Investigating variability in morphological processing with Bayesian distributional models

Laura Anna Ciaccio[1] & João Veríssimo[2]

[1] Potsdam Research Institute for Multilingualism
University of Potsdam
[2] Department of Linguistics
University of Potsdam

Abstract

We investigated the processing of morphologically complex words adopting an approach that goes beyond estimating average effects and allows testing predictions about variability in performance. We tested masked morphological priming effects with English derived ('printer') and inflected ('printed') forms priming their stems ('print') in non-native speakers, a population that is characterized by large variability. We modelled reaction times with a shifted-lognormal distribution using Bayesian distributional models, which allow assessing effects of experimental manipulations on both the mean of the response distribution ('mu') and its standard deviation ('sigma'). Our results show similar effects on mean response times for inflected and derived primes, but a difference between the two on the sigma of the distribution, with inflectional priming increasing response time variability to a significantly larger extent than derivational priming. This is in line with previous research on non-native processing, which shows more variable results across studies for the processing of inflected forms than for derived forms. More generally, our study shows that treating variability in performance as a direct object of investigation can crucially inform models of language processing, by disentangling effects which would otherwise be indistinguishable. We therefore emphasise the importance of looking beyond average performance and testing predictions on other parameters of the distribution rather than just its central tendency.

Keywords: RT distribution, distributional models, masked priming, visual word recognition
Word count: 3569 words (excluding Method section)

Introduction

Psycholinguistic research, and cognitive science more generally, is primarily concerned with estimating average effects for population samples in a given task. The underlying assumption is that human performance is largely homogeneous—across participants, items, or trials—and, consequently, average effects can be taken to reflect the cognitive mechanisms involved in performing the task. Hence, variability in performance has not traditionally been seen as a potential source of information about the structure of the cognitive system. Instead, it has been mostly ignored or dismissed as noise (Andrews, 2012). However, recent research has witnessed an increasing interest in variability in language processing, with a growing consensus that this is a reflection of a flexible system, and therefore an intrinsic aspect of language processing mechanisms (see Amenta & Crepaldi, 2016; Kidd et al., 2018). This makes investigating variability crucially informative for building theoretical models of language processing.

The present work focuses on variability in morphological processing, that is, in processing complex words such as *player* [play][-er] or *played* [play][-ed], during visual word recognition. A well established finding from research on morphological processing are so-called morphological priming effects. We speak of a morphological priming effect when responses to target words are facilitated (i.e., faster) when they are preceded by a morphologically related prime, such as in *walked-WALK*, as compared to when they are preceded by a prime unrelated in form or meaning, such as *kissed-WALK* (e.g. Rastle et al., 2000; Feldman & Soltano, 1999; see Amenta & Crepaldi, 2012 for a review). Morphological priming studies generally employ the *masked priming* technique coupled with a lexical decision task. In this paradigm, primes are only presented very briefly (around 50 ms) and are preceded by a visual mask, thus preventing their conscious recognition (Baayen, 2014; Kinoshita & Lupker, 2004). Different interpretations have been proposed for this speed-up effect from morphologically related primes. On the one hand, this has been explained by positing a mechanism of morphological decomposition stripping off affixes from complex words, for example *-ed* from *walked*, thereby pre-activating the stem *walk*, leading to faster reaction times when subsequently encountering the stem as target word (Rastle et al., 2004; Taft & Forster, 1975). Alternatively, it has been suggested that morphological decomposition relies on a stem-activation mechanism, which extracts edge-aligned embedded stems from letter strings (Grainger & Beyersmann, 2017). Finally, morphological priming effects have been also explained in terms of shared semantic and orthographic properties of prime and target, hence without positing a level of morphological representation (Baayen et al., 2011; Baayen & Smolka, 2020; Feldman, 2000).

Both authors contributed equally to this work.

This work has been partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 317633480 – SFB 1287, Projects B04 and Q.

We thank Yara Amer and Jerley Castro for their support with data collection.

Correspondence concerning this article should be addressed to Laura Anna Ciaccio, University of Potsdam, Potsdam Research Institute for Multilingualism, Campus Golm, Haus 2, Karl-Liebknecht-Straße 24–25, 14476 Potsdam, Germany. +49 (0)331/977-2638. E-mail: ciaccio@uni-potsdam.de

Many masked priming studies have consistently reported morphological priming effects from typologically different languages and different linguistic phenomena (see Ciaccio et al., 2020). At the same time, morphological priming studies have also provided evidence for a rather flexible cognitive structure, which is sensitive to the specific characteristics of a given language, leading to different patterns of priming effects (see e.g., Boudelaa & Marslen-Wilson, 2005; Günther et al., 2019). Even more interestingly, morphological priming effects have been shown to also vary *within* languages, as a result of speakers' characteristics. For example, Andrews and Lo (2013) reported similar priming effects for truly morphologically related (e.g., *farmer-FARM*) and pseudo-morphologically related (e.g., *corner-CORN*) prime-target pairs in speakers with better spelling than vocabulary skill, but larger priming effects for the former than the latter type in speakers with better vocabulary than spelling. Similarly, Beyersmann et al. (2016) showed that the size of embedded stem priming effects with non-morphological pseudo-word primes (e.g., **flexint-FLEX*) is modulated by reading proficiency; comparable results were obtained by Beyersmann et al. (2015) with a composite language proficiency measure. Importantly, as also underlined by Andrews (2012), focusing on variability in morphological processing can be crucial to understanding why different studies led to apparently inconsistent results (compare e.g., Longtin & Meunier, 2005; and Heathcote et al., 2018 on priming with **flexint-FLEX* pair types).

A population that lends itself particularly well to the investigation of variability in language processing are non-native (L2) speakers of language. Performance in an L2 is characterized by larger heterogeneity than in native language, both within and across individuals (Bialystok & Hakuta, 1999; Hopp, 2013; White, 2003). When it comes to morphological processing, results from L2 speakers tend to be more variable across studies than results from native speakers. Masked morphological priming studies with L2 speakers have generally reported robust priming effects from derived forms (Ciaccio & Clahsen, 2020; Diependaele et al., 2011; Heyer & Clahsen, 2015; Li et al., 2017), but smaller priming effects that vary substantially across studies in the case of inflected forms (Jacob et al., 2018; Veríssimo et al., 2018; Feldman et al., 2010). This is in line with previous work suggesting persistent vulnerabilities in L2 speakers for inflectional morphology, especially with respect to the ability of consistently and reliably accessing inflected forms and morphosyntax (Blom et al., 2006; Hopp, 2013; White, 2003). Moreover, some, but not all, L2 masked morphological priming studies additionally reported significant priming effects for purely orthographically related prime-target word pairs (e.g., *scandal-SCAN*; Heyer & Clahsen, 2015; Diependaele et al., 2011; Feldman et al., 2010; Li et al., 2017), an effect otherwise consistently absent in L1 speakers (Crepaldi et al., 2010; Longtin et al., 2003; Rastle et al., 2004), which raises the question of whether the observed morphological priming effects in an L2 may in fact be due to prime-target orthographic overlap.

Similarly to research on native language processing, the few studies that have directly investigated or discussed variability in L2 morphological processing (e.g., Veríssimo et al., 2018; Basnight-Brown et al., 2007; Bosch et al., 2019; Feldman et al., 2010; Li et al., 2017) have done so with a focus on inter-individual variability, asking to what extent factors such as age of acquisition or proficiency can account for the different priming effects reported for L1 and L2 speakers. An alternative approach for investigating variability in morphological processing (and visual word recognition more in general) consists in looking at the whole reaction time (RT) distribution, thus going beyond estimates of average effects for a given population to test the amount of variability around that effect. Despite recommendations of looking beyond central tendencies of distributions in RT analyses (e.g., Heathcote et al., 1991), this rationale has hardly been applied to visual word recognition studies and it is virtually absent from morphological processing research (but see Balota et al., 2008; Hasenäcker et al., 2016; and Yap et al., 2006).

The present study

The experiment reported below contrasted the processing of inflected and derived forms using the masked priming paradigm. We investigated the effect of these two types of morphologically complex words on the visual recognition of their bases. Specifically, we assessed priming effects elicited by inflected *-ed* past-tense forms (e.g., *printed*) and derived *-er* nominalisations (e.g., *printer*) on lexical decision times to the same target stems (e.g., *print*). The two conditions were created as to be comparable in a range of morphological and non-morphological properties: (a) both *-ed* and *-er* are regular, productive and semantically transparent suffixes; (b) the two types of morphological primes were equivalent in their amount of orthographic overlap with their targets; and (c) priming conditions were well-matched in other lexical measures (see below). Importantly, priming effects were estimated not only on average response speed, but also on the variability of responses.

We made use of Bayesian *distributional models* to assess the effects of morphologically related primes. In distributional models, estimates refer not only to the mean of responses in different experimental conditions, but can also describe additional features of the response distribution, for example, its variability. Importantly, these models allow the different parameters of a distribution to depend on a set of explanatory variables. For example, RTs can be estimated to be shorter or longer in terms of their central tendency, but also narrower (less variable) or wider (more variable), depending on a given predictor (Bürkner, 2018; Kneib & Umlauf, 2017). As a result, distributional models provide better estimates of differences between means (because they relax the common assumption of equal variances), but more importantly, they allow going beyond the mean in order to draw inferences about how the whole shape of a response distribution is affected by any experimental manipulations (see Balota et al., 2008; Balota & Yap, 2011).

Another advantage of Bayesian models is that they allow fitting virtually any kind of distribution in a straightforward way (Nicenboim et al., 2016). Here, we modelled RTs as a 'shifted lognormal' distribution. This distribution incorporates a *shift* parameter, which moves the whole RT distribution by an estimated amount of milliseconds (Logan, 1992; Rouder, 2005); this allows log-RTs to more closely follow a normal distribution, thus better satisfying the assumptions of linear models (Baayen & Milin, 2010). The shifted-lognormal is described by two other parameters, besides the shift: (a) *mu*, the mean of (normally-distributed) log-RTs, and (b) *sigma*, the standard deviation of (normally-distributed) log-RTs. The mean *mu* is a 'location' parameter, expressing central tendency, and can be taken as an index of difficulty (Wagenmakers & Brown, 2007): experimental manipulations that slow down responses, for example, can be empirically described as increasing the mean of a lognormal distribution (corresponding to the *median* of RTs in milliseconds), and thus dispersing the RT distribution in the direction of longer times. The standard deviation *sigma* is a 'scale' parameter, which stretches or squeezes the RT distribution around the same centre (i.e., around the median RT). In this way, the *sigma* parameter captures the variability of RTs, and importantly, does so independently of the central tendency of the distribution, that is, over and above the effects of condition difficulty (Wagenmakers & Brown, 2007).

We made use of such distributional models to assess priming effects on both the mean and *sigma* parameters of the shifted-lognormal distribution of lexical decision RTs. In line with some of the previous morphological priming studies with L2 groups (Jacob et al., 2018; Kirkici & Clahsen, 2013; Veríssimo et al., 2018), we expected a difference between derivational and inflectional priming on the mean of log-RTs. Specifically, derived forms should produce facilitation effects on the recognition of their constituent stems, but the masked priming effects elicited by inflected forms may be smaller in magnitude or absent. Additionally, if L2 speakers show particular difficulties with consistently accessing and decomposing inflected forms in written word recognition, then the presentation of inflected primes may produce more inconsistent benefits, possibly leading to more variable lexical decision responses. In that case, inflected primes are expected to increase the *sigma* (standard deviation) of log-RTs in comparison to derived (and possibly unrelated) primes.

Method

Participants

Sixty-nine intermediate to advanced non-native speakers of English (54 women; 15 men) took part in the experiment in exchange for payment or course credits. Their mean age was 26.09 years (SD = 5.27, range = 18–37). All participants started learning English after the age of 4 (mean age of acquisition = 8.71, SD = 1.95, range = 4–13; 3 NAs). Participants were all native speakers of German, three of whom additionally spoke Russian as a native language. They all lived in Germany at the time of testing. All participants reported reading in English to some extent in their daily lives (mean use of English, as compared to other languages = 32.14%, SD = 18.10, range = 2–80%). Skill in English was tested by means of a 50-item multiple-choice grammar test adapted from the Oxford Placement Test 1 (Allan, 2004). Participants' mean score was 38.62/50 (SD = 5.91; range = 21–48), suggesting that their proficiency level was between B2 and C2 of the Common European Framework of Reference for Languages (Council of Europe, 2001). All participants additionally took the LexTALE test (Lemhöfer & Broersma, 2012), a standardized vocabulary test consisting of an un-speeded visual lexical decision task. The LexTALE score is a percentage score, calculated as the percentage of correct responses corrected for the proportion of existing and non-existing words in the test. The group achieved a mean score of 78.91% (SD = 9.28; range = 61.25–98.75%, again roughly corresponding to B2 to C2 level). Prior to testing, all participants signed a written consent.

Materials

The experiment included 102 English monomorphemic verbs used as targets (e.g., *print*). These were preceded by their *-ed* past-tense form (e.g., *printed*) as the inflected prime, their *-er* nominalization (e.g., *printer*) as the derived prime, or by an unrelated prime. Unrelated primes were dissimilar in form and meaning from their corresponding targets; half of them were *-ed* inflected forms and half of them were *-er* derived words. The distributions of word-form frequency, lemma frequency, and length (in letters) of the three prime types (inflected, derived, unrelated) were kept as similar as possible. Word-form and lemma frequency were extracted from the SUBTLEX-UK database (van Heuven et al., 2014) and are provided in the Zipf scale, which approximately spans from 1 to 7; values below 3 indicate relatively low frequency, while values above 4 indicate high frequency. Item characteristics are provided in Table 1.

Table 1

Summary of the item characteristics (mean, SD, range)

Item type	Length (letters)	Word-form frequency	Lemma frequency
Target	5.09 (1.24)	3.93 (0.62)	4.04 (0.53)
	3–8	1.9–5.37	2.55–5.21
Derived prime	6.73 (1.07)	3.03 (0.64)	3.2 (0.65)
	4–9	1.3–4.22	1.6–4.46
Inflected prime	6.73 (1.07)	3.29 (0.73)	3.88 (0.58)
	4–9	1.17–4.76	2.23–5.11
Unrelated prime	6.73 (1.07)	3.24 (0.73)	3.66 (0.84)
	4–9	1.3–4.73	1.3–5.57

The 102 experimental targets and their corresponding inflected, derived, and unrelated primes were distributed across three presentation lists, following a Latin-Square design, so that each participant saw 34 targets associated with each of the three prime types. Three additional lists were created by reversing the order of the items in each list, for a total of six presentation lists. The 102 experimental prime-target pairs were mixed with 438 filler pairs, for a total of 540 items in each list. All filler primes were existing words, while 270 targets were non-existing words, requiring a 'no' response in 50% of the trials. Non-words were generated from existing English words using the software Wuggy (Keuleers & Brysbaert, 2010). Of the 438 filler pairs, 102 pairs included an inflected *-ed* prime (51) or a derived *-er* prime (51) combined with a non-existing word target (e.g., *barked-LEAMS*). This way, the presentation of *-ed* and *-er* primes did not represent a cue for lexicality of the target. Additionally, 68 of the filler targets were non-existing words that were orthographically embedded in their primes (e.g., *sincere-SINCH*), so that form overlap was also not a cue for a 'yes' response. Overall, 25.2% of the prime-target pairs in each presentation list were related, either morphologically or orthographically.

Procedure

Participants were tested in a quiet laboratory room. They were homogeneously assigned to one of the six lists. Participants' accuracy and RTs in milliseconds were measured using the experimental software DMDX (Forster & Forster, 2003). We informed the participants that they would see a series of existing English words and invented words, and that they would have to indicate as quickly and as accurately as possible whether the word on screen is an existing word ('lexical decision'). The 'yes' or 'no' responses were provided by pressing one of two buttons on a gamepad. Participants provided 'yes' responses with their dominant hand. Each trial started with a blank screen presented for 500 ms, followed by a mask consisting of ten hashes, also presented for 500 ms. Next, the prime appeared, remaining on screen for 50 ms and directly followed by the target. The target was displayed until a button press or until the timeout, which was set at 2,000 ms. The next trial automatically began right after a button press or timeout. Primes were always presented in lowercase letters, and targets in uppercase letters.

Data analysis

Incorrect responses (7.47%) and timeouts (0.18%) were excluded from analysis. No participants or items were excluded. RTs were analysed with Bayesian mixed-effects distributional regression models, assuming a 'shifted lognormal' response distribution. Bayesian models combine prior information with the evidence from the data in order to obtain a probability distribution over a parameter's possible values—its *posterior distribution*. In this way, an experimental effect can be quantified in terms of the likelihood of different magnitudes, which is more informative than a binary statement about whether an effect exists or not (McElreath, 2020; Vasishth et al., 2018). We chose wide, weakly informative priors for all parameters. Thus, only very extreme and unreasonable values were ruled out a priori, while still allowing a wide range of possible inferences regarding the size of effects (see Table A1 in the Appendix for the full specification of prior distributions). Analyses were performed with the *brms* package in R (Bürkner, 2017; R Core Team, 2020). The procedures for fitting Bayesian models and assessing their convergence followed recent recommendations (Schad et al., 2020; Vasishth et al., 2018).

Models included fixed effects for prime type (unrelated, inflected, derived), as well as trial position (centered). The prime type predictor was coded with treatment contrasts (i.e., 0/1), comparing each of the morphological priming conditions to the unrelated condition. Fixed effects were estimated both on the mean of log-RTs and on the *sigma* parameter of the distribution, corresponding to the standard deviation of log-RTs. Models included random effects for participant and item (for both mean and *sigma*). Random slopes for prime type were not included, because model comparisons on the basis on ELPD (a Bayesian measure of predictive accuracy) showed that none of the random slopes (by participant and by item, for mean and for *sigma*) provided meaningful improvements in goodness-of-fit (better fit was defined here as an ELPD difference larger than 2 standard errors; Bürkner, 2017; Vasishth et al., 2018).

Results

Accuracy rates were very high (at least 90%) in all conditions. No further analyses of accuracy were conducted. Overall median RTs in the three prime type conditions were 661 ms (unrelated), 625 ms (inflected), 625 ms (derived).

A summary of the statistical model is shown in Table 2. For each estimate in Table 2, we report the mean of its posterior distribution together with its 95% credible interval (i.e., the range within which a parameter falls with 95% probability). With regards to the mean of log-RTs (see the four top rows in Table 2), both morphological primes were found to speed up responses (cf. negative coefficients). That is, the mean of log-RTs was shorter following inflected and derived primes than following unrelated primes (by 43 ms for inflected, and 39 ms for derived, in back-transformed estimates). These effects show relatively narrow 95% credible intervals, with upper bounds that are far away from zero, and thus indicate strong support for facilitation effects on the speed of lexical decision responses for both types of morphological primes.

Table 2

Summary of a distributional model, with both mean and standard deviation (sigma) of log-RTs predicted by prime type (unrelated, inflected, derived)

	Estimate	L-95% CI	U-95% CI
Intercept (unrelated)	5.935	5.874	5.995
Prime type (derived vs. unrelated)	-0.109	-0.128	-0.091
Prime type (inflected vs. unrelated)	-0.121	-0.141	-0.102
Trial (centered)	-0.198	-0.247	-0.147
Sigma $\sim$ Intercept (unrelated)	0.314	0.296	0.332
Sigma $\sim$ Prime type (derived vs. unrelated)	0.010	-0.004	0.026
Sigma $\sim$ Prime type (inflected vs. unrelated)	0.025	0.010	0.042
Sigma $\sim$ Trial (centered)	0.013	-0.022	0.056

The model also revealed effects of morphological primes on the *sigma* parameter, that is, on the standard deviation of log-RTs (see the four bottom rows in Table 2). Specifically, the positive coefficients for both morphological priming conditions suggest an increase in variability relative to the unrelated condition. Figure 1, panels a and b, show the full posterior distributions for these effects. For derived primes (panel a), a large proportion of the posterior distribution is on the positive side (i.e., greater standard deviation in the derived than in the unrelated condition). However, there is still some mass on the negative side, and the 95% credible interval for this effect crosses zero, indicating that very small or even negative values are not completely implausible (Lindley, 1970; Rouder et al., 2018). There is much clearer support for an effect of inflected primes (again, relative to the unrelated condition), with almost all of its posterior distribution on the positive side (panel b). Critically, the effect of inflected and derived primes can also be directly compared; in a Bayesian framework, this can be achieved from the same statistical model by simply subtracting pairs of posterior samples for one and the other effect. The resulting posterior distribution is shown in Figure 1, panel c. Although the difference between the two conditions is relatively small in magnitude, the exclusively positive 95% credible interval provides evidence that inflected primes increased the standard deviation of log-RTs more than derived primes.

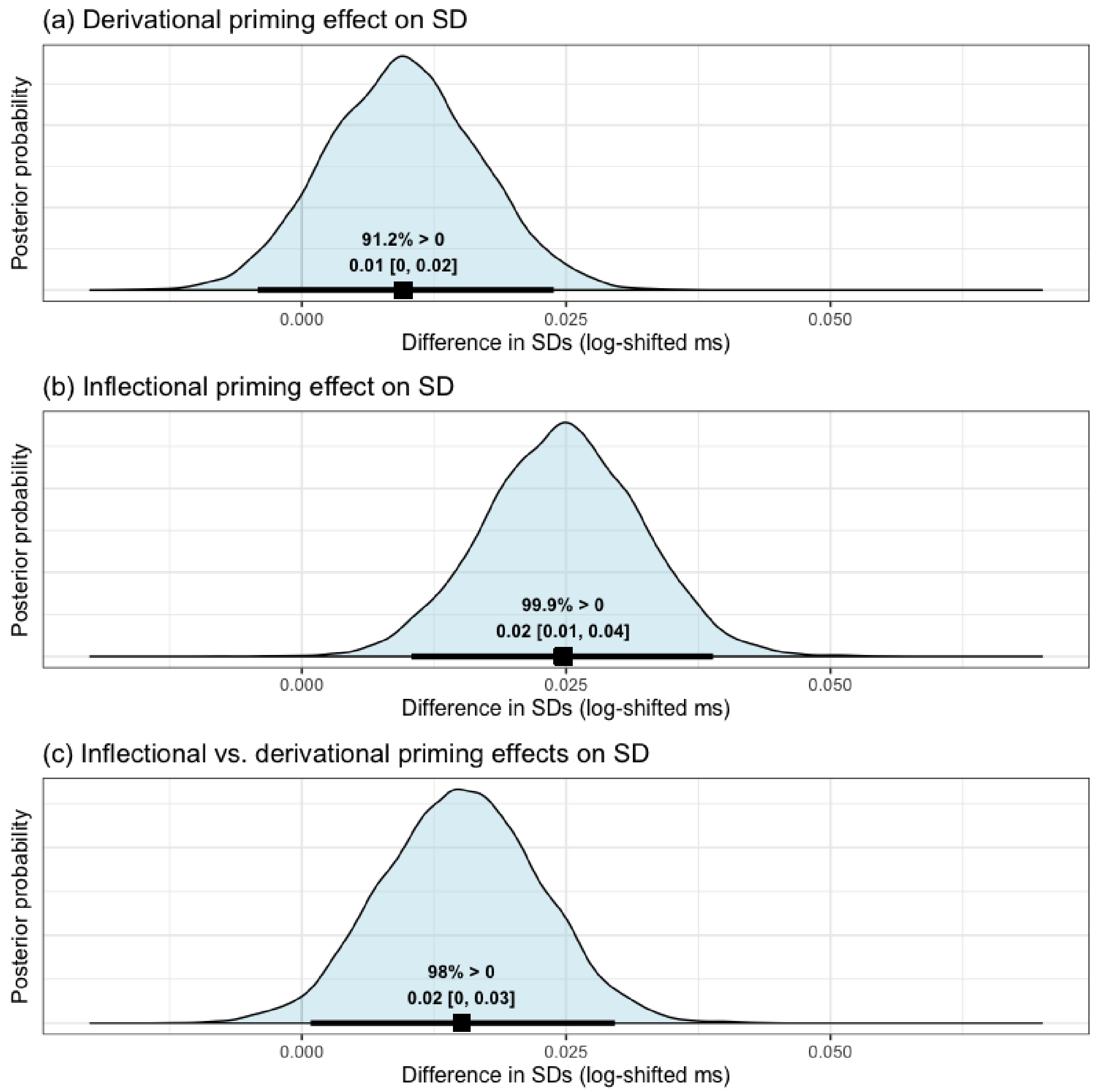

Figure 1

Posterior distributions for the effects of derived (panel a) and inflected primes (panel b) on the SDs of log-RTs, relative to unrelated primes, and for the comparison between inflected and derived primes (panel c). Each posterior distribution contains its mean and 95% credible interval displayed in numerical and graphical form, as well as the percentage of posterior samples that are on the positive side.

Discussion

In the present study, we investigated masked morphological priming effects both on mean RTs to target words and on RT variability. Concerning mean RTs, based on the previous masked priming literature involving L2 speakers (Jacob et al., 2018; Kirkici & Clahsen, 2013; Silva & Clahsen, 2008; Veríssimo et al., 2018), we predicted priming effects from derived words, but reduced or absent priming from inflected forms. However, our results showed that both types of morphologically related primes sped up word recognition latencies (and to a similar extent). Although unexpected, the finding of a priming effect with inflected words in L2 speakers has been previously reported in at least two other masked priming studies (Feldman et al., 2010; Foote, 2017), though these did not include derived primes. Such priming effects from inflected forms may be the result of morphological processing (e.g., decomposition into stems and affixes) or may arise from purely orthographic overlap between primes and targets, which is likely to play a role in L2 processing (Diependaele et al., 2011; Heyer & Clahsen, 2015; Li et al., 2017). In general, the available masked priming literature involving L2 speakers shows quite unstable results when it comes to inflected forms. This contrasts with the current evidence on derived words, for which priming effects have been found to be particularly robust across different types of derived words (e.g., prefixed and suffixed), even for L2 speakers (e.g., Ciaccio & Clahsen, 2020; Diependaele et al., 2011; Silva & Clahsen, 2008). From this, we can infer that examining variability in performance may be particularly informative with regard to the mechanisms of morphological processing.

When analysing the effects of morphologically related primes on the standard deviation (*sigma*) of log-RTs, which was our measure of variability in responses, we predicted that the presentation of inflected primes would produce more inconsistent benefits as compared to derived primes, leading to more variable lexical decision times. We found that the presentation of both derived and inflected primes led to larger RT variability as compared to unrelated primes, but inflected primes indeed increased RT variability more so than derived primes. While all major morphological processing accounts—i.e., affix stripping (Rastle et al., 2004; Taft & Forster, 1975), embedded word activation (Grainger & Beyersmann, 2017), and form-and-meaning approaches (Baayen et al., 2011; Baayen & Smolka, 2020; Feldman, 2000)—can theoretically explain the priming effect on mean RTs, they all need some refinement in order to accommodate for differences in RT variability following different types of morphologically complex primes. For this, it is useful to look at theories of lexical processing that specifically tried to capture variability in response times.

Variability in performance has been related to automaticity in lexical processing, especially in the context of language learning research. In a series of lexical decision experiments, Segalowitz and Segalowitz (1993) and Segalowitz et al. (1998) showed that, when lexical representations become well-established, for example with practice, lexical processing becomes more efficient, or automatized, and, as a consequence, RTs are overall less variable. Instead, the process of establishing new representations leads to an increase in RT variability, as observed by Solovyeva and DeKeyser (2018) in a series of novel word learning tasks. While several studies have looked at variability in performance in the context of second or artificial language learning, including some testing morphosyntax (e.g., Rodgers, 2011; for a review, see Segalowitz, 2008), none of them have specifically looked at online morphological processing.

By extending this theoretical account of variability in lexical processing to morphological processing and masked priming, we can take variability in RTs to reflect how efficiently, or automatically, speakers access the lexical representation of a stem, for example 'print', following the presentation of its morphological complex forms: our results suggest that, at least in an L2, accessing 'print' given 'printed' works less automatically than accessing 'print' given 'printer'. What remains to be understood is what specifically causes this difference in processing automaticity, and therefore in variability, between inflected and derived words. First, although reliance on orthographic cues possibly plays a more prominent role in L2 morphological processing as compared to native processing (e.g., Heyer & Clahsen, 2015; Li et al., 2017), this cannot explain our effect on RT variability, since the two morphological conditions were perfectly pairwise matched with regard to prime–target orthographic overlap. Therefore, the effect we report is more likely to result from *morphological* differences between inflected and derived words. In spite of their superficial similarity, derived and inflected words modify their stems in different ways: while derivational operations create new lexical entries, inflectional operations are purely grammatical, in that they spell out morphosyntactic features of the stem, such as number, case, or tense (see Anderson, 1992). Following the framework sketched by Segalowitz and Segalowitz (1993) and Solovyeva and DeKeyser (2018), we suggest that the different lexical status of inflected and derived words generates the differences in processing variability we observed: processing complex forms which have established lexical representations (in our case, derived words) and accessing the representations of their stems works more automatically, or efficiently, than processing complex forms that are bare spell-outs of grammatical properties of their stems (inflected forms). This should be particularly true in L2 populations, given that morphosyntax remains a particularly vulnerable domain in L2 acquisition, and even proficient L2 speakers may not be able to consistently access the information contained in inflected forms (Blom et al., 2006; Hopp, 2013; White, 2003). In this way, the framework proposed by Segalowitz and Segalowitz (1993) and Solovyeva and DeKeyser (2018) can very well—and quite economically—account for our findings.

An open question is to what extent what we found also applies to native processing. Theoretically, the different lexical nature of derived and inflected words may also affect the variability of responses in native speakers. Alternatively, the effects we have obtained may be specific to non-native or less proficient speakers, given their particularly vulnerability in the domains of inflection and morphosyntax. Future studies should examine morphological priming effects on RT variability across different populations, including native speakers and speakers at different proficiency levels. The detection of such effects in native speakers may nevertheless prove to be particularly challenging, considering that native speakers display higher language proficiency, less variable RTs, and show consistent and robust priming effects across all types of morphologically related primes, at least in the masked priming paradigm.

A relevant aspect of the study concerns the type of RT distribution that we modelled. Similarly to the studies on RT variability reported above, our goal was to investigate variability in responses as measured independently from the central tendency of the RT distribution. However, in RT distributions, the mean and standard deviation of the distribution have a linear relationship (Wagenmakers & Brown, 2007), such that manipulations that slow down responses (i.e., lead to larger mean RTs in milliseconds) also increase the variability (i.e., lead to larger SDs). By modeling a (shifted) log-normal distribution, we could not only test effects on both parameters of interest, that is, *mu*—the central point of a normally distributed log-RT distribution—and *sigma*—its standard deviation—, but we could also test the effect of morphological priming on these two parameters *independently* of each other, thus being able to isolate effects on variability from bare speed-up effects. This is very well illustrated by the fact that derived and inflected primes had different effects on *sigma* although the two prime types were associated with very similar mean log-RTs.

Overall, we have shown that morphological type (derivation vs. inflection) modulates the processing of complex words as tested with the masked priming paradigm, with inflected primes in particular producing larger increases in the variability of responses. By examining only the central tendency (i.e., the mean) of the response times, we would have been unable to distinguish priming effects produced by derived and inflected words. However, the differences between the two prime types emerged when going beyond mean RTs by specifically investigating the variability of responses. Besides the relevance of the present results for models of morphological processing, the current study therefore emphasises the importance of testing predictions on other parameters of the RT distribution rather than just its central tendency (Balota et al., 2008; Heathcote et al., 1991), and more generally, of treating variability in performance as a direct object of investigation in psycholinguistics.

Supplementary files

Appendix: Table A1. Prior distributions on each estimate of the Bayesian distributional model.

Open practices statement

The materials, the raw data, and the analysis scripts related to this manuscript are available at https://osf.io/4zwty/.

References

Allan, D. (2004). *Oxford Placement test 1*. Oxford University Press.

Amenta, S., & Crepaldi, D. (2016). Editorial: The variable mind? How apparently inconsistent effects might inform model building. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00185

Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Psychology*, *3*. https://doi.org/10/ggvfm5

Anderson, S. R. (1992). *A-Morphous morphology*. Cambridge University Press.

Andrews, S. (2012). Individual differences in skilled visual word recognition and reading: The role of lexical quality. In *Visual word recognition: Meaning and context, individuals and development, Vol. 2* (pp. 151–172). Psychology Press.

Andrews, S., & Lo, S. (2013). Is morphological priming stronger for transparent than opaque words? It depends on individual differences in spelling and vocabulary. *Journal of Memory and Language*, *68*(3), 279–296. https://doi.org/10.1016/j.jml.2012.12.001

Baayen, H. (2014). Experimental and psycholinguistic approaches. In R. Lieber & P. Štekauer (Eds.), *The Oxford Handbook of Derivational Morphology* (pp. 95–117).

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12. https://doi.org/10/gddhjp

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118*(3), 438. https://doi.org/10.1037/a0023851

Baayen, R. H., & Smolka, E. (2020). Modeling morphological priming in German with naive discriminative learning. *Frontiers in Communication*, *5*. https://doi.org/10.3389/fcomm.2020.00017

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science*, *20*(3), 160–166. https://doi.org/10/bkcjq6

Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language*, *59*(4), 495–523. https://doi.org/10/fs3gpg

Basnight-Brown, D. M., Chen, L., Hua, S., Kostić, A., & Feldman, L. B. (2007). Monolingual and bilingual recognition of regular and irregular English verbs: Sensitivity to form similarity varies with first language experience. *Journal of Memory and Language*, *57*(1), 65–80. https://doi.org/10/fp6mz7

Beyersmann, E., Casalis, S., Ziegler, J. C., & Grainger, J. (2015). Language proficiency and morpho-orthographic segmentation. *Psychonomic Bulletin & Review*, *22*(4), 1054–1061. https://doi.org/10.3758/s13423-014-0752-9

Beyersmann, E., Cavalli, E., Casalis, S., & Colé, P. (2016). Embedded stem priming effects in prefixed and suffixed pseudowords. *Scientific Studies of Reading*, *20*(3), 220–230. https://doi.org/10.1080/10888438.2016.1140769

Bialystok, E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis*. Erlbaum.

Blom, E., Polisšenská, D., & Weerman, F. (2006). Effects of age on the acquisition of agreement inflection. *Morphology*, *16*(2), 313–336. https://doi.org/10/c56gqf

Bosch, S., Veríssimo, J., & Clahsen, H. (2019). Inflectional morphology in bilingual language processing: An age-of-acquisition study. *Language Acquisition*, *26*(3), 339–360. https://doi.org/10.1080/10489223.2019.1570204

Boudelaa, S., & Marslen-Wilson, W. D. (2005). Discontinuous morphology in time: Incremental masked priming in Arabic. *Language and Cognitive Processes*, *20*(1-2), 207–260. https://doi.org/10.1080/01690960444000106

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10/gddxwp

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/10/gfxzpn

Ciaccio, L. A., & Clahsen, H. (2020). Variability and consistency in first and second language processing: A masked morphological priming study on prefixation and suffixation. *Language Learning*, *70*(1), 103–136. https://doi.org/10/ggvfr9

Ciaccio, L. A., Kgolo, N., & Clahsen, H. (2020). Morphological decomposition in Bantu: A masked priming study on Setswana prefixation. *Language, Cognition and Neuroscience*, 1–15. https://doi.org/10.1080/23273798.2020.1722847

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.

Crepaldi, D., Rastle, K., & Davis, C. J. (2010). Morphemes in their place: Evidence for position-specific identification of suffixes. *Memory & Cognition*, *38*(3), 312–321. https://doi.org/10.3758/MC.38.3.312

Diependaele, K., Duñabeitia, J. A., Morris, J., & Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language*, *64*(4), 344–358. https://doi.org/10/fjghsn

Feldman, L. B. (2000). Are morphological effects distinguishable from the effects of shared meaning and shared form? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1431–1444. https://doi.org/http://dx.doi.org/10.1037/0278-7393.26.6.1431

Feldman, L. B., Kostić, A., Basnight-Brown, D. M., Đurđević, D. F., & Pastizzo, M. J. (2010). Morphological facilitation for regular and irregular verb formations in native and non-native speakers: Little evidence for two distinct mechanisms. *Bilingualism: Language and Cognition*, *13*(02), 119. https://doi.org/10.1017/S1366728909990459

Feldman, L. B., & Soltano, E. G. (1999). Morphological priming: The role of prime duration, semantic transparency, and affix position. *Brain and Language*, *68*(1-2), 33–39. https://doi.org/10.1006/brln.1999.2077

Foote, R. (2017). The storage and processing of morphologically complex words in L2 Spanish. *Studies in Second Language Acquisition*, *39*(4), 735–767. https://doi.org/10.1017/S0272263115000376

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 116–124. https://doi.org/10/d5dfns

Grainger, J., & Beyersmann, E. (2017). Edge-aligned embedded word activation initiates morpho-orthographic segmentation. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 67, pp. 285–317). Academic Press. https://doi.org/10.1016/bs.plm.2017.03.009

Günther, F., Smolka, E., & Marelli, M. (2019). "Understanding" differs between English and German: Capturing systematic language differences of complex words. *Structure in Words: The Present and Future of Morphological Processing in a Multidisciplinary Perspective*, *116*, 168–175. https://doi.org/10.1016/j.cortex.2018.09.007

Hasenäcker, J., Beyersmann, E., & Schroeder, S. (2016). Masked morphological priming in German-speaking adults and children: Evidence from response time distributions. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00929

Heathcote, A., Popiel, S. J., & Mewhort, D. J. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, *109*(2), 340–347. https://doi.org/10.1037/0033-2909.109.2.340

Heathcote, L., Nation, K., Castles, A., & Beyersmann, E. (2018). Do "blacheap" and "subcheap" both prime "cheap"? An investigation of morphemic status and position in early visual word processing. *Quarterly Journal of Experimental Psychology*, 17470218.2017.1. https://doi.org/10.1080/17470218.2017.1362704

Heyer, V., & Clahsen, H. (2015). Late bilinguals see a scan in scanner AND in scandal: Dissecting formal overlap from morphological priming in the processing of derived words. *Bilingualism: Language and Cognition*, *18*(03), 543–550. https://doi.org/10.1017/S1366728914000662

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56. https://doi.org/10/f4rnbm

Jacob, G., Heyer, V., & Veríssimo, J. (2018). Aiming at the same target: A masked priming study directly comparing derivation and inflection in the second language. *International Journal of Bilingualism*, *22*(6), 619–637. https://doi.org/10/ggvfsr

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. https://doi.org/10/cfxkgw

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

Kinoshita, S., & Lupker, S. J. (Eds.). (2004). *Masked priming: The state of the art*. Psychology Press.

Kirkici, B., & Clahsen, H. (2013). Inflection and derivation in native and non-native language processing: Masked priming experiments on Turkish. *Bilingualism: Language and Cognition*, *16*(4), 776–791. https://doi.org/10/ggvjsg

Kneib, T., & Umlauf, N. (2017). *A primer on Bayesian distributional regression* (Working Papers in Economics and Statistics Nos. 2017-13). University of Innsbruck, Research Platform Empirical and Experimental Economics (eeecon).

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10/c9f897

Li, J., Taft, M., & Xu, J. (2017). The processing of English derived words by Chinese-English bilinguals. *Language Learning*, *67*(4), 858–884. https://doi.org/10.1111/lang.12247

Lindley, D. V. (1970). *Introduction to probability and statistics from a Bayesian viewpoint. Part 2: Inference* (2nd ed.). Cambridge University Press.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 883–914. https://doi.org/10/bwv74n

Longtin, C.-M., & Meunier, F. (2005). Morphological decomposition in early visual word processing. *Journal of Memory and Language*, *53*(1), 26–41. https://doi.org/10.1016/j.jml.2005.02.008

Longtin, C.-M., Segui, J., & Hallé, P. A. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes*, *18*(3), 313–334. https://doi.org/10.1080/01690960244000036

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (Second). CRC Press.

Nicenboim, B., Logačev, P., Gattei, C., & Vasishth, S. (2016). When high-capacity readers slow down and low-capacity readers speed up: Working memory and locality effects. *Frontiers in Psychology*, *7*. https://doi.org/10/ggks56

Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes*, *15*(4-5), 507–537. https://doi.org/10.1080/01690960050119689

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11*(6), 1090–1098. https://doi.org/10/ddq274

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rodgers, D. M. (2011). The automatization of verbal morphology in instructed second language acquisition. *IRAL - International Review of Applied Linguistics in Language Teaching*, *49*(4). https://doi.org/10.1515/iral.2011.016

Rouder, J. N. (2005). Are unshifted distributional models appropriate for response time? *Psychometrika*, *70*(2), 377–381. https://doi.org/10/c9cx9s

Rouder, J. N., Haaf, J. M., & Vandekerckhove, J. (2018). Bayesian inference for psychology, part IV: Parameter estimation and Bayes factors. *Psychonomic Bulletin & Review*, *25*(1), 102–113. https://doi.org/10/gc9qfx

Schad, D. J., Betancourt, M., & Vasishth, S. (2020). Toward a principled Bayesian workflow in cognitive science. *Psychological Methods*. https://doi.org/10.1037/met0000275

Segalowitz, N. (2008). Automaticity and second languages. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (Vol. 27, pp. 382–408). Blackwell.

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, *14*(3), 369–385. https://doi.org/10.1017/S0142716400010845

Segalowitz, S. J., Segalowitz, N. S., & Wood, A. G. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics*, *19*(1), 53–67. https://doi.org/10.1017/S0142716400010572

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language & Cognition*, *11*, 245–260. https://doi.org/10.1017/S1366728908003404

Solovyeva, K., & DeKeyser, R. (2018). Response time variability signatures of novel word learning. *Studies in Second Language Acquisition*, *40*(1), 225–239. https://doi.org/10.1017/S0272263117000043

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 638–647. https://doi.org/10.1016/S0022-5371(75)80051-X

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10/ggrw22

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, *71*, 147–161. https://doi.org/10/gfzq3c

Veríssimo, J., Heyer, V., Jacob, G., & Clahsen, H. (2018). Selective effects of age of acquisition on morphological priming: Evidence for a sensitive period. *Language Acquisition*, *25*(3), 315–326. https://doi.org/10/ggffzk

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, *114*(3), 830–841. https://doi.org/10/btnjz6

White, L. (2003). *Second language acquisition and Universal Grammar* (First). Cambridge University Press. https://doi.org/10.1017/CBO9780511815065

Yap, M. J., Balota, D. A., Cortese, M. J., & Watson, J. M. (2006). Single- versus dual-process models of lexical decision performance: Insights from response time distributional analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1324–1344. https://doi.org/10.1037/0096-1523.32.6.1324

Appendix

Table A1

Prior distributions on each estimate of the Bayesian distributional model

Parameter	Type	Coefficient	Prior	Units
Mu (mean)	Fixed effect	Intercept	Normal(6.0, 1)	Log-ms
Mu (mean)	Fixed effect	Slopes	Normal(0, 0.25)	Log-ms
Mu (mean)	Random effect	Intercept	Normal(0, 0.5)	Log-ms
Mu (mean)	Random effect	Slopes	Normal(0, 0.25)	Log-ms
Sigma (SD)	Fixed effect	Intercept	Normal(-3, 2)	Log of log-ms
Sigma (SD)	Fixed effect	Slopes	Normal(0, 0.25)	Log of log-ms
Sigma (SD)	Random effect	Intercept	Normal(0, 0.5)	Log of log-ms
Sigma (SD)	Random effect	Slopes	Normal(0, 0.25)	Log of log-ms
Shift			Normal(250, 250)	ms

Note. Prior distributions are in the notation Normal(mean, SD). These are expressed in different units due to the parametrizations used by the brms R-package: estimates on mu (mean) are in log-ms because a (shifted-)lognormal response distribution was assumed; estimates on sigma (SD) are modelled in the log-scale by default because they must be strictly positive (hence, in log of log-ms). Note that all effects on sigma that are reported in the paper were back-transformed to log-ms, for easier interpretation and for consistency with the other estimates.