d> will likely be processed faster than forms written with <dt>. This means that there are three factors that can all be either facilitating or inhibitory for the processing of the verb form, and these factors are sometimes in conflict with each other. For instance, a highly frequent form spelled with <dt> leads to a conflict between the preference for this highly frequent form and the preference for the form spelled with <d>. Similarly, when a form is incorrectly spelled but highly frequent, the preferred form based on the frequency information is in conflict with the correct form according to the grammatical information. This means that, depending on the properties of a given form, it is possible to have both facilitating and inhibitory effects resulting from a single form, and I could only speculate on how the effects of the different factors will exactly unfold over time during the reading process.

2 Method

2.1 Participants

Sixty-four participants (41 female) participated in the experiment. Most of them were bachelor's or master's students at Radboud University; a small number was recently graduated. The mean age was 22.7 years (SD: 3.5; range: 18-34). All participants were native speakers of Dutch. None of them reported to suffer from dyslexia, severe eye abnormalities, or other reading problems. Participants with glasses or soft contact lenses were allowed to participate if their vision was corrected-to-normal; hard contact lenses were not allowed as these lead to problems with eye-tracker calibration. Participants were rewarded with a €10 gift card; a single participant received course credit instead of a gift card.

2.2 Design

Two categorical variables were used in the experiment to describe the verb form: *correctness* (correctly or incorrectly spelled) and *correct suffix* (<d> or <dt>)[2], leading to a total of four conditions summarized in Table 1.

Table 1: Overview of the experimental conditions

	Correctly spelled	Incorrectly spelled
Correct suffix: <d>	written: <d>, correct: <d>	written: <dt>, correct: <d>
Correct suffix: <dt>	written: <dt>, correct: <dt>	written: <d>, correct: <dt>

A third variable of interest was the *relative frequency* of the written verb form compared to its homophone counterpart. This numerical variable was calculated using formula (1); frequencies were taken from the Dutch Morphology Wordforms from CELEX (Baayen, Piepenbrock, & van Rijn, 1995).

$$relativefrequency = \log(\frac{frequency_{written}}{frequency_{homophone}} + 1) \tag{1}$$

2.3 Materials

The items used in the experiment were real tweets. They were partially collected from TwiNL, a database of Dutch tweets posted from December 2010 onwards (Tjong Kim Sang & van den Bosch, 2013), and partially via the Twitter search engine (https://twitter.com/search-advanced?lang=nl).

Prior to the main experiment, I conducted a small pilot study where five participants (who did not participate in the main experiment afterwards) had to read twenty tweets. This pilot study showed that, taking into account a maximum experiment duration of one hour, it would be possible to use a total of ±336 tweets in the main experiment. To ensure that the participants would have enough time to finish the experiment, I decided to limit the number of tweets to 320.

[2] I am aware of the fact that, strictly speaking, <d> is the stem-final segment of the verb form in the homophonous verb forms I use, rather than (part of) the suffix. However, for the sake of simplicity and to stress the homophony between <d> and <dt> as word-final segments, I chose to refer to both of them as 'suffix', rather than, for instance, using the levels <Ø> and <t>.

As was demonstrated by Schmitz et al. (2018-in press), about 10% of Dutch tweets contain a misspelled homophonous verb form, and to keep the experiment as natural as possible, I maintained this percentage in the materials. I used 64 target items, which originally all contained a misspelled verb form. I then created a correctly spelled variant of each target item as well. The target items were counterbalanced by correctness of the verb form, which means that 32 out of the 320 tweets presented to a given participant were target items with a misspelled verb form, and 32 were target items with a correctly spelled verb form (but participants never saw the same target item in different conditions). The remainder of 256 tweets in the experiment were fillers. The mean length of the target items was 98.5 characters; the mean length of the fillers was 93.7 characters.

All target items had the same structure, consisting of six Interest Areas (IAs). The structure is illustrated in Table 2. To ensure that the verb form was roughly in the middle of the screen, the target items always started with an introductory part. The subject was always directly followed by the verb form, and after the verb form two words followed to catch possible spillover effects. After the spillover region, a concluding part ended the target item.

Table 2: Example of the structure of a target item

IA1	IA2	IA3	IA4	IA5	IA6
introductory part	subject	verb form	spillover 1	spillover 2	final part
Ik stel af en toe echt onmogelijk domme vragen aan mijn docent, maar	*hij*	*[beantwoord/ beantwoordt]*	*ze*	*altijd*	*lief en rustig*
'Sometimes I ask really stupid questions to my teacher, but	he	answers	them	always	nicely and quietly'

As explained in examples (3) and (4) in the Introduction, the spelling rule marking second/third person singular adds <t> to the verb stem, and leads to a homophone pair with the first person singular in verbs with stem-final <d> (e.g., *word/wordt* 'become(s)'). The verb forms used in the target items were all homophones of this type. Half of the target items were first person singular (corresponding to forms ending in <d>) and the other half were third person singular (corresponding to forms ending in <dt>). Each verb was only used once, meaning that in each target item, the verb form was unique.

It was important to make sure that the materials would not evoke a bias towards forms ending in <d> or forms ending in <dt>. A first way in which I aimed to prevent this bias is that the fillers did not contain any verbs with a stem ending in <d>, to avoid the occurrence of other homophonous verb forms in the experiment. Furthermore, I carefully selected and balanced out the verb forms used in the target items. First, I made sure that for half of the verbs the <d>-form was more frequent than the <dt>-form, and vice versa for the other half of the verbs. This means that the verb forms used in the materials were 50% <d>-dominant verbs and 50% <dt>-dominant verbs. I balanced out the average relative frequency (of the more frequent versus the less frequent form of a given homophone pair) between these two groups of verbs. Second, I also made sure that the average relative frequency did not differ between items with first person singular verb forms (correct spelling with <d>) and items with third person singular verb forms (correct spelling with <dt>). Thus, in all cases, the average relative frequencies were equal between the different groups, so that participants would remain unbiased with respect to the suffix of the verb forms. An overview of the average relative frequencies over the different groups is given in Table 3; for an overview of the verb forms used in the experiment, see the Appendix.

Table 3: Average relative frequencies of more frequent versus less frequent verb forms (after logarithmic transformation) in the materials

	<d>-dominant	<dt>-dominant
correct: <d>	0.66 (range: 0.05-1.07)	0.69 (range: 0.03-0.98)
correct: <dt>	0.62 (range: 0.04-1.23)	0.65 (range: 0.04-1.17)

A further important factor in the selection of the materials was that the items did not contain any features which might attract the attention of the participants and in this way influence the natural reading process. Therefore, hashtags and user-tags were not included in the experiment. For the target items, I only selected tweets without tags, and in the fillers I removed the tags if present (N=17). Furthermore, the tweets did not contain other spelling mistakes, nor did they contain emojis. The only adjustments made to the tweets were:

- If the tweet started with a lower case letter, I capitalized this letter.
- If the tweet did not end with a period, I added a period at the end of the tweet.
- The Dutch possessive pronoun *mijn* 'my' is often abbreviated to (correct) *m'n* or (incorrect) *mn* in informal written language. When the incorrect abbreviation was used, I consistently changed it to the correct *m'n* (N=5).

25% of the stimuli were followed by yes-/no-content questions. The questions were proportionally distributed over fillers and targets, meaning that 16 of the 64 target items and 64 of the 256 fillers were followed by questions. These questions targeted the general content of the items, to make sure participants would pay attention while reading. Two examples of target items with questions, to be answered with respectively *yes* and *no*, are given in (7) and (8).

(7) *Toiletten in de trein zijn altijd vies, dus ik [mijd/mijdt] ze als het maar even kan.*
Question: *Gaat deze persoon weleens met de trein?*
'Train toilets are always dirty, so I avoid them if it's remotely possible.'
Question: 'Does this person travel by train sometimes?'
Answer: Yes

(8) *Ik wil eigenlijk tv kijken, maar ik [verbied/verbiedt] het mezelf tot ik klaar ben met leren.*
Question: *Is deze persoon al klaar met leren?*
'I actually want to watch TV, but I won't let myself until I have finished studying.'
Question: 'Has this person already finished studying?'
Answer: No

I used three pairs of lists, resulting in a total of six lists. Each list consisted of 320 items (64 targets, half of which contained a misspelled verb form, and 256 fillers). In each pair of lists, the second list was an exact copy of the first list, except that the correctness of the verb forms was mirrored: If the correct version of an item was used in the first list, the incorrect version was used in the second list and vice versa. I made sure that each list contained an equal number of <d>-forms and <dt>-forms, and that both forms contained an equal number of errors. Furthermore, the average relative frequency was comparable in each of the six lists and over correctly- and incorrectly spelled verb forms per list.

Each list started with a practice block. The practice block had a fixed order and consisted of four practice items, two of which were followed by a content question (one to be answered with *yes* and one with *no*). The practice items did not contain homophonous verb forms and were not counted as a part of the 320 items used in the experiment.

The main experiment consisted of three experimental blocks. The order in which the items were presented in each list pair was pseudo-randomized using the program Mix (van Casteren & Davis, 2006). Each target item was followed by at least four fillers. No more than two items with questions could follow each other, and items with questions could be separated by maximally seven items. After a question, minimally one filler followed before a target item appeared. Additionally, the first block started with at least five fillers and the subsequent blocks with at least two fillers.

2.4 Procedure

The participants performed an eye-tracking task in the Centre for Language Studies Lab at Radboud University. The experiment was programmed in Experiment Builder (version 2.1.140) and the eye-tracking system used was Eyelink 1000, combined with a fixed desk mount with adjustable chin rest. Participants were seated in front of a PC monitor in a dimly lit, sound-proof booth. After they read the study information document and signed the consent form, the chair and chin rest were adjusted to the appropriate height, so that the participants were in a comfortable position and the camera was able to track their eyes. They were instructed not to move during the experiment as this would invalidate the calibration.

After the participants read the instruction screen, the first calibration and validation were performed, followed by the practice block. After completion of the practice block, participants had a final opportunity to ask questions before starting with the main experiment. In the breaks in between the three blocks of the main experiment, participants were allowed to move and it was checked whether they were still comfortable. At the beginning of each block, a new calibration and validation were performed. In total, the experiment took 35-55 minutes, depending on reading times, duration of the breaks, and the ease with which calibration and validation could be performed.

The stimuli were horizontally aligned to the left of the screen and vertically centered, such that all stimuli started at the same position on the screen, independent of their length. The left and right margins were 1.5 centimeters. The font used was Calibri, size 22. None of the stimuli exceeded one line on the screen. Before each stimulus was presented, a fixation dot appeared at the coordinates of the beginning of the stimuli. This fixation dot contained a fixation trigger, which ensured that the stimulus was only presented when a fixation on the dot of at least 80 milliseconds was registered. In this way it was ensured that participants were looking at the position of the beginning of the stimulus when it appeared. Participants were instructed to press the space bar when they finished reading the stimulus. After each five items, a drift correction was performed before the fixation trigger appeared. At each drift correction, a recalibration could be performed if necessary.

When an item was followed by a content question, the word *vraag* 'question' and the question itself were presented on the screen, such that the question itself was horizontally and vertically centered on the screen and the word *vraag* 'question' was presented on the line above. The x-key on the keyboard corresponded to 'no' and the period-key corresponded to 'yes'. This information was visible at the bottom of the screen for all questions, to prevent confusion. Additionally, 3D-foam stickers were applied to the corresponding keys, to make sure the participants would not lose track of these keys.

2.5 Data-analysis

Four participants were excluded from the analysis due to poor calibration and/or poor performance (<80% of content questions answered correctly, while the remaining participants all scored above 90%). I manually checked the fixation data and, if necessary, corrected them for drift using the program Eyelink Dataviewer (version 3.1.97). Only the grid of the IAs was visible during this process and not the words themselves. As the content of the sentence was invisible during the annotation, the fixation results remained independent of expectations based on the theory.

In the entire analysis, I only focused on IAs 2 (the subject), 3 (the verb form), and 4 and 5 (spillover regions). On these IAs, I removed fixations shorter than 50 ms (213 data points, 2.2%). To analyze the data, I used Mixed Effects Regression Analysis. I used four dependent variables, resulting in four separate analyses: Fixation Probability, Fixation Count, First Fixation Duration, and Total Fixation Duration.

In the analysis of the first dependent variable, *Fixation Probability*, I performed a logistic mixed-effects regression analysis on the presence/absence of fixations on all IAs (i.e., IAs 2-5). In the analyses of the other dependent variables I only took into account the data points where an IA actually contained a fixation. When a given participant did not fixate on a given IA, I defined that data point as missing. Outliers, defined as values of more than 2.5 SD below/above the grand mean of each dependent variable were removed as well.

In the second analysis, *Fixation Count*, I analyzed the number of fixations on all IAs (given that they contained at least one fixation). In the removal of outliers on this variable, 393 data points (4.1%) were deleted. After this procedure, *Fixation Count* only contained the values 1, 2, or 3. Due to the lack of a normal distribution, I decided to perform a logistic mixed-effects regression analysis testing whether there were one or multiple fixations on an IA, rather than a linear mixed-effects analysis.

In the third analysis, *First Fixation Duration*, I performed a linear mixed effects regression analysis on the duration of the first fixation on each IA. To make the data approximately normally distributed, I log-transformed the values and subsequently removed outlying data points (175, 1.8%). The same procedure was used in the analysis of the last dependent variable, *Total Fixation Duration*: I performed a log-transformation on the data and removed the outliers (230 data points, 2.4%) and again performed a linear mixed effects regression analysis.

In all models, I tested four fixed factors and interactions between them: Interest Area (with the levels *IA2*, *IA3*, *IA4*, and *IA5*), Correctness (with the levels *correct* and *incorrect*), Correct Suffix (with the levels <*d*> and <*dt*>), and Relative Frequency (continuous). I also tested three random intercepts: *participant* was included to control for individual variation, *verb* was used to control for the variation between the items, and *current word* was used to control for the word corresponding to each IA. Furthermore, I tried to include random slopes for correctness by participant to control for individual variation in the reaction to incorrectly spelled verb forms, and for relative frequency by participant to account for individual variation in the effect of relative frequency, but these random slopes caused problems in converging the models. The final models only included fixed factors and interaction effects with p-values below .05 and random effects that significantly improved the model fit, based on likelihood ratio tests at the .05 α-level. When the fixed and random effects of the models were established, I additionally removed outliers from the lmer-models by deleting all data points with absolute standardized residuals exceeding 2.5 standard deviations and refitting the final models.

3 Results

3.1 Analysis 1: Fixation Probability

The first analysis was a logistic mixed-effects regression analysis of *Fixation Probability*, testing the presence/absence of fixations in the full dataset (i.e., Interest Areas 2-5). Table 4 presents the final model in an analysis of deviance table, produced by the Anova function from the Car package (Fox & Weisberg, 2011) for R (R Core Team, 2017). I found fixed effects of Interest Area and Correctness, as well as several interactions, including two three-way interactions: Interest Area * Correctness * Correct Suffix and Interest Area * Correct Suffix * Relative Frequency. To further investigate these effects, I performed additional analyses on subsets of the data split by Interest Area.

Table 4: Analysis of Deviance table (Type II Wald chi-square tests) for the fixed effects in the final overall model of *Fixation Probability*, predicting the presence/absence of fixations on each Interest Area. The standard deviation for the random intercept of *verb* was estimated at 0.282, that for the *participant* random intercept at 0.464, and that for *current word* at 0.805.

Fixed effects	χ^2	Df	p
Interest Area	1697.57	3	$<.001$
Correctness	7.62	1	$<.01$
Correct Suffix	0.34	1	$>.05$
Relative Frequency	1.72	1	$>.05$
Interest Area * Correctness	8.57	3	$<.05$
Interest Area * Correct Suffix	87.14	3	$<.001$
Correctness * Correct Suffix	0.47	1	$>.05$
Interest Area * Relative Frequency	2.31	3	$>.05$
Correct Suffix * Relative Frequency	1.44	1	$>.05$
Interest Area * Correctness * Correct Suffix	30.40	3	$<.001$
Interest Area * Correct Suffix * Relative Frequency	18.30	3	$<.001$

First, I analyzed the subset of the data consisting of Interest Area 2, the grammatical subject of the verb form. After initial inspection of the data in this subset, it is worth to remark that the data suggest that fixations on the grammatical subject are often the result of a regression (i.e., there has already been a fixation on a further point in the sentence). Table 5 shows the final model for the presence/absence of a fixation on Interest Area 2. The main effect of Correctness shows that, when a verb form is incorrectly spelled, the probability that people fixate on the grammatical subject preceding the verb form becomes larger. The main effect of Correct Suffix shows that, when the correct suffix is <dt>, the probability of a fixation becomes larger as well, regardless of whether the verb form was correctly spelled or not.

Table 6 shows the final model for the presence/absence of a fixation on Interest Area 3 (the verb form). The main effects of Correctness and Correct Suffix and their interaction show that the effect of Correct Suffix differs for correctly and incorrectly spelled verb forms. In order to investigate this interaction in more detail, I split the data of Interest Area 3 by correctness. A separate analysis of correctly and incorrectly spelled verb forms showed that the direction of the effect of Correct Suffix is opposite for correctly and incorrectly spelled verb forms. For correctly spelled verb forms, the probability of a fixation is higher when it

Table 5: Statistical model for the presence/absence of a fixation on Interest Area 2. The intercept represents correctly-spelled verb forms spelled with <d>. The standard deviation for the random intercept of *verb* was estimated at 0.217 and that for the *participant* random intercept at 0.527. With the random intercept of *current word* the model failed to converge and therefore *current word* was not included in the model.

Fixed effects	β	z	p
Intercept	-0.51	-5.48	<.001
Correctness (incorrect)	0.30	2.60	<.01
Correct Suffix (<dt>)	0.42	4.87	<.001

is (correctly) spelled with <dt>, compared to when it is (correctly) spelled with <d> (β = 0.41, t = 2.94, p < .01). For incorrectly spelled verbs forms, in contrast, the probability of a fixation is lower when the form should end in <dt> but instead is spelled with <d> (β = -0.37, t = -2.34, p < .05), compared to when the form should end in <d> but is spelled with <dt>. This means that, although I found an interaction of Correct Suffix and Correctness, in both correctly and incorrectly spelled verb forms the probability of a fixation is higher when the form is *written* with <dt>, irrespective of the correct suffix.

Besides splitting the data of Interest Area 3 by Correctness, I additionally split the data by Correct Suffix to further investigate the interaction between Correctness and Correct Suffix. This revealed that the effect of Correctness is only found when the correct suffix is <d> (β = 0.68, t = 4.55, p < .001), but not when the correct suffix is <dt> (β = -0.09, t = -0.64, p > .05). This means that when <dt> is incorrectly written instead of <d>, the probability of a fixation is higher, but when <d> is incorrectly written instead of <dt>, no difference in fixation probability is found.

Table 6: Statistical model for the presence/absence of a fixation on Interest Area 3. The intercept represents correctly-spelled verb forms spelled with <d>. The standard deviation for the random intercept of *verb* was estimated at 0.855 and that for the *participant* random intercept at 0.683. *Current word* was not included as random intercept, as it was identical to the verb form of interest.

Fixed effects	β	z	p
Intercept	1.57	3.82	<.001
Correctness (incorrect)	1.19	3.64	<.001
Correct Suffix (<dt>)	1.08	2.01	<.05
Correctness (incorrect) * Correct Suffix (<dt>)	-1.34	-3.07	<.01

The results for Interest Area 4 (the word following the verb form) will not be reported, as no statistically significant effects were found. Table 7 shows the final model for the presence/absence of a fixation on Interest Area 5 (the second word after the verb form). The main effects of Correctness, Correct Suffix, and Relative Frequency and the interactions between Correct Suffix and both Correctness and Relative Frequency show that the effects of Correctness and Relative Frequency differ for verb forms with <d> and with <dt> as correct suffix. In order to investigate these interactions in more detail, I split the data of Interest Area 5 by Correct Suffix.

Table 7: Statistical model for the presence/absence of a fixation on Interest Area 3. The intercept represents correctly-spelled verb forms spelled with <d>. The standard deviation for the random intercept of *verb* was estimated at 0.232, that for the *participant* random intercept at 0.511, and that for *current word* at 0.845.

Fixed effects	β	z	p
Intercept	1.57	6.59	<.001
Correctness (incorrect)	-6.79	-3.31	<.001
Correct Suffix (<dt>)	-1.07	-3.56	<.001
Relative Frequency	-0.43	-3.09	<.01
Correctness (incorrect) * Correct Suffix (<dt>)	1.01	3.73	<.001
Correct Suffix (<dt>) * Relative Frequency	0.67	3.68	<.001

First, I analyzed the interaction between Correctness and Correct Suffix. When the verb form had <d> as correct suffix, the probability of a fixation on Interest Area 5 was higher when the verb form was correctly spelled (with <d>), compared to when it was incorrectly spelled (with <dt>) (β = 1.37, t = 2.02, p < .05). When the correct suffix was <dt>, the probability of a fixation on Interest Area 5 was higher when the verb form was incorrectly spelled (with <d>), compared to when the verb form was correctly spelled (with <dt>) (β = 0.51, t = 2.06, p < .05). This means that, although I found an interaction of Correct Suffix and Correctness, the probability that participants fixated on the verb form was larger when it was *written* with <d>, regardless of whether this was the correct suffix or not. Note that this is opposite to the effect found on Interest Area 3.

To investigate the interaction between Correct Suffix and Relative Frequency, I again performed separate analyses on subsets of the data of Interest Area 5 split by Correct Suffix. The analysis of verb forms with <d> as correct suffix showed that the probability of a fixation on Interest Area 5 was lower when the relative frequency of the verb gets higher (β = -0.60, t = 2.02, p < .05). For verbs with <dt> as correct suffix, the effect of Relative Frequency was reversed (β = 0.28, t = 2.22, p < .05), which means that when <dt> was the correct suffix, the probability of a fixation on Interest Area 5 was higher when the relative frequency of the verb gets higher.

3.2 Analysis 2: Fixation Count

The second analysis I performed was a logistic mixed-effects regression analysis of the number of fixations (one or multiple) per Interest Area. Again, I first performed statistical analyses on the full dataset (i.e., IAs 2-5, given that they contained a fixation). Table 8 presents the final model in an analysis of deviance table, produced by the Anova function from the Car package (Fox & Weisberg, 2011) for R. I found fixed effects of Interest Area, Correctness, and Relative Frequency, as well as interactions of Correct Suffix with both Correctness and Interest Area.

The main effect of Relative frequency shows that, independent of the Interest Area, there were in general fewer fixations when the relative frequency was higher (β = -0.04, z = -4.17, p < .001). The interaction effect of Correct Suffix and Correctness was independent of Interest Area as well. To investigate this interaction, I performed separate analyses of items with correctly spelled and incorrectly spelled verb forms. These analyses showed that the effect of Correct Suffix on correctly spelled verb forms (β = 0.09, z = 0.77, p > .05) was in the opposite direction of the effect of Correct Suffix on incorrectly spelled verb forms (β = -0.12, z = -0.93, p > .05). Although neither of the two effects reached significance in isolation, the interaction between them implies that on correctly spelled verb forms, there tend to be more fixations when the correct suffix is <dt>, compared to incorrectly spelled verb forms, where there tend to be more fixations when the correct suffix is <d>. This means that, irrespective of the *correct* suffix, items with verb forms *written* with <dt> receive more fixations, compared to items with verb forms written with <d>.

Table 8: Analysis of Deviance table (Type II Wald chi-square tests) for the fixed effects in the final overall model predicting the fixation count (1 or multiple) on each Interest Area. The standard deviation for the random intercept of *verb* was estimated at 0.368, that for the *participant* random intercept at 0.540, and that for *current word* at 0.434.

Fixed effects	χ^2	Df	p
Interest Area	525.37	3	<.001
Correctness	29.81	1	<.001
Relative Frequency	17.36	1	<.001
Correct Suffix	0.07	1	>.05
Interest Area * Correct Suffix	32.45	3	<.001
Correctness * Correct Suffix	3.85	1	<.05

To further explore the two-way interaction between Interest Area and Correct Suffix, I performed additional analyses on subsets of the data split by Interest Area. Table 9 shows the final model for Fixation Count on Interest Area 2, the grammatical subject of the verb form. The main effect of Correct Suffix shows that the fixation count is higher when the correct suffix is <dt>, regardless of whether the verb form was correctly spelled or not.

Table 9: Statistical model for Fixation Count on Interest Area 2. The intercept represents verb forms ending in <d>. The standard deviation for the random intercept of *verb* was estimated at 0.376 and that for the *participant* random intercept at 0.676. With the random intercept of *current word* the model failed to converge and therefore *current word* was not included in the model.

Fixed effects	β	z	p
Intercept	-2.53	-11.17	<.001
Correct Suffix (<dt>)	0.72	2.88	<.01

Table 10 shows the final model for Fixation Count on Interest Area 3, the verb form. The main effects of Correctness and Correct Suffix and the interaction between them show that the effect of Correct Suffix differs for correctly spelled and incorrectly spelled verb forms. In order to investigate this interaction in more detail, I additionally split the data of Interest Area 3 by Correctness. A separate analysis of correctly spelled verb forms showed a statistically significant effect of Correct Suffix (β = 0.25, z = 2.37, p < .05). This means that the number of fixations on correctly spelled verb forms was higher when the verb form ended in <dt>, than when the verb form ended in <d>. For incorrectly spelled verb forms, the effect of Correct Suffix was not statistically significant (β = -0.18, z = -0.95, p > .05).

Table 10: Statistical model for Fixation Count on Interest Area 3. The intercept represents correctly-spelled verb forms ending in <d>. The standard deviation for the random intercept of *verb* was estimated at 0.635 and that for the *participant* random intercept at 0.634. *Current word* was not included as random intercept, as it was identical to the verb form of interest.

Fixed effects	β	z	p
Intercept	-0.63	-3.55	<.01
Correctness (incorrect)	0.64	4.02	<.001
Correct Suffix (<dt>)	0.25	2.38	<.05
Correctness (incorrect) * Correct Suffix (<dt>)	-0.43	-2.88	<.01

The full models of Interest Areas 4 and 5 will not be reported, as no statistically significant effects were found.

3.3 Analysis 3: First Fixation Duration

The third analysis was that of the duration of the first fixation on each Interest Area. As in the first two analyses, I first performed a statistical analysis on all Interest Areas taken together (i.e., IAs 2-5, given that they contained a fixation). Table 11 presents the final model in an analysis of deviance table, produced by the Anova function from the Car package (Fox & Weisberg, 2011) for R. I found a fixed effect of Interest Area, as well as a three-way interaction between Interest Area, Correctness, and Relative Frequency. To further explore this interaction, I performed additional analyses on subsets of the data split by Interest Area.

Table 11: Analysis of Deviance table (Type II Wald chi-square tests) for the fixed effects in the final overall model predicting the duration of the first fixation on each Interest Area. The standard deviation for the random intercept of *verb* was estimated at 0.011, that for the *participant* random intercept at 0.093, and that for *current word* at 0.023. The residual standard deviation was 0.309.

Fixed effects	χ^2	Df	p
Interest Area	23.04	3	<.001
Correctness	0.16	1	>.05
Relative Frequency	0.03	1	>.05
Interest Area * Correctness	3.23	3	>.05
Interest Area * Relative Frequency	0.43	3	>.05
Correctness * Relative Frequency	0.17	1	>.05
Interest Area * Correctness * Relative Frequency	8.50	3	<.05

For Interest Area 3 (the verb form), *current word* was not included as random intercept as it was identical to the verb form of interest. No statistically significant effects were found on Interest Area 2 and 3, and therefore the corresponding final models will not be reported.

Table 12 shows the final model for the first fixation durations on Interest Area 4 (the word following the verb form). The interaction effect of Correctness and Relative Frequency shows that the effect of Relative Frequency differs for correctly and incorrectly spelled verb forms. In order to investigate this interaction in more detail, I split the data of Interest Area 4 by Correctness. A separate analysis of correctly spelled verb forms showed that the effect of Relative Frequency was not statistically significant (β = -0.02, t = -0.85, p > .05). For incorrectly spelled verb forms, the effect of Relative Frequency was statistically significant (β = 0.13, t = 2.04, p < .05), which means that the duration of the first fixation on a word following an incorrectly spelled verb form increases when the relative frequency of the verb form is higher.

Table 12: Statistical model for the duration of the first fixation on Interest Area 4. The intercept represents correctly-spelled verb forms. The standard deviation for the random intercept of *verb* was estimated at 0.001, that for the *participant* random intercept at 0.081, and that for *current word* at 0.028. The residual standard deviation was 0.321.

Fixed effects	β	t	p
Intercept	5.33	249.54	<.001
Correctness (incorrect)	-0.02	-0.42	>.05
Relative Frequency	-0.01	-0.44	>.05
Correctness (incorrect) * Relative Frequency	0.14	2.00	<.05

The final model of Interest Area 5 (the second word after the verb form) will not be reported as it revealed no statistically significant effects.

3.4 Analysis 4: Total Fixation Duration

The last analysis I performed was that of Total Fixation Duration. Again, I first performed a statistical analysis on the full dataset. Table 13 presents the final model in an analysis of deviance table, produced by the Anova function from the Car package (Fox & Weisberg, 2011) for R. I found fixed effects of Interest Area and Correctness, as well as several interactions, including three-way interactions between Interest Area, Correctness, and Correct suffix, and between Interest Area, Correctness, and Relative Frequency. To further explore these effects, I again performed additional analyses on subsets of the data split by Interest Area.

Table 13: Analysis of Deviance table (Type II Wald chi-square tests) for the fixed effects in the final overall model predicting the total fixation duration on each Interest Area. The standard deviation for the random intercept of *verb* was estimated at 0.102, that for the *participant* random intercept at 0.148, and that for *current word* at 0.089. The residual standard deviation was 0.469.

Fixed effects	χ^2	Df	p
Interest Area	326.84	2	<.001
Correctness	11.27	1	<.001
Correct Suffix	1.03	1	>.05
Relative Frequency	0.71	1	>.05
Interest Area * Correctness	2.66	3	>.05
Interest Area * Correct Suffix	12.94	3	<.01
Interest Area * Relative Frequency	2.04	3	>.05
Correctness * Correct Suffix	6.67	1	<.01
Correctness * Relative Frequency	1.80	1	>.05
Interest Area * Correctness * Correct Suffix	8.30	3	<.05
Interest Area * Correctness * Relative Frequency	7.56	3	<.05

Table 14 shows the final model for the total fixation durations on Interest Area 2, the grammatical subject of the verb form. The main effect of Correct Suffix shows that the total fixation duration on the subject is longer when the correct suffix of the verb form is <dt>, regardless of whether the verb form is correctly spelled or not. The effects of Correctness and Relative Frequency and the interaction between them show that the effect of Relative

Frequency differs for correctly and incorrectly spelled verb forms. In order to investigate this interaction in more detail, I split the data of Interest Area 2 by correctness. A separate analysis of correctly spelled verb forms revealed no statistically significant effects of Relative Frequency (β = 0.02, t = 0.87, p > .05). For incorrectly spelled verb forms, however, the effect of Relative Frequency was statistically significant (β = -0.20, t = -2.34, p < 0.05). This means that the total fixation duration on grammatical subjects followed by an incorrectly spelled verb form is shorter when the relative frequency of this verb form is higher. Note that this effect is reversed to the interaction effect of Correctness and Relative Frequency on the word following the verb form I found in the analysis of the first fixation duration.

Table 14: Statistical model for total fixation duration on Interest Area 2. The intercept represents correctly-spelled verb forms ending in <d>. The standard deviation for the random intercept of *verb* was estimated at 0.303, that for the *participant* random intercept at 0.137, and that for *current word* at 0.001. The residual standard deviation was 0.400.

Fixed effects	β	t	p
Intercept	5.38	104.30	<.001
Correct Suffix (<dt>)	0.10	3.41	<.001
Correctness (incorrect)	0.10	1.69	>.05
Relative Frequency	0.01	0.50	>.05
Correctness (incorrect) * Relative Frequency	-0.21	-2.32	<.05

Table 15 shows the final model for the total fixation durations on Interest Area 3, the verb form. The main effect of Correctness and its interaction with Correct Suffix show that the effect of Correctness differs for verb forms with <d> as correct suffix and verb forms with <dt> as correct suffix. In order to investigate this interaction in more detail, I split the data of Interest Area 3 by Correctness. A separate analysis of correctly spelled verb forms showed no statistically significant effect of Correct Suffix (β = 0.04, t = 0.84, p > .05). For incorrectly spelled verb forms, however, the effect of Correct Suffix was statistically significant (β = -0.09, t = -2.07, p < .05). This means that the total fixation duration on incorrectly spelled verb forms was shorter when the verb was written with a <d> (instead of the correct <dt>), than when the verb was written with <dt> (instead of the correct <d>).

Table 15: Statistical model for Total Fixation Duration on Interest Area 3. The intercept represents correctly-spelled verb forms ending in <d>. The standard deviation for the random intercept of *verb* was estimated at 0.143 and that for the *participant* random intercept at 0.175. *Current word* was not included as random intercept, as it was identical to the verb form of interest. The residual standard deviation was 0.465.

Fixed effects	β	t	p
Intercept	5.64	81.98	<.001
Correctness (incorrect)	0.28	3.24	<.01
Correct Suffix (<dt>)	0.03	0.77	>.05
Correctness (incorrect) * Correct Suffix (<dt>)	-0.13	-3.94	<.001

Table 16 shows the final model for the total fixation durations on Interest Area 4, the word following the verb form. The main effect of Correctness shows that the total fixation duration is longer when the verb form is incorrectly spelled. The main effect of Correct Suffix shows that the total fixation duration is shorter when the correct suffix is <dt>, regardless of whether the verb form was correctly spelled or not. This means that the effect of Correct Suffix on Interest Area 4 is reversed with respect to the effect I found on Interest Area 2.

Table 16: Statistical model for total fixation duration on Interest Area 4. The intercept represents correctly-spelled verb forms ending in <d>. The standard deviation for the random intercept of *verb* was estimated at 0.074, that for the *participant* random intercept at 0.136, and that for *current word* at 0.065. The residual standard deviation was 0.466.

Fixed effects	β	t	p
Intercept	5.50	87.26	<.001
Correctness (incorrect)	0.09	1.20	<.01
Correct Suffix (<dt>)	-0.08	-2.52	<.05

The final model for Interest Area 5 will not be reported as no statistically significant effects were found. This section will be concluded with an overview of all experimental results (see Table 17).

Table 17: Summary of the experimental results per Interest Area

<table>
<tr><th></th><th>IA2</th><th>IA3</th><th>IA4</th><th>IA5</th></tr>
<tr><td>FP[3]</td><td>**Correctness**: Higher fixation probability when verb is incorrectly spelled

Correct Suffix: Higher fixation probability when correct suffix is <dt></td><td>**Correctness * Correct Suffix**:
-Correctly spelled verbs: Higher fixation probability when correct suffix is <dt>
-Incorrectly spelled verbs: Higher fixation probability when correct suffix is <d></td><td>-</td><td>**Correct Suffix * Correctness**:
-<d> as correct suffix: Higher fixation probability when verb is correctly spelled
-<dt> as correct suffix: Higher fixation probability when verb is incorrectly spelled

Correct Suffix * Relative Frequency:
-<d> as correct suffix: Higher fixation probability when relative frequency is lower
-<dt> as correct suffix: Higher fixation probability when relative frequency is higher</td></tr>
<tr><td rowspan="2">FC</td><td colspan="4">Independent of IA:
More fixations when **Relative Frequency** is higher;
Correct Suffix * Correctness: More fixations when correctly spelled verbs have <dt> as correct suffix; more fixations when incorrectly spelled verbs have <d> as correct suffix</td></tr>
<tr><td>**Correct Suffix**: More fixations when correct suffix is <dt></td><td>**Correctness * Correct Suffix**:
-Correctly spelled verbs: More fixations when correct suffix is <dt>
-Incorrectly spelled verbs: No effect</td><td>-</td><td>-</td></tr>
<tr><td>FFD</td><td>-</td><td>-</td><td>**Correctness * Relative Frequency**
-Correctly spelled verbs: No effect
-Incorrectly spelled verbs: Longer first fixation duration when relative frequency is higher</td><td>-</td></tr>
<tr><td>TFD</td><td>**Correct Suffix**: Longer total fixation duration when correct suffix is <dt>

Correctness * Relative Frequency:
-Correctly spelled verbs: No effect
-Incorrectly spelled verbs: Longer total fixation duration when relative frequency is lower</td><td>**Correctness * Correct Suffix**:
-Correctly spelled verbs: No effect
-Incorrectly spelled verbs: Longer total fixation duration when correct suffix is <d> (written: <dt>)</td><td>**Correctness**: Longer total fixation duration on incorrectly spelled verbs

Correct Suffix: Longer total fixation duration when correct suffix is <d></td><td>-</td></tr>
</table>

[3] FP: Fixation Probability, FC: Fixation Count, FFD: First Fixation Duration, TFD: Total Fixation Duration

4 Discussion

In this section, I will first discuss the results per Interest Area in detail, before ending with a general discussion of the results and their implications.

4.1 Discussion of the results per Interest Area

4.1.1 Interest Area 2 (the grammatical subject preceding the verb form)

The first Interest Area I investigated was the grammatical subject preceding the verb form. Already here, I found an effect of the correctness of the verb form on the probability that a fixation occurred, namely that this probability was larger when the verb form was incorrectly spelled. As I already mentioned in the Results section, fixations on the subject are often the result of regressions, which means that the subject is sometimes initially skipped. As I discussed in the Introduction, during a fixation on a word, people already start processing information from the upcoming word(s), which are either in foveal or parafoveal vision (e.g., Rayner, 1998). Especially short and frequent words are processed quickly via this (para)foveal preview benefit (Engbert et al., 2002; Veldre & Andrews, 2018). As the grammatical subjects in this experiment were all very short and highly frequent words (namely, personal pronouns), it is likely that people already started processing the information from the subject without actually fixating on it. When they later on in processing the sentence register a mismatch between the spelling of the verb form and the grammatical information from the subject, however, they are more likely to fixate on the subject to check whether their grammatical analysis was correct. This explains the finding that the probability of a fixation on the subject is higher when the verb form was incorrectly spelled.

The interaction effect of Correctness and Relative Frequency on the total fixation duration on the subject shows that the effect of Relative Frequency differs for correctly and incorrectly spelled verb forms. For correctly spelled verb forms, I found no effect of Relative Frequency. Apparently, processing the subject of correctly spelled verb forms was not facilitated by a relatively high frequency of the verb form itself. However, for incorrectly spelled verb forms I found an effect of Relative Frequency in the expected direction: When the relative frequency of the written verb form versus its homophone counterpart was larger, the total fixation duration on the subject was smaller. Apparently, participants pay more attention to (grammatical information from) the subject when an incorrectly spelled verb form is less frequent. There are several possible explanations for this result. First, as we already saw, participants tend to fixate on the subject more often when there is a mismatch with the grammatical information from the verb form. It appears that these fixations become longer when the verb form is a low-frequent form. While a highly frequent verb form is more easily accepted and processed faster, this is the other way around for verb forms with a lower frequency. The low frequency does not inherently lead to a preference for the verb form as it was written, and as a consequence, the form cannot be immediately accepted based on the frequency information. As a result, it is more likely that participants become aware of the mismatch between the subject and the verb form, which explains the increase in total fixation duration on the subject for low-frequent, incorrectly spelled verb forms. A second explanation has to do with (para)foveal preview benefit effects again. Irrespective of whether a fixation on the subject is the result of a regression or not, it is likely that participants were able to see the verb form as well when they fixated on the subject, especially in the case of short verb forms (e.g., *vind* 'find', *red* 'rescue').

As a result, the total fixation duration on the subject increases when the verb form is both incorrect and low-frequent. A third, related explanation is that retrieval of multi-word units is likely to play a role as well. The subject and the verb form can be seen as a single unit. The more frequent such a unit consisting of multiple words is, the stronger it will be represented in the mental lexicon and the faster it will be processed (e.g., Arnon & Snider, 2010; Sprenger et al., 2006). This does not only hold for grammatically correct word combinations: When an ungrammatical combination is perceived frequently, it is processed faster as well, and is possibly even represented in the mental lexicon. This means that, when participants have often seen an ungrammatical combination of a certain subject and verb form, they will process it faster. When the ungrammatical combination is relatively new to them, the processing will be slower. This explains why the combination of a subject with a relatively infrequent verb form receives more attention than when the verb form is highly frequent. Note, additionally, that the increased fixation duration in this case not necessarily means that people are aware of the ungrammaticality of the combination.

The duration of the first fixation is believed to reflect factors that have a strong impact on the process of early lexical access (Rayner, 1998; Rayner & Duffy, 1986), whereas the total fixation duration reflects a more advanced processing stage, where the word is integrated into the rest of the sentence (Staub et al., 2010). The fact that, on the subject, the interaction of correctness and relative frequency only influences the total fixation duration, thus implies that this effect reflects a more advanced stage of processing, rather than a process of early lexical access of the subject. This is compatible with the explanations given above, which all involve the use of information from both the subject and the verb form. In the stage of processing where this effect is found, the accessed lexical form of the subject has to be integrated with the grammatical information from the verb form. Especially for lower frequent verb forms, a mismatch between the grammatical properties of the verb form and the subject may be noticed in this stage, explaining the longer fixation durations for incorrectly spelled, low-frequent verb forms, whereas a mismatch may more often go unnoticed for highly frequent verb forms, at least in this stage of processing.

In the analysis of Interest Area 2, I also found effects of Correct Suffix. In general, the probability that participants fixated on the subject was larger when the verb form had <dt> as correct suffix, compared to when the correct suffix was <d>. The number of fixations was larger as well, as was the total fixation duration. Again, several explanations can be given for these findings. In the current experiment, <d> corresponds to first person singular verb forms, while <dt> corresponds to third person singular verb forms. First person singular verb forms can only have one subject form, namely *ik* 'I'. For third person singular verb forms, there is much more variation in possible subject forms, such as *hij* 'he', *zij* 'she', *het* 'it', *dit* 'this', *er* 'there', proper names, and so on. As a result of this large variation, it takes more time to process third person singular subjects. In first person singular verb forms, in contrast, there is only one possible subject form, which makes this form easier to process. The link between what the correct suffix of the verb form would be and which possible subjects correspond to this suffix, can explain why I found shorter and fewer fixations on subjects corresponding to verb forms with <d> as correct suffix (first person singular, only one possible subject), compared to subjects corresponding to verb forms with <dt> as correct suffix (third person singular, many possible subjects).

Another explanation that can account for the increased attention to subjects preceding verb forms that have to be spelled with <dt>, could be that <dt> as correct suffix is more difficult to process compared to <d>. As Ernestus and Mak (2005) explain, <d> is much more frequent in the inflectional paradigm, compared to other word-final segments. Based on this explanation, it can be assumed that <dt> is more salient and attracts more attention, and it is relatively more difficult and time-consuming to determine that <dt> is the correct suffix (compared to <d>). As a result, when <d> is written while <dt> would have been correct, the error is more often overlooked, while errors are more salient when <dt> is written while <d> would have been correct. In other words, <d> is more easily accepted, even when it is not the correct suffix. This is in accordance with the preference for analogy to <d> in the entire inflectional verb paradigm (Ernestus & Mak, 2005).

4.1.2 Interest Area 3: The verb form

The next Interest Area I analyzed was the verb form itself. Here, I found interaction effects of Correctness and Correct Suffix on both the probability of a fixation, the fixation count, and the total fixation duration. These interactions show that the effect of Correct Suffix differed for correctly spelled and incorrectly spelled verb forms. When the verb form was correctly spelled, the probability that participants fixated on it was higher when the form was spelled with <dt>, compared to when it was spelled with <d>. However, when the verb form was incorrectly spelled, this effect was reversed: When the verb form should be spelled with <dt> (but was incorrectly spelled with <d>), I found *fewer* fixations, compared to when the form should be spelled with <d> (and was incorrectly spelled with <dt>). This means that, although the results show an interaction of Correct Suffix and Correctness, the effect of the *written* suffix is independent of the correctness of the verb: The probability that participants fixated on the verb form was larger when it was *written* with <dt>, regardless of whether this was the correct spelling or not. Apparently, verb forms written with <dt> attract more attention than verb forms written with <d>, which again implies that the presence of <dt> is more salient than the 'default' <d>.

For the number of fixations on the verb form, I found an interaction between Correctness and Correct Suffix as well. Again, when the verb form was spelled correctly, I found a larger number of fixations when the verb was (correctly) spelled with <dt>, compared to when it was (correctly) spelled with <d>. However, when the verb form was incorrectly spelled, I did not find an effect of Correct Suffix, meaning that the number of fixations on incorrectly spelled verb forms did not depend on the (correct or written) suffix.

For the total fixation duration, I again found an interaction between Correctness and Correct Suffix. There was no effect of correct suffix on correctly spelled verb forms, which means that, when a verb form was correctly spelled, the total time people fixated on it did not depend on the suffix. However, when the verb form should end in <d> and was incorrectly spelled with <dt>, the total fixation duration was longer than when the verb form should end in <dt> and was incorrectly spelled with <d>. In general, therefore, the results show that forms written with <dt> attract attention initially, leading to a larger probability that participants will fixate on it. When the verb form is incorrectly spelled, the subsequent number of fixations does not depend on the spelling of the verb form, but the total fixation duration increases when the verb form is incorrectly written with <dt>, which means that the fixations are longer on average in that case.

Several things can be concluded from these interaction effects of Correct Suffix and Correctness. First, the results support previous studies which found that there is a preference for

<d> in writing homophonous verb forms (e.g., Bosman, 2005; Frisson & Sandra, 2002; Sandra et al., 1999). As I hypothesized, forms that are produced more easily, are processed more easily as well. It appears that, indeed, the <d>-preference found in production exists for reading as well. The fact that verb forms written with <d> receive fewer and shorter fixations than verb forms written with <dt>, suggests that processing verb forms with <d> is faster and/or easier. The preference for this form might have people ‘accept’ the spelling more easily than when the form is written with <dt>. Related to this explanation, it can also be assumed that a mismatch between the subject and the verb form is registered more often and/or more quickly when <dt> is incorrectly used than when <d> is incorrectly used. As forms written with <dt> receive more attention in general, and the <d>-form can be seen as the ‘default’ form and therefore tends to be accepted easier, an erroneously written <dt> is more salient than an erroneously written <d>, leading to longer fixations in case of an erroneously written <dt>.

4.1.3 Interest Area 4: The word following the verb form

Since eye-tracking data often show spillover effects (e.g., Bertram, Hyönä, et al., 2000; Witzel et al., 2012), I also analyzed the first and second word following the verb form. The first word following the verb form is the first Interest Area where I found an effect on the duration of the First Fixation, namely an interaction effect of Correctness and Relative Frequency. When the verb form was correctly spelled, there was no effect of Relative Frequency, but when the verb form was incorrectly spelled, the effect of relative frequency was opposite to the effect I found on Interest Area 2: When the written verb form was more frequent compared to its homophone counterpart, the duration of the first fixation on the word following the verb form became longer. Previous research (e.g., Assink, 1985; Verhaert, 2016) showed that more errors are made with highly frequent verb forms and I hypothesized that, consequently, these errors are more often overlooked as well. When the relative frequency of a verb form is low, in contrast, fewer errors are made and it is more likely that participants become aware of an error (and that this happens in an earlier stage of processing). This explains the direction of the effect of Relative Frequency when participants read the word following the verb form. A higher relative frequency makes that the verb form is processed faster. As a consequence of this high frequency, it is more likely that an error is not immediately registered while processing the verb form itself. Later on, when participants become aware of the fact that the highly frequent form they have read forms a mismatch with the subject of the sentence, this leads to a delay in reading the following word. When the verb form had a lower frequency, in contrast, it is more likely that the error is already registered in an earlier stage, which makes the processing delay on the following word smaller.

The fact that the effect of Relative Frequency is manifested on the duration of the first fixation on Interest Area 4, is likely due to an overlap of a more advanced processing stage of the verb form (Interest Area 3) with an early processing stage of the following word (Interest Area 4). Participants already fixate on the next word before the processing of the verb form is entirely finished, and as a result, the duration of the first fixation on the next word is influenced by the remaining processing costs for the verb form. When these costs are higher (as is the case when participants become aware that the highly frequent form mismatches the grammatical information from the subject), the delay on the following word is consequently larger.

In general, the total fixation duration on the word following the verb form was longer when the verb form was incorrectly spelled, which means that an incorrectly spelled verb form not only causes a delay on the verb form itself, but also at the following word. Additionally, I found an effect of Correct Suffix on the total fixation duration, but opposite to the effects I found on Interest Area 2: The total fixation duration on the word following the verb form increased when the correct suffix of the verb form was <d>, compared to when it was <dt>. The reversal of this effect, compared to Interest Area 2, can be explained as a spillover effect. As the effect of Correct Suffix on the total fixation duration on Interest Area 4 is independent of the correctness of the verb form, there are two possible scenarios in interpreting this result. In the first scenario, the verb form was incorrectly spelled. This would mean that, when the <d>-form would have been correct, the verb form was written with <dt> and vice versa. In this case, the effect means that the total fixation duration was longer when the verb form was written with <dt>, compared to when it was written with <d>. This is the same outcome as we already saw for Interest Area 3 and can be explained by the general preference for verb forms written with <d> and the salience of verb forms written with <dt>, combined with longer fixation durations when the verb form is incorrectly spelled.

In the second scenario, the verb form was correctly spelled, which means that the word following verb forms (correctly) written with <d> received more attention than the word following verb forms (correctly) written with <dt>. In this case, another explanation is needed for the longer fixation durations when <d> was the correct suffix. As we already saw, initially forms with <d> receive less attention, indicating that these forms are initially accepted more easily. It has been demonstrated that forms that become available (too) quickly, can in some cases be labeled as 'suspicious' and therefore be rejected in a second stage (e.g., Balota, Law, & Zevin, 2000; Finkbeiner, Almeida, Janssen, & Caramazza, 2006). In the current experiment, this could have happened both for highly frequent forms, and for forms ending in <d>. In both cases, the processing is initially speeded up because the form is quickly available. As a result of this processing advantage, participants then sometimes unconsciously distrust the quickly available form and tend to rely on the rule-based outcome, which involves additional processing. This is relatively time-consuming and explains the delay after reading highly frequent verb forms and forms ending in <d>. In this way, effects that are initially facilitating in the reading process, may in a second stage become inhibitory.

4.1.4 Interest Area 5: The second word following the verb form

The final Interest Area I investigated, Interest Area 5, was the second word after the verb form. Here I only found effects on the probability of a fixation, but not on fixation count or duration. An interaction between Correct Suffix and Correctness showed that the effect of Correctness differed for forms with <d> and <dt> as correct suffix. When the form should end in <d>, I found more fixations when the verb form was correctly spelled (with <d>), compared to when it was incorrectly spelled (with <dt>). When the <dt>-form was correct, I found more fixations when the verb form was incorrectly spelled (with <d>), compared to when it was correctly spelled (with <dt>). This means that, although the interaction effect with Correctness depends on the *correct* suffix, the probability of a fixation on Interest Area 5 was higher when the verb form was *written* with <d>, regardless of its correctness. This is opposite to the effects at Interest Area 3 (the verb form itself), where I found more and longer fixations when the verb form was written with <dt>. This means that, initially, verb forms written with <d> are more often skipped and verb forms written with <dt> are more often fixated on. Only later on, the probability of a fixation becomes higher when the verb form was written with <d>. Again, this can be explained by the fact that verb forms written with <d> are initially accepted more easily, but are sometimes labelled as 'suspicious' due to their quick availability. The effect then turns into an inhibitory effect, explaining the delay in reading the following words.

The last effect I found on Interest Area 5 is an interaction between Correct Suffix and Relative Frequency, which means that the effect of Relative Frequency differed for the two suffixes. When the form should end in <d>, the probability of a fixation was higher when the verb form was less frequent. This means that processing verb forms that have to be spelled with <d> is easier when the verb form is more frequent, and that the processing becomes more difficult and time-consuming when the verb form is less frequent, resulting in a delay in reading the following words. Conform the results of Interest Area 4, the <d>-form might have been rejected because it was labelled as suspicious due to its quick availability. However, when it appears that the <d>-form would have been correct after all, participants may become aware of this more quickly when the form is more frequent, as a lower frequency might give them 'confirmation' that their rejection of this form was justified.

When the correct suffix was <dt>, I found the reverse: The probability of a fixation on Interest Area 5 was higher when the verb form was *more* frequent. This can be accounted for by the following explanation. At the very beginning of the processing, people have a <d>-preference. As we have seen, this effect reverses in a later stage, leading to a delay in processing verb forms ending in <d>, compared to verb forms ending in <dt>. Now, the <dt>-form is in advantage. However, when that form is the most frequent form, the problem is that it might have been labelled as suspicious as well in this stage of the processing, while the lower frequent form does not have this label. As a consequence, the processing delay is smaller when the verb form has to be spelled with <dt> and has a lower frequency, and the delay is larger when the frequency is higher. An alternative explanation is that, when the form ending in <dt> is the more frequent form of the two homophones, this leads to a conflict with the preference for forms ending in <d> in an earlier stage of processing. This causes a processing delay on the following words, which explains why I found more fixations after highly frequent forms with <dt> as correct suffix, compared to <dt>-forms with a lower frequency, where the conflict with the <d>-preference is smaller.

4.2 General discussion

In this study, I aimed to investigate the cognitive processes underlying the perception of homophones during reading. To this purpose, I investigated the effect of relative frequency and suffix on how people read (in)correctly spelled homophonous verb forms occurring in Dutch everyday language behavior. I carried out an eye-tracking experiment with tweets, containing a homophonous verb form that was correctly spelled in one condition and incorrectly spelled in the other condition. This study provided a good opportunity to investigate whether cognitive processes underlying spelling production also underlie perception during reading. Whereas Schmitz et al. (2018-in press) conducted a corpus study on tweets containing homophonous verb forms, I investigated whether factors important for spelling production are important in reading tweets containing homophonous verb forms as well. The fact that tweets were used in both studies enhances the ecological validity of the results, as tweets are spontaneously produced, real language data. I expected that homophonous verb forms that are more easily produced, are processed faster during reading as well. In terms of spelling errors, this would mean that errors that are made more often during spelling, are overlooked more often during reading. My results provide support for this hypothesis. Importantly, factors that led to a faster initial processing, caused a processing delay later on, especially when the verb form was incorrectly spelled.

The independent variables used in the experiment were *Interest Area* (i.e., the grammatical subject, the verb form, and the two words thereafter), *Correctness* (i.e., correct/incorrect), *Correct Suffix* (i.e., <d>/<dt>), and *Relative Frequency* (between the written verb form and its homophone counterpart). I tested four dependent variables: *Fixation Probability* (the probability that a given participant fixated on a given interest area), *Fixation Count* (the number of fixations on a given interest area, given that a participant fixated on it), *First Fixation Duration* (the duration of the first fixation of a given participant on a given interest area), and *Total Fixation Duration* (the sum of the durations of all fixations by a given participant on a given interest area).

In general, as I expected, sentences with incorrectly spelled verb forms received more attention than when the verb form was correctly spelled. The increase in fixations on both the subject and the verb form could be explained in several ways. The first option is that people indeed rely on grammatical rules in processing the grammatical properties of homophonous verb forms, and that they need (and use) information from the subject to be able to do so. When a mismatch between the subject and the grammatical properties of the verb form is registered, the reading process is delayed, resulting in more and longer fixations on both the subject and the verb form.

Another explanation that can account for the increased number and duration of fixations when the verb form was incorrectly spelled, is that the retrieval of multi-word units plays a role. It has been demonstrated that word combinations that are frequently used have a stronger representation in the mental lexicon, and that retrieval of an entire combination is easier when it is more frequent (see, e.g., Arnon & Snider, 2010; Sprenger et al., 2006). In the current experiment, the subject and the verb form were always adjacent, which makes it entirely possible that they are (at least in some cases) accessed as a unit instead of two separate forms. When the combination of subject and verb form occurs frequently in written language, it is more strongly represented in the mental lexicon, which means that it can be accessed more easily, resulting in shorter processing times. When the combination of subject and verb form is ungrammatical, however, the combination obviously has a much weaker (if any) representation in the mental lexicon, which makes it more difficult or even impossible to access the combination as a unit. Thus, when the verb form was incorrectly spelled, it is less likely that participants could access the combination of the subject and the verb form as a unit, and consequently, the time to process the combination increased, even if participants were not aware of the error.

It is likely that the increased attention to sentences with incorrectly spelled verb forms is due to a combination of the two explanations given above. If subject and verb form are initially processed as a unit and this unit appears to have a low frequency, this might unconsciously lead to suspicion of the grammaticality of this combination. As a result, it is more likely that people (also) rely on the grammatical rules and information from the subject and the verb form as independent words. In this way, the reading process can be delayed by both the low frequency of the incorrectly spelled combination as a whole, and by the mismatch that is more likely to be registered by subsequent application of the grammatical rules.

Another important finding supporting and complementing previous research is the initial preference for verb forms written with <d>, which later results into a delay. Many studies report a preference for <d> over other word-final segments. In writing, this means that people more often write <d> instead of <dt> or <t> than the other way around (e.g., Bosman, 2005; Sandra et al., 1999, Schmitz et al., 2018-in press). As I hypothesized, it appears that verb forms incorrectly written with <d> are (at least initially) more often overlooked in the reading process. In the current study, the <d>-preference comes to expression in multiple ways. First, on all interest areas taken together, we see a general increase in the number of fixations when the verb form was written with <dt>, even if this was the correct suffix. Second, the <d>-preference was also clearly visible at the verb form itself: It was more often and longer fixated on when it was written with <dt>. Third, irrespective of the suffix that was used in the written form, especially the subject of the verb form received more attention when the *correct* suffix was (or would have been) <dt>, compared to <d>. Fourth, on the words following the verb form, I found a delay when the verb form was written with <d>, even when this was the correct suffix.

The combination of these results suggests that, initially, forms with <d> are 'accepted' more easily, but later on cause a delay. One of the reasons for the initial preference for <d> is that it is much more frequent as word-final segment in the inflectional paradigm of verbs, compared to other word-final segments (Ernestus & Mak, 2005). In a later stage of processing, when the grammatical information of the subject and the verb form are integrated, participants may become aware of the mismatch in grammatical information, and that their preference for the <d>-form was unjustified. However, apparently, abandoning this preference is a costly process, which explains the processing delay.

Another important outcome of this study, showing how the cognitive stages of processing written language unfold over time, are the effects of the relative frequency between the written verb form and its homophone counterpart. In the main analysis with all Interest Areas together, I found a larger number of fixations when the verb form was less frequent, indicating that – as I expected – a higher frequency facilitates processing. More interesting, however, are the interactions with the correctness of the verb form I found on the separate Interest Areas, as the effect reverses during the reading process. On the grammatical subject, I found that incorrectly spelled verb forms with a higher frequency led to a shorter total fixation duration, which means that they were initially processed more quickly. A lower frequency, in contrast, allows for a quicker registration of the mismatch between the subject and the verb form, explaining the delay in processing in that case.

However, on the word following the verb form, this pattern reverses. The processing speed of words following incorrectly spelled verb forms is again influenced by the frequency of the verb form, but when the verb form is more frequent, it now results in *longer* fixations. This perfectly follows from my explanation of the previously discussed effect. When a verb form is higher in frequency, it is more likely to be accepted as the correct form initially. Only when the grammatical processing reaches a more advanced stage, participants may realize that the grammatical properties of the highly frequent form do not match the intended grammatical function, and this consequently leads to a processing delay.

Additionally, an interaction of Relative Frequency and Correct Suffix on the spillover region showed that a high frequency facilitated processing of the words following the verb form when the correct suffix was <d>, but that a high frequency was inhibitory when the correct suffix was <dt>. This shows that verb forms with <d> as correct suffix are easier to process, especially when they are highly frequent. Due to the general <d>-preference, it is easily accepted as the correct suffix, especially when the form is highly frequent. However, for verb forms with <dt> as correct suffix, the correct spelling is in conflict with the preferred <d>-spelling. Especially when the verb form is highly frequent, it is difficult to suppress both the <d>-preference and the frequency information. This explains the processing delay after reading highly frequent verb forms with <dt> as correct suffix.

The frequency-effects I found provide us with important insights into the online process of reading homophonous verb forms. However, while it could be expected that the frequency effect would directly influence the speed of lexical access (and, consequently, the duration of the first fixation on the accessed form), I did not find a main effect of Relative Frequency on the duration of the first fixation on the verb form. Verhaert (2016) performed a partially comparable eye-tracking experiment on spelling errors in homophonous verb forms and did not find a robust effect of Relative Frequency at all (at least, not in this specific study, but see the other experiments in Verhaert, 2016). One of the explanations she gives is that the sampling rate of 300Hz she used, could have been too low to measure an effect due to too fast saccades between different words. However, in the current experiment I used a sampling rate of 1000Hz, which is demonstrated to be more than sufficient for eye-tracking on a reading task (e.g., Holmqvist et al., 2011). Therefore, the sampling rate cannot explain the absence of more effects of Relative Frequency.

A more likely explanation for the fact that I did not find a main effect of Relative Frequency on the first fixation duration is that this is due to the verb forms I used in this experiment (to be found in the Appendix). In selecting the materials, I had to take into account many criteria to prevent a bias towards one of the two suffixes. This left me with a rather limited choice of suitable verbs, especially because I also had to find tweets where the selected verb form was incorrectly spelled and which additionally fitted the general structure of the experimental items. I had to keep the average relative frequency of the homophone pairs roughly constant over the different conditions of the materials, and as a result, the relative frequency in some homophone pairs was rather small. Furthermore, sometimes the relative frequency of a homophone pair might have been large enough, but the absolute frequencies of both verb forms were still too low to find an effect. For instance, when a given homophone pair has frequencies of 10 and 40, their relative frequency is identical to that of a homophone pair with 1000 and 4000 as frequencies, but the effect will be much smaller as both forms have a relatively weak representation in the mental lexicon. If I would have had more choice in verbs, I could have used (i) a larger range in the relative frequency of the homophone pairs, and (ii) verbs with a higher absolute frequency.

Alternatively, the experiment could be repeated as a proofreading task. That is, the participants would do the same experiment, but their task would not be reading the stimuli in the way they would normally read, but to actively look for errors. It has been shown that frequency effects are larger in proofreading than in reading for comprehension (Schotter, Bicknell, Howard, Levy, & Rayner, 2014). In reading for comprehension, spelling errors can generally be ignored, as long as they do not interfere with the comprehension process of the text's intended meaning. In proofreading, however, spotting errors is the focus of the task. While this task would obviously be less natural and would reveal less about the cognitive principles underlying 'spontaneous reading', it could give us more insights in the role frequency plays in people's competence in detecting spelling errors.

What particularly stands out in the results, is that many of the effects reverse in the spillover regions. More specifically: Factors that initially lead to a processing advantage, result in a delay later on. It has been demonstrated before that effects that are facilitating in one task, can have an inhibitory effect in another task (e.g., Balota et al., 2000; Finkbeiner et al., 2006). Producing a highly frequent form is easier than producing a form with a lower frequency. In contrast, in tasks where the highly frequent form has to be suppressed, this often leads to a processing delay, compared to when a lower frequent form has to be suppressed. In the current experiment, this could describe the different stages of the reading process. Initially, a form with a high frequency is recognized quickly, leading to a facilitated processing. However, this form may be labelled as ‘suspicious’ due to its quick availability and is therefore suppressed, so that the grammatical information from the subject and the verb form can be integrated to determine the correct form. This explains the reversal of the frequency effect and the suffix preference effect, and the delay both of them cause in a more advanced stage of the reading process.

To my knowledge, the phenomenon of effects that reverse within the same task has not yet received much attention in the literature, because a situation as in the current experiment is quite unique (combining several facilitating and inhibitory factors within a single form). I did, however, find another study where the spillover effect did not match the initial effect. Acha and Perea (2008) investigated transposed-letter effects that result in pairs of two real words that are orthographically very similar, but differ in frequency (e.g., *silver* – *sliver*). Participants had to read sentences containing one of these two words and when the word that was used had a lower frequency, readers often initially misperceived it as the higher frequency variant. In early measures (i.e., the duration of the first fixation on the word), this facilitated the reading process. However, due to the initial misperception, the facilitating effect turned into an inhibitory effect later on, when the participants became aware of this misperception, resulting in more regressions and a longer total fixation time. This can be compared with what happened in my experiment. The expected effect of a high frequency and of verb forms ending in <d> indeed facilitated processing initially, but consequently, spelling errors are more often missed. In that case, it is possible that participants become aware of the grammatical mismatch between the subject and the verb form in a later stage, explaining the delayed inhibitory effect.

In general, the results of this study are well-explainable in terms of a Parallel Dual Route Model. In such a model, the computational procedure and the retrieval procedure are constantly in competition and the procedure that is fastest determines the outcome. In spelling production, this results in a single output form, at the moment that one of the two routes ‘wins’ and the resulting output form is produced. Word recognition, and reading in general, have been argued to rely on these two routes as well (see, e.g., Coltheart, Curtis, Atkins, & Haller, 1993; Fiebach, Friederici, Müller, & van Cramon, 2002; for an overview see Jobard, Crivello, & Tzourio-Mazoyer, 2003). However, in perception, unlike production, there is not a fixed moment where the competition between the two procedures has to terminate. As I have shown (and as is known in general), the processing of a word is not immediately ended when the word is outside a fixation. This suggests that, in terms of a Parallel Dual Route Model of perception, the preferred form could be seen as more dynamic and can change over time. That is, the preferred form might be a different form at different stages in the reading process. In an earlier processing stage, the preferred form is likely to be mainly based on frequency information and general preferences (such as the preference for <d>), while in a later processing stage additional information (such as the grammatical properties of the subject) can point towards the other form and change the preference to that form. This dynamicity of the preferred form implies that in perception, the two spelling routes do not as clearly lead to a ‘winner’ as is the case in production.

In the current experiment, the interplay between the computational and the retrieval procedure is visible in multiple ways. As I have shown, highly frequent but incorrectly spelled homophonous verb forms are initially processed faster than incorrectly spelled forms with a lower frequency. This suggests that, at least for highly frequent forms, the output from the retrieval procedure is in advantage initially, which means that the retrieval of whole word forms plays an important role in the early stages of processing homophonous verb forms. Obviously, the advantage of the retrieval procedure is likely to be smaller when the word that is being processed is less frequent. Application of the computational procedure involves more cognitive effort and is therefore only finished in a more advanced stage of processing. The bias towards the most frequent form resulting from the retrieval procedure can be confusing in the application of the computational procedure, as the outputs of the two procedures can be in conflict with each other. The higher the frequency of a form, the more difficult it is to ignore the frequency information and instead rely on the output of the computational procedure. As a result, errors involving frequent forms are more often overlooked, and the delay in processing the following words indicates that the processing costs for the computational procedure were higher due to the saliency of the highly frequent form that had to be suppressed.

To develop a full picture of the cognitive processes underlying the perception of homophones, further research should be undertaken to investigate the effect of other factors known to be important in the spelling process of homophones. These factors include adjacency of the verb form and the word determining its suffix, as cognitive load has been shown to increase when relevant information is more distant. This leads to more errors and is, in line with the current findings, likely to impede the recognition of these errors. This would mean that errors are less often registered when the distance between the subject and the verb form increases, especially when the form is highly frequent and/or ends in <d>. If participants register an erroneously spelled verb form that is distant from the grammatical subject, the delay I already found in the current experiment will probably become even larger, and it is likely that more regressions will be made as well. Furthermore, the present study only investigated one type of homophone pair in the inflectional paradigm of Dutch verbs. For a more complete picture, it would be relevant to involve other homophone pair types as well, such as weak-prefix verbs (*gebeurt*/*gebeurd*), and sublexical homophony (*krabde*/*krapte*, Ernestus & Mak, 2005). Nevertheless, the current study gives us valuable new insights into the cognitive processes underlying the perception of homophones and the way they are spelled.

5 Conclusion

By means of an eye-tracking experiment, I sought to get insight into the online process of the perception of correctly and incorrectly spelled homophonous verb forms, and in this way start unraveling the cognitive principles underlying this process and contribute to our understanding of the spelling process in general. I have shown that factors important in the production of homophonous forms are also relevant in perception, but that they have a different effect at different stages of the reading process. Errors that are made more often during spelling, are initially overlooked more often during reading, but lead to a processing delay in a later stage. These results show the interplay between frequency information on the one hand, and the application of grammatical rules on the other hand, and are therefore well explainable in terms of a Parallel Dual Route Model. However, I have argued that, unlike production, in reading the output of such a model could be seen as a dynamic output that can change over time, rather than a single winning form. Overall, this study shows that, in order to obtain a full understanding of the cognitive principles underlying spelling, it is important to approach these principles both from a productional and from a perceptional perspective, and that it should be kept in mind that these two perspectives are not always necessarily each other's mirror.

6 References

Acha, J., & Perea, M. (2008). The effect of neighborhood frequency in reading: Evidence with transposed-letter neighbors. *Cognition*, *108*(1), 290-300.

Angele, B., Slattery, T. J., & Rayner, K. (2016). Two stages of parafoveal processing during reading: Evidence from a display change detection task. *Psychonomic bulletin & review*, *23*(4), 1241-1249.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of memory and language*, *6*(1), 67-82.

Assink, E. M. (1985). Assessing spelling strategies for the orthography of Dutch verbs. *British Journal of Psychology*, *76*(3), 353-363.

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

Baayen, R. H., McQueen, J., Dijkstra, T., & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: revisiting Dutch plurals. In R. H. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 355-390). Berlin: Mouton de Gruyter.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1995). *The CELEX database*. Nijmegen: Center for Lexical Information, Max Planck Institute for Psycholinguistics, CD-ROM.

Balota, D. A., Law, M. B., & Zevin, J. D. (2000). The attentional control of lexical processing pathways: Reversing the word frequency effect. *Memory & Cognition*, *28*(7), 1081-1089.

Bertram, R., Hyönä, J., & Laine, M. (2000). The role of context in morphological processing: Evidence from Finnish. *Language and Cognitive Processes*, *15*(4-5), 367-388.

Bertram, R., Laine, M., Baayen, R. H., Schreuder, R., & Hyönä, J. (2000). Affixal homonymy triggers full-form storage, even with inflected words, even in a morphologically rich language. *Cognition*, *74*(2), B13-B25.

Bertram, R., Schreuder, R., & Baayen, R. H. (2000). The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy, and productivity. *Journal of experimental psychology: Learning, memory, and cognition*, *26*(2), 489-511.

Bonin, P., & Fayol, M. (2002). Frequency effects in the written and spoken production of homophonic picture names. *European Journal of Cognitive Psychology*, *14*(3), 289-313.

Bosman, A. M. (2005). Development of rule-based verb spelling in Dutch students. *Written language & literacy*, *8*(1), 1-18.

Burani, C., Salmaso, D., & Caramazza, A. (1984). Morphological structure and lexical access. *Visible Language*, *18*(4), 342-352.

Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1430-1450.

Carpenter, P. A., & Just, M. A. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 275-307). New York: Academic Press.

Colé, P., Segui, J., & Taft, M. (1997). Words and morphemes as units for lexical access. *Journal of Memory and Language*, *37*(3), 312-330.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological review*, *100*(4), 589-608.

Dell, G. S., & Gordon, J. K. (2003). Neighbors in the lexicon: Friends or foes?. In N. O. Schiller & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 9-38). Berlin: Mouton de Gruyter.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193-210.

Drieghe, D., Rayner, K., & Pollatsek, A. (2008). Mislocated fixations can account for parafoveal-on-foveal effects in eye movements during reading. *The Quarterly Journal of Experimental Psychology*, *61*(8), 1239-1249.

Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of verbal learning and verbal behavior*, *22*(1), 75-87.

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision research*, *42*(5), 621-636.

Ernestus, M., & Baayen, R. H. (2004). Analogical effects in regular past tense production in Dutch. *Linguistics*, *42*(5), 873-903.

Ernestus, M., & Mak, W. M. (2005). Analogical effects in reading Dutch verb forms. *Memory & cognition*, *33*(7), 1160-1173.

Fayol, M., Largy, P., & Lemaire, P. (1994). Cognitive overload and orthographic errors: When cognitive overload enhances subject–verb agreement errors. A study in French written language. *The Quarterly Journal of Experimental Psychology Section A*, *47*(2), 437-464.

Fiebach, C. J., Friederici, A. D., Müller, K., & Cramon, D. Y. V. (2002). fMRI evidence for dual routes to the mental lexicon in visual word recognition. *Journal of Cognitive Neuroscience*, *14*(1), 11-23.

Finkbeiner, M., Almeida, J., Janssen, N., & Caramazza, A. (2006). Lexical selection in bilingual speech production does not involve language suppression. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1075-1089.

Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression (2nd ed)*. Thousand Oaks, CA: Sage.

Frisson, S., & Sandra, D. (2002). Homophonic forms of regularly inflected verbs have their own orthographic representations: A developmental perspective on spelling errors. *Brain and Language*, *81*(1-3), 545-554.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of memory and language*, *28*(5), 501-518.

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*, *140*(2), 186-209.

Hanssen, E., Schreuder, R., & Neijt, A. (2015). From t-bias to d-bias in Dutch: Evidence from children's spelling and pronunciation. *Written Language & Literacy*, *18*(1), 104-120.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). Eye tracking: A comprehensive guide to methods and measures. Oxford: Oxford University Press.

Hutzler, F., Fuchs, I., Gagl, B., Schuster, S., Richlan, F., Braun, M., & Hawelka, S. (2013). Parafoveal X-masks interfere with foveal word recognition: evidence from fixation-related brain potentials. *Frontiers in systems neuroscience*, *7*, 33.

Jobard, G., Crivello, F., & Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: a metanalysis of 35 neuroimaging studies. *Neuroimage*, 20(2), 693-712.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, *87*(4), 329.

Kapatsinski, V. (2010). What is it I am writing? Lexical frequency effects in spelling Russian prefixes: Uncertainty and competition in an apparently regular system. *Corpus Linguistics and Linguistic Theory*, *6*(2), 157-215.

Katz, L., Rexer, K., & Lukatela, G. (1991). The processing of inflected words. *Psychological Research*, *53*(1), 25-32.

Kliegl, R., Hohenstein, S., Yan, M., & McDonald, S. A. (2013). How preview space/time translates into preview cost/benefit for fixation durations during reading. *The Quarterly Journal of Experimental Psychology*, *66*(3), 581-600.

Largy, P., Fayol, M., & Lemaire, P. (1996). On confounding verb/noun inflections. A study of subject-verb agreement errors in French. *Language and Cognitive Processes*, *11*, 217-255.

Laudanna, A., & Burani, C. (1985). Address mechanisms to decomposed lexical entries. *Linguistics*, *23*(5), 775-792.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, *22*(1), 1-38.

McCormick, S. F., Brysbaert, M., & Rastle, K. (2009). Is morphological decomposition limited to low-frequency words?. *The Quarterly Journal of Experimental Psychology*, *62*(9), 1706-1715.

Meyer, A. S., Huettig, F., & Levelt, W. J. (2016). Same, different, or closely related: What is the relationship between language production and comprehension? [Special issue]. *Journal of Memory and Language*, *89*, 1-254.

Neijt, A., & Schreuder, R. (2007). Asymmetrical phoneme–grapheme mapping of coronal plosives in Dutch. *Written Language* & *Literacy*, *10*(2), 219-234.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist, 38*(1), 63-71.

Pacton, S., & Fayol, M. (2003). How do French children use morphosyntactic information when they spell adverbs and present participles?. *Scientific studies of reading*, *7*(3), 273-287.

R Core Team (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372-422.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory* & *cognition*, *14*(3), 191-201.

Roelofs, A. (2003). Modeling the relation between the production and recognition of spoken word forms. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production*: *Differences and similarities* (pp. 115-158). Berlin: Mouton de Gruyter.

Rubenstein, H., & Pollack, I. (1963). Word predictability and intelligibility. *Journal of Memory and Language*, *2*(2), 147.

Sandra, D., & Fayol, M. (2003). Spelling errors with a view on the mental lexicon: Frequency and proximity effects in misspelling homophonous regular verb forms in Dutch and French. *Trends in Linguistic Studies and Monographs*, *151*, 485-514.

Sandra, D., & Van Abbenyen, L. (2009). Frequency and analogical effects in the spelling of full-form and sublexical homophonous patterns by 12 year-old children. *The Mental Lexicon*, *4*(2), 239-275.

Sandra, D. (2010). Homophone dominance at the whole-word and sub-word levels: Spelling errors suggest full-form storage of regularly inflected verb forms. *Language and speech*, *53*(3), 405-444.

Sandra, D., Frisson, S., & Daems, F. (1999). Why simple verb forms can be so difficult to spell: The influence of homophone frequency and distance in Dutch. *Brain and language*, *68*(1-2), 277-283.

Sandra, D., Frisson, S., & Daems, F. (2004). Still errors after all those years...: Limited attentional resources and homophone frequency account for spelling errors on silent verb suffixes in Dutch. *Written Language & Literacy*, *7*(1), 61-77.

Schmitz, T., Chamalaun, R., & Ernestus, M. (2018-in press). The Dutch verb-spelling paradox in social media: A corpus study. *Linguistics in the Netherlands*, *2018*.

Schotter, E. R., Bicknell, K., Howard, I., Levy, R., & Rayner, K. (2014). Task effects reveal cognitive flexibility responding to frequency and predictability: Evidence from eye movements in reading and proofreading. *Cognition*, *131*(1), 1-27.

Sereno, J. A., & Jongman, A. (1997). Processing of English inflectional morphology. *Memory & Cognition*, *25*(4), 425-437.

Slattery, T. J. (2009). Word misperception, the neighbor frequency effect, and the role of sentence context: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1969.

Sprenger, S. A., Levelt, W. J., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of memory and language*, *54*(2), 161-184.

Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1280-1293.

Stemberger, J. P., & MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory & Cognition*, *14*(1), 17-26.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, *7*(4), 263-272.

Tjong Kim Sang, E., & van den Bosch, A. (2013). Dealing with big data: The case of Twitter. *Computational Linguistics in the Netherlands Journal*, *3*, 121-134.

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior research methods*, *38*(4), 584-589.

Veldre, A., & Andrews, S. (2018). How does foveal processing difficulty affect parafoveal processing during reading?. *Journal of Memory and Language*, *103*, 74-90.

Verhaert, N. (2016). *Rules Or Regularities? The Homophone Dominance Effect in Spelling and Reading Regular Dutch Verb Forms.* PhD dissertation, Universiteit Antwerpen, Belgium.

Verhaert, N., Danckaert, E., & Sandra, D. (2016). The dual role of homophone dominance. Why homophone intrusions on regular verb forms so often go unnoticed. *The Mental Lexicon*, *11*(1), 1-25.

Whaley, C. P. (1978). Word-nonword classification time. *Journal of Memory and Language*, *17*(2), 143.

White, K. K., Abrams, L., Zoller, S. M., & Gibson, S. M. (2008). Why did I right that? Factors that influence the production of homophone substitution errors. *The Quarterly Journal of Experimental Psychology*, *61*(7), 977-985.

Witzel, N., Witzel, J., & Forster, K. (2012). Comparisons of online reading paradigms: Eye tracking, moving-window, and maze. *Journal of psycholinguistic research*, *41*(2), 105-128.

Zwitserlood, P. (2003). The internal structure of words: Consequences for listening and speaking. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production*: *Differences and similarities* (pp. 79–114). Berlin: Mouton de Gruyter.

Appendix: Verb forms used in the experimental materials

Verb	Translation	Dominant form	Relative Frequency (log)
aanvaarden	'to accept'	d	0.34
beantwoorden	'to answer'	d	0.14
begeleiden	'to supervise'	d	0.47
behoeden	'to save'	d	0.11
beïnvloeden	'to influence'	d	0.39
bekleden	'to clothe'	d	0.21
belanden	'to end up'	d	0.34
bespieden	'to spy'	d	0.15
besteden	'to spend'	d	0.58
bevreemden	'to amaze'	d	0.78
bevrijden	'to free'	d	0.72
doorgronden	'to understand'	d	0.51
misleiden	'to mislead'	d	0.83
ontbranden	'to ignite'	d	0.05
onthouden	'to remember'	d	0.04
ontleden	'to dissect'	d	0.25
verantwoorden	'to justify'	d	1.23
verblinden	'to dazzle	d	0.72
verbranden	'to burn'	d	0.71
vergoeden	'to reimburse'	d	0.50
verkleden	'to dress'	d	0.82
verleiden	'to seduce'	d	0.31
vermelden	'to mention'	d	0.40
vermoeden	'to suspect'	d	0.28
vermoorden	'to murder'	d	1.07
verspreiden	'to spread'	d	0.44
vervreemden	'to alienate'	d	0.61
verwonden	'to wound'	d	0.48
verwoorden	'to articulate'	d	0.58
voorbereiden	'to prepare'	d	1.15
wedden	'to bet'	d	0.82
afleiden	'to distract'	dt	0.82
antwoorden	'to answer'	dt	0.98
behouden	'to keep'	dt	0.90
bestrijden	'to fight'	dt	0.89
betreden	'to enter'	dt	0.53
bidden	'to pray'	dt	0.06
binden	'to bind'	dt	0.56
dulden	'to tolerate'	dt	0.57
glijden	'to slide'	dt	1.17
houden	'to keep'	dt	0.50
lijden	'to suffer'	dt	0.91

Verb	Translation	Dominant form	Relative Frequency (log)
melden	‘to report’	dt	0.71
mijden	‘to avoid’	dt	0.15
ondervinden	‘to experience’	dt	0.98
ontaarden	‘to degenerate’	dt	0.52
redden	‘to rescue’	dt	0.30
rijden	‘to drive’	dt	0.51
schelden	‘to scold’	dt	0.24
schudden	‘to shake’	dt	0.62
smeden	‘to forge’	dt	0.52
snijden	‘to cut’	dt	0.70
spreiden	‘to spread’	dt	0.49
strijden	‘to fight’	dt	0.20
verbeelden	‘to imagine’	dt	0.02
verbieden	‘to forbid’	dt	0.67
verbinden	‘to connect’	dt	0.91
vermijden	‘to avoid’	dt	0.04
vinden	‘to find’	dt	0.09
voeden	‘to feed’	dt	0.88
volharden	‘to persevere’	dt	0.74
wijden	‘to dedicate’	dt	0.50
zenden	‘to send’	dt	0.54