Quaderni di Linguistica e Studi Orientali- QULSO
Working papers in Linguistics and Oriental Studies

Firenze University Press
https://oajournals.fupress.net/index.php/bsfm-qulso

Citation: R. Petrosino, J. Sprouse, D. Almeida (2023) Asymmetries in the stem and suffix masked priming response in a large-scale online study. *Qulso* 9: pp. 177-194. doi: http://dx.doi.org/10.13128/QULSO-2421-7220-15154

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Asymmetries in the Stem and Suffix Masked Priming Response in a Large-Scale Online Study

Roberto Petrosino, Jon Sprouse, Diogo Almeida
New York University Abu Dhabi
(<robpetro@nyu.edu>; <jon.sprouse@nyu.edu>; <diogo@nyu.edu>)

Abstract:

Models of visual word processing incorporate morphological decomposition as a step in the recognition process, but they vary as to when this step happens, and what kind of information is used in it. In particular, the affix stripping model proposes that words are accessed through their stems, after affixes are automatically stripped off. This dichotomy between stems and affixes seems to be mirrored in masked priming. Masked stem priming is quite robust and comparable to masked repetition priming, whereas masked affix priming is often null, or very small. However, the literature on masked affix priming is much smaller than the one on masked stem priming. This study investigates the stem vs suffix asymmetry in masked priming by running an online experiment with a large sample size (N=161) to ensure higher statistical power. For comparison and validation, the same experiment was also conducted in a standard lab setting. In addition, we ran a follow-up experiment with two additional suffix priming conditions with an even larger sample size (N=400) to assess the influence of orthographic and strategic confounds. The three experiments show significant stem priming, but null or very small suffix priming, thus supporting the asymmetry between stem activation and affix stripping.

Keywords: *Affix Stripping, Lexical Access, Morphological Decomposition, Suffix Priming, Visual Word Processing*

1. Introduction

Models of visual word processing often posit that morphologically complex lexical entries are decomposed into their morphemic constituents in the process of lexical access, although details vary as to whether the decomposition happens early or late in the recognition process (Baayen *et al*. 1997; Taft 2004), and whether morphological decomposition is blind to semantics or not (Feldman *et al*. 1999; Rastle et al. 2000; Feldman *et al*. 2004). Evidence for an early, semantic-blind decomposition stage as claimed in the affix stripping model decomposition (Taft 1994) comes from a series of masked priming findings wherein a monomorphemic word (called *target*: e.g., *DRIVE*) is recognized faster when preceded by a morphologically related word (called *prime*: e.g., *driver*) that is masked (i.e., visually presented for such a short time, 40-60 ms, so that subjects do not consciously perceive it), but also wherein a monomorphemic target (e.g., *BROTH*) is recognized faster when preceded by a seemingly, but not actually morphologically related prime (morphologically "opaque" words; e.g., *brother*, which does not mean "someone who broths", even though it displays the *-er* agentive suffix). Crucially, a similar facilitation fails to obtain when the target (*BROTH*) is only orthographically similar, with no possible morphological parse of the prime (e.g., *brothel*, where *-el* is not an extant English suffix; Rastle et al. 2004). This pattern of results has been argued to support a procedure of MORPHOLOGICAL DECOMPOSITION (henceforth, MD) occurring early in word processing and prior to lexical access, as it seems to rely on the morpho-orthographic form of morphemes (that is, the phonological realization and the orthographic sequence of letter strings associated with a given morpheme; in this sense, *-er* would elicit decomposition, but *-el* would not), and does not depend on semantic interpretation (since *-er* elicits decomposition even in morphologically opaque, monomorphemic words like *brother*). However, the affix stripping model makes the extra assumption that only *stems* are used for lexical search, whereas affixes are just stripped off (Taft and Forster 1975; Taft 1981). This would predict that masked affix priming may either not occur or be very weak under the same circumstances where masked stem priming is robust. This prediction seems to be borne out. On the one hand, the literature of masked stem priming has reported robust effects across languages, even with different systems of word formation (e.g., Semitic languages: Frost et al. 2001; Boudelaa and Marslen-Wilson 2001). On the other hand, masked affix priming, while comparatively understudied, provides less robust results across different languages (among others, English: Crepaldi *et al.* 2016, Petrosino 2020; French: Giraudo and Grainger 2003; Italian: Giraudo and Dal Maso 2016; Spanish: Dominguez *et al.* 2010, Andoni Duñabeitia *et al.* 2008; for a review, Amenta and Crepaldi 2012). However, while masked affix priming results seem to be less robust than masked stem priming, it is hard to ascertain whether they are really null, as the affix stripping model would predict, or just much significantly smaller than stem priming effects. The results in the literature could be used to support either conclusion, and there is an unfortunate trend in the literature to consider small effect sizes as null if they are not statistically significant, which is never a warranted conclusion, even when statistical power is taken in consideration (which is rarely done anyway).

Here, we turn to internet-based online experimentation to provide a well-powered test for the asymmetry between affix masked priming response – in particular, the *suffix* masked priming response – compared to the stem masked priming response. Online experimentation has three main advantages: (i) an order of magnitude increase in the potentially recruitable sample size; (ii) dramatic decrease in the time needed to test a desired sample size; and (iii) easy access to populations that may not be easily available in lab based experiments. In recent years, the number of cognitive science labs and departments capitalizing on the HTML5 capabilities has rapidly increased, also thanks to the proliferation of stable and powerful software packages for stimulus delivery and data collection. Here we provide a test of *Labvanced* (Finger *et al.* 2017), a GUI-based web app that allows researchers to dispense with local installation issues (thus preventing potential incompatibilities with the CPU of the local machine and ensuring cross-platform consistency) and yet another programming language to learn (thus facilitating experimental designing and deployment). The same experiment was conducted online with a larger sample size than ever reported before (experiment 1a; N=161). At the same time, the very same experiment was also conducted in the traditional setting of a controlled room, albeit with a smaller sample size (*N*=58), as a way to further assess the reliability of the online environment (experiment 1b). Both experiments show comparable results, and in particular no priming effects in the affix priming conditions tested. We then carried out a follow-up study, in which we tested additional affix priming conditions (experiment 2), with an even larger sample size (*N*=400). The combined results show little to no evidence of there being masked suffix priming effects, and therefore seem to be compatible with an affix-stripping model of MD.

2. Experiment 1

In experiment 1, we tested the suffix masked priming response to bimorphemic word targets preceded by bimorphemic word primes sharing the same suffix. We tested two suffix priming conditions, each with the same suffix recurring throughout. This was done to assess the potential decomposition and activation of each suffix, rather than mixing them up in a single condition (as usually done in stem priming conditions, in which words with different stems and affixes are grouped together). We chose two of the most productive English suffixes: *-er* (agentive suffix; derivational morpheme), and *-s* (plural suffix; inflectional morpheme). All words tested were nouns, to avoid the potential overlap in the activation of homophonic morphemes (i.e., *-er* is also the surface representation of the comparative suffix; *-s* is also the surface representation of the 3SG.PRES verbal suffix). As a way to assess the effect size of the suffix priming response, we additionally tested the typical spectrum of relatedness between prime and target words reported in the literature: repetition or identity (in which the same word is presented as both prime and target), morphological (in which the target is the morphological stem of the corresponding prime), orthographic (in which prime and target words are orthographically similar, but not related morphologically nor semantically), and semantic (in which prime and target words are semantically connected, but not related orthographically nor morphologically). The same experiment was conducted both online (experiment 1a) and in lab (experiment 1b), so to (i) validate the experimental results of the data collected online and (ii) assess replicability across different data collection settings and sample sizes.

2.1 Methods

Two-hundred and forty words were selected from the English Lexicon Project database (ELP; Balota *et al*. 2007) to construct twenty word pairs of each of the six conditions tested. In the identity condition, the same monomorphemic word was presented as both prime and target (e.g., *shrimp-SHRIMP*). In the stem condition, a bimorphemic word was presented as prime and its corresponding stem as a target (*boneless-BONE*). In the plural condition, regular plural nouns were presented (*worlds-HEAVENS*). In this condition, we made sure that all plural forms had the [z] allomorph of the suffix to avoid potential confounds due to the phono-orthographic differences with the other allomorphs (i.e., [ɪz, s]). We also made sure that all plural forms were less frequent than the corresponding singular forms, to avoid potential confounding effects due to relative whole-word frequency (*dominance*; Baayen *et al*. 1997). In the *er*-condition, words ending with the suffix *-er* were presented (*driver-RUNNER*). In the semantic condition, monomorphemic word pairs involving a semantically (but not morphologically or orthographically) transparent relationship were presented (*squid-OCTOPUS*). Finally, in the rhyme condition word pairs sharing the rightmost phono-orthographic portion were presented (*casket-BASKET*). Because we are interested in the priming effects within each condition, it was not necessary to control lexical properties such as frequency (HAL) and orthographic length across all conditions. The conditions were split into two different groups. The first group of conditions included the identity, semantic, and plural conditions, since we wanted to compare the priming effects to the first two conditions (which we expect to elicit large and small-to-null effects, respectively) to the plural condition (which is one of the crucial experimental conditions tested here). We matched the frequency of prime words of the plural and semantic conditions, and the target words of the plural, identity, and semantic conditions (primes: $F(2, 57)=0.878$, $p=.42$; targets: $F(3, 76)=0.784$, $p=.51$). Orthographic length was not controlled due to interdependency between the several constraints applied in word selection. For this reason, the length of the prime words significantly varied across conditions, but it did not after removing the semantic condition ($F(1, 38)=0.018$, $p=.89$). The length of all target words did not vary across conditions ($F(3, 76)=1.90$, $p=.14$). The second group of conditions included the stem, *er*-, and rhyme condition, as a way to gauge potential asymmetries in the priming response to stem and suffix priming; the rhyme condition was tested to parse out the potential effects of phono-orthographic priming. While frequency did not vary across conditions (primes: $F(2, 57)=1.50$, $p=.23$; targets: $F(2, 57)=0.008$, $p=.99$), orthographic length could not be controlled across conditions due to the idiosyncrasies of the conditions tested: the *er*-condition consisted of bimorphemic primes and targets, the rhyme condition consisted of monomorphemic primes and targets, and the stem condition consisted of bimorphemic primes and monomorphemic targets. Descriptive statistics of the lexical properties of each condition are reported in Table 1 below. An additional subset of one hundred and twenty words that were unrelated to each target along all possible dimensions were also chosen from the ELP and were used as unrelated primes. They were matched with the corresponding related prime as much as possible in frequency ($t(139)=-0.43$, $p=.66$) and orthographic length ($t(139)=-1.64$, $p=.10$). The complete word list is reported in the appendix.

	item type	condition	HAL log frequency	orthographic length
GROUP 1	related primes	plural	6.74 (1.90)	7.4 (1.05)
		semantic	7.37 (1.49)	5.85 (2.08)
	targets	plural	7.34 (1.50)	6.8 (1.20)
		identity	7.52 (0.33)	6.45 (0.51)
		semantic	7.30 (0.61)	6 (1.12)
GROUP 2	related primes	stem	6.72 (0.55)	6.3 (0.80)
		er	6.50 (0.11)	6.55 (0.76)
		rhyme	6.46 (0.694)	5.85 (0.49)
	targets	stem	8.44 (1.24)	4 (0.73)
		er	8.41 (0.543)	6.5 (0.69)
		rhyme	8.43 (1.24)	5.95 (0.61)

Table 1. Mean and sd (in parenthesis) of the lexical properties of the items selected for experiment

Finally, one-hundred and twenty non-words were selected from the ELP database, so that they matched in length with the word target items of the first group condition (mean: 6.22, sd: 1.30; $F(4, 215)=1.71$, $p=.15$). One hundred and twenty bimorphemic words from the ELP (different from the words used in the word-word conditions described above) were selected as unrelated primes of the non-word target items.

We then constructed two different word lists. In one list, the six word conditions had half of the targets preceded by the corresponding related prime; and the other half preceded by the unrelated prime. In the other list, the order was reversed. The two lists presented the same set of target words and non-words (120 word-word pairs + 120 word-nonword pairs = 240 pairs in total). Labvanced automatically assigned each participant a list, so that all participants would see the same target items, but differed in the list being assigned.

2.2 Procedure

2.2.1 Experiment 1a (online)

Experiment 1a was conducted online. One hundred and sixty-one native English speakers (71 females; age mean: 41.32, age sd: 13.73) were recruited on Cloud Research and Prolific. All participants were located in the U.S., had English as their first and only language; none of them reported any neuro-cognitive impairment. Participants were asked to perform a lexical decision task on Labvanced by pressing keys on their keyboard. Each trial consisted of three different stimuli appearing at the center of the screen: a series of hashes (#####) presented for 500 ms, followed by a prime word and finally the corresponding target word, which disappeared from the screen at button press. The prime duration was set to 33 ms because a pilot study had shown that setting prime duration to a longer duration would lead to a higher number of trials with the prime duration fluctuating beyond the subliminal threshold (>60 ms). Participants were also given a few breaks throughout the experiment to avoid exertion. The experiment lasted around 11 minutes on average.

2.2.2 Experiment 1b (in-lab)

Experiment 1b was conducted in a sound-shielded room at New York University Abu Dhabi. Fifty-five native English speakers (39 females; age mean: 28.27, age sd: 9.73) participated in the experiment with the same inclusion criteria indicated above. The task was the same as described above for experiment 1a, with the only difference being that the items were presented on PsychoPy (Pierce *et al*. 2019). Each participant received either course credit or a gift voucher.

2.3 Analysis & results

The two datasets were analyzed separately as follows. We will discuss the results of both experiments in a single section below.

2.3.1 Experiment 1a (online)

Prior to analysis, the dataset (consisting with a total of 45,080 observations) went through three steps to remove outliers. Current online stimulus delivery programs offer sub-optimal timing precision, which may dramatically depend on a number of uncontrollable variables, such as the system specifications (operative system, CPU), the browser used by each participant, as well as the number and the type of programs remaining active in the background.

Therefore, as a first step we removed all trials in which the duration of the prime word was out of an acceptable range to elicit masked priming. The target duration was 33 ms (corresponding to 2 refresh cycles in a standard 60-Hz monitor). The upper bound was set to 60 ms, as priming subliminality has been shown to arise as long as the prime item is presented below that threshold (Forster 1999; Forster *et al*. 2003). This meant that we kept trials with a prime word being presented for up to 3 complete refresh cycles in a standard 60-Hz monitor (roughly corresponding to 48 ms). The lower bound was set to 25 ms, which is half a cycle (i.e., 8 ms) away from the target duration; this was done to remove trials with subliminally undetectable primes. This meant that 15% of the trials were removed from analysis. In the second step, we removed 4 subjects whose error performance was above 30%. We also removed 2 items whose overall error rate was above 30% (*dogmas, wean*). After this step, we removed one extra subject because they had at least one condition with a number of trials that was lower than 5. Finally, in the third step, we removed 0.62% trials with RTs that were below 200 ms or above 1800 ms (as suggested by Ratcliff 1993). The final shape of the dataset had 20,460 observations (word trials only) and 157 subjects.

Priming calculations and statistical analyses involved the word trials only. For each condition, we calculated priming as the difference between the RTs to related trials and RTs to unrelated trials. The mean response time (RT), standard deviations (SD), Pearson's *r*, raw priming magnitudes, and standardized effect sizes (ES; Cohen's *d*) of the conditions tested are reported in Table 2 below. As Figure 1 shows, the identity and stem conditions elicited medium-size priming (~20 ms), the rhyme condition elicited small priming (10 ms), and the semantic condition negative priming (-3 ms). Both suffix conditions (the plural and the *er*-conditions) elicited small priming effects (3 and 5 ms, respectively), thus being more comparable to semantic effects than to morphological effects.

	CONDITION	EXAMPLE	MEAN RELATED RT (SD)	MEAN UNRELATED RT (SD)	*PEARSON'S* CORRELATION	PRIMING (SD)	ES
GROUP 1	identity	*glance-GLANCE*	642 (110)	661 (110)	0.79	19 (72)	0.26
	plural	*worlds-HEAVENS*	677 (113)	680 (112)	0.76	3 (79)	0.04
	semantic	*squid-OCTOPUS*	659 (111)	656 (107)	0.84	-3 (61)	-0.05
GROUP 2	stem	*driver-DRIVE*	622 (100)	640 (99)	0.79	18 (65)	0.28
	er	*driver-RUNNER*	635 (109)	640 (98)	0.80	5 (66)	0.08
	rhyme	*casket-BASKET*	634 (107)	653 (96)	0.77	10 (69)	0.15

Table 2. Summary of the priming effects elicited in experiment 1a.

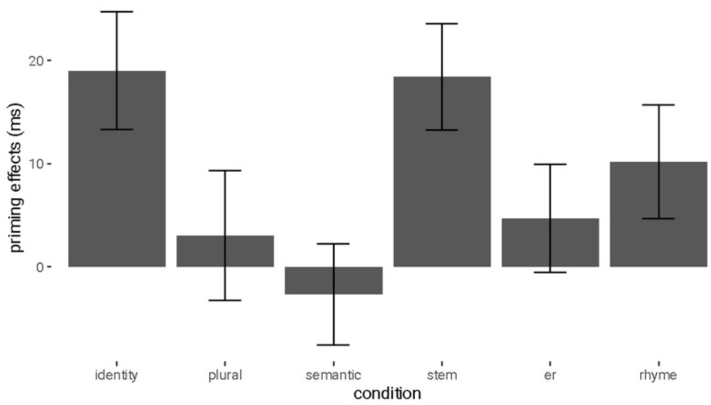

Figure 1. Priming effects for experiment 1a. Error bars represent one standard error from the mean in either direction.

For each condition we ran a *t*-test for RELATEDNESS (2 levels: related vs. unrelated). The details of the statistical results (*t*-values, dfs, and *p*-values) are reported in Table 3 below. The identity and stem conditions elicited significant priming, and the semantic condition did not. These results are reassuring, in that they are in line with the previous literature on masked priming and validate the reliability of the results of the other conditions. The rhyme condition elicited trending-to-significant priming effects, whereas neither suffix priming condition (i.e., the plural condition nor the *er*-condition) elicited significant effects.

	CONDITION	*T* (DF)	*P*-VALUE	SIGN.
GROUP 1	identity	-3.30 (2963)	0.001	*
	plural	-0.489 (2743)	0.625	n.s.
	semantic	0.449 (2937)	0.653	n.s.
GROUP 2	stem	-3.22 (2781)	0.001	*
	er	-1.05 (3016)	0.294	n.s.
	rhyme	-1.85 (2975)	0.065	(*)

Table 3. Summary of the statistical results for experiment 1a.

2.3.2 Experiment 1b (in-lab)

Prior to analysis, the dataset (consisting with a total of 15,400 observations) went through the same outlier rejection pipeline described for experiment 1a. Experiment 1b was carried out in the lab, so the prime duration could be maintained constant at exactly 33 ms. Therefore, no trial had to be removed because of out-of-range durations. None of the subjects had an error score higher than 30%, so no subject was removed from analysis because of it. We removed three items whose overall error rate was above 30% (*dogmas, vane, wean*). Finally, we removed 0.33% trials with RTs that were below 200 ms or above 1800 ms. The final shape of the dataset had 7,236 observations (word trials only) and 55 subjects.

Priming calculations and statistical analyses involved the word trials only and were performed in the same way as described above. The mean RTs, standard deviations, Pearson's *r*, raw priming magnitudes, and standardized effect sizes (ES; Cohen's *d*) of the conditions tested are reported in Table 4 below. As Figure 2 shows, the identity and stem conditions elicited medium-size positive priming (30 and 13 ms), along with the rhyme condition (19 ms), and the semantic condition negative small priming (-7). Both suffix conditions (the plural and the *er*-conditions) elicited small priming effects (2 ms for each), thus being more comparable to semantic effects than to morphological effects.

	CONDITION	EXAMPLE	MEAN RELATED RT (SD)	MEAN UNRELATED RT (SD)	*PEARSON'S* CORRELATION	PRIMING (SD)	ES
GROUP 1	identity	*glance-GLANCE*	617 (86)	647 (77)	0.56	30 (77)	0.39
	plural	*worlds-HEAVENS*	648 (92)	650 (74)	0.49	2 (85.7)	0.02
	semantic	*squid-OCTOPUS*	645 (82)	638 (74)	0.72	-7 (59)	-0.12
GROUP 2	stem	*driver-DRIVE*	616 (95)	629 (75)	0.59	13 (79)	0.17
	er	*driver-RUNNER*	625 (79)	627 (82)	0.54	2 (77)	0.03
	rhyme	*casket-BASKET*	619 (81)	638 (92)	0.77	19 (59.7)	0.32

Table 4. Summary of the priming effects elicited in experiment 1b.

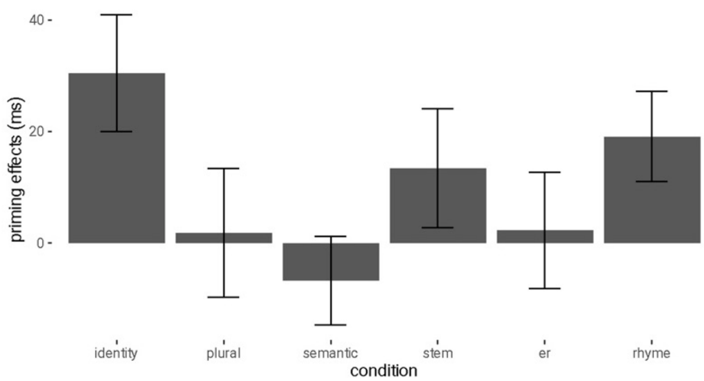

Figure 2. Priming effects for experiment 1b.
Error bars represent one standard error from the mean in either direction.

Statistical analyses were performed in the same way as described for experiment 1a. The details of the statistical analysis are reported in Table 5 below. The identity elicited significant priming. The other conditions did not.

	CONDITION	*T* (DF)	*P*-VALUE	SIGN.
GROUP 1	identity	-3.14 (1059)	0.002	*
	plural	-0.2 (970)	0.841	n.s.
	semantic	0.711 (1040)	0.477	n.s.
GROUP 2	stem	-1.27 (920)	0.203	n.s.
	er	-0.205 (1065)	0.838	n.s.
	rhyme	-1.76 (1060)	0.079	n.s.

Table 5. Summary of the statistical results for experiment 1b.

2.4 Discussion

The goal of Experiment 1 was to address the asymmetry between the (consistently large) stem priming effects and the unreliable affix priming effects, by eliciting them both in an experiment with a larger-than-usual sample size. As a way to further assess the reliability and replicability of the data collected independently of the setting, we ran the same experiment both online (experiment 1a) and in-lab (experiment 1b). In both experiments, the two suffix priming conditions (i.e., the *er*- and the plural conditions) showed no significant priming along with the semantic condition and the rhyme condition. The identity condition elicited significant priming effects in both experiments, whereas stem priming was not significant in experiment 1b. The unexpected null effect (and small effect size) for stem priming in experiment 1b was likely due to the low sample size and the short prime duration.

The general consistency of the results for these conditions provides further evidence for the asymmetry reported in the literature between stem and affix priming: the average stem priming effect in the two experiments was 15.5ms, whereas the average suffix priming effect in the two experiments was five times smaller, only 3 ms. This is compatible with the proposal from the affix stripping model of lexical access. Nonetheless, two potentially interacting confounding factors were identified that could explain the absence of suffix priming in this experiment. First, the numerically much smaller effect size in the suffix priming conditions could be due to the unusually short prime duration of 33 ms. Perhaps the morphological parser does not have the time necessary to get to the end of the suffixed prime word (i.e., where the suffixal orthographic unit occurs) and activate the corresponding lexical entry, thus inhibiting suffix priming. Second, the lack of suffix priming effects could just be due to the minimal morpho-orthographic overlap within the pairs of these conditions. The two suffix priming conditions only involve a one-letter (*-s*) and a two-letter (*-er*) suffixes, respectively. Therefore, such minimal overlap might have just been not enough for priming to arise in these conditions, especially at such a short prime duration. Similarly, the unexpected significant rhyme priming effects could be explained as triggered by the overly large proportion of orthographic overlap in the rhyme condition. For this reason, we ran a follow-up experiment that could overcome such potential confounds.

3. *Experiment 2 (online)*

Experiment 2 explores the extent to which either (or both) of the two confounds identified above might have impinged on the results of experiment 1. Both the short prime duration and the minimal overlap size might have somehow brought about suffix priming inhibition. In this experiment, we made three crucial modifications from experiment 1. First, we set a longer prime duration (48 ms, roughly equal to three full cycles in a standard 60-Hz monitor), to provide the parser more time to fully process the prime while still ensuring a subliminal prime presentation. Second, we tested two additional suffix conditions involving the two derivational suffixes that are comparably productive, but orthographically longer than the two used in the experiment 1 (4 letter each): *-able* and *-ness*. This was done to gauge the extent to which the orthographic overlap size impinges on the priming response. Should either/both confounds be playing a role in masked priming elicitation, priming would arise in all, or at least in a subset of the four suffix priming conditions tested. Finally, to further enhance the statistical power of the experiment and therefore the reliability of the results, we conducted the experiment online and increased the sample size.

3.1 Methods & procedure

Materials and procedure were the same as in experiments 1a-b. In addition to the six conditions above, we prepared two more suffix priming conditions, consisting of 20 related word pairs each. In the *able*-condition, prime and target words shared the suffix *-able* (*washable-NOTABLE)*; in the *ness*-condition, prime and target words shared the suffix *-ness* (*weakness-SICKNESS*). The words were all chosen from the ELP database. The prime and the targets were matched across the two conditions in frequency (primes: $F(1, 38)=0.024$, $p=.88$; targets: $F(1, 38)=2.22$, $p=.15$) and length (primes: $F(1, 38)=0.18$, $p=.66$; targets: $F(1, 38)=0.02$, $p=.90$), but could not be matched with any of the other conditions tested, in particular because of the different orthographic length of the suffixes involved. Therefore, these two conditions were constructed as an additional, separate group (see Table 2 for the descriptive statistics). Forty unrelated bimorphemic words were used as unrelated primes. They were matched with the corresponding related prime as much as possible in frequency ($t(139)=-0.43$, p=.66) and orthographic length ($t(139)=-1.64$, p=.10). The words of these two conditions are reported in the appendix below. The same word-nonwords pairs used in experiment 1 and 2 were also used for experiment 3, while adding forty more pairs of the same length. The two lists were constructed as described above, with a total of 280 pairs for each list.

	ITEM TYPE	CONDITION	HAL LOG FREQUENCY	ORTHOGRAPHIC LENGTH
GROUP 3	related primes	*able*	6.43 (1.82)	9.11 (0.99)
		ness	6.51 (1.57)	8.95 (1.28)
	targets	*able*	8.39 (1.04)	8.7 (1.38)
		ness	7.82 (1.34)	8.5 (0.95)

Table 6. Mean and sd (in parenthesis) of the lexical properties of the items selected for experiment 2.

The experiment procedure was the same as reported for experiment 1a, with the only difference that the target prime duration was longer and set to 48 ms (corresponding to three full refresh cycles in a standard 60-Hz monitor), instead of 33 ms. Four hundred native U.S. English speakers (200 female; age mean: 42.4; age sd: 13.1) were recruited on Prolific with the same inclusion criteria as indicated for experiment 1a. The average completion time was about 13 minutes, and participants were compensated for their participation.

3.2 Analysis & results

Prior to analysis, the dataset (consisting with a total of 108,000 observations) went through the same three outlier detection steps described above. First, 24% of the trials had a prime duration below 30 ms or above 60 ms and thus had to be removed from analysis. Second, 5 subjects were removed because their error performance was above 30%. We also removed 3 items because their overall error rate was above 30% (*dogmas, dilemmas, wean*). Finally, we removed 0.87% of the trials whose RT was below 200 ms or above 1800 ms. This left 37,623 observations (word trials only) and 295 subjects suitable for analysis.

Priming calculations and statistical analyses involved the word trials only as described above. The mean RTs, raw priming magnitudes, standard deviation, and standardized effect sizes (ES; Cohen's *d*) of the conditions tested are reported in Table 7 below. As Figure 3 shows, the identity and stem conditions elicited medium-size priming (47 and 31 ms, respectively), the rhyme condition elicited medium priming (10 ms), and the semantic condition elicited small priming (7 ms). All suffix priming conditions elicited small priming effects (ranging between -3 and 5 ms).

	CONDITION	EXAMPLE	MEAN RELATED RT (SD)	MEAN UNRELATED RT (SD)	*PEARSON'S* CORRELATION	PRIMING (SD)	ES
GROUP 1	identity	*glance-GLANCE*	635(127)	682 (120)	0.78	47 (83)	0.57
	plural	*worlds-HEAVENS*	682 (128)	687 (123)	0.72	5 (94)	0.05
	semantic	*squid-OCTOPUS*	662 (114)	669 (115)	0.72	7 (86)	0.08
GROUP 2	stem	*driver-DRIVE*	619 (107)	650 (105)	0.73	31 (79)	0.39
	er	*driver-RUNNER*	651 (108)	648 (107)	0.78	-3 (72)	-0.04
	rhyme	*casket-BASKET*	634 (112)	653 (112)	0.79	10 (74)	0.17
GROUP 3	*able*	*washable-NOTABLE*	668 (115)	671 (117)	0.80	3 (74)	0.04
	ness	*weakness-SICKNESS*	674 (129)	679 (129)	0.64	5 (109)	0.05

Table 7. Summary of the priming effects elicited in experiment 2.

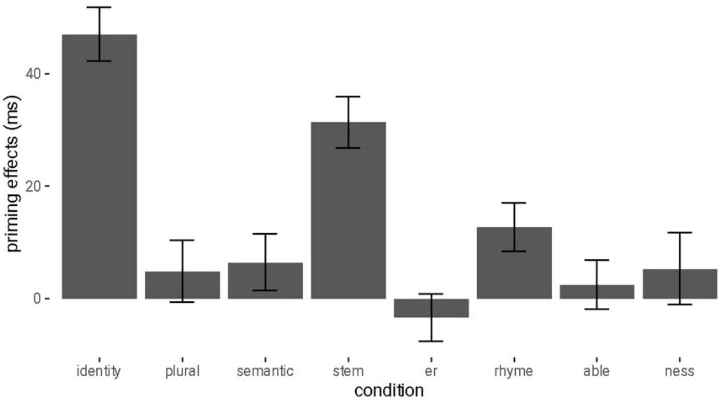

Figure 3. Priming effects for experiment 2.
Error bars represent one standard error from the mean in either direction.

Statistical analyses were performed separately for each of the 8 conditions as described above. The details of the analysis are reported in Table 8 below. The results of experiment 2 are essentially the same as the results of experiment 1.[1] The identity, stem, and rhyme conditions elicited significant priming. Along with the semantic condition, all suffix priming conditions (i.e., the plural, *er-, able-,* and *ness*-conditions) did not elicit significant priming effects.

	CONDITION	*T* (DF)	*P*-VALUE	SIGN.
GROUP 1	identity	-8.28 (4282)	< 1.59e-16	*
	plural	-0.568 (3731)	0.57	n.s.
	semantic	-1.47 (4239)	0.141	n.s.
GROUP 2	stem	-5.91 (3979)	3.8e-9	*
	er	0.06 (4352)	0.95	n.s.
	rhyme	-1.98 (4254)	0.048	*
GROUP 3	*able*	-0.237 (4407)	0.812	n.s.
	ness	-1.05 (4078)	0.292	n.s.

Table 8. Summary of the statistical results for experiment 2.

[1] We take the numerical and statistical discrepancies in the effects elicited across experiments as resulting from natural sampling variance, and therefore are not discussed further.

3.3 Discussion

Experiment 2 was specifically designed to address two potential confounding variables that could have played a role in the results of experiment 1. First, we increased the prime duration to 48 ms, to provide the parser more time to fully process the prime, while still ensuring its subliminal presentation. Second, we added two suffix priming conditions wherein suffixes were orthographically longer than those tested in experiment 1. Third, we increased the sample size to ensure higher statistical power. The results for this experiment were in line with the results of experiment 1: significant priming effects were found in the identity, stem, and rhyme conditions, but in none of the suffix priming conditions (the plural and *-er* conditions, as well as the *-able* and *-ness* conditions). The increase in priming duration (33ms to 48ms) did not increase the affix masked priming effects for the two conditions also tested in experiment 1 (M_{suffix} = 3 ms in experiment 1a-b, M_{suffix} = 1 ms in experiment 2, for the *-er* and *plural* conditions), unlike what was observed for instance in the identity condition ($M_{identity}$ = 25 ms in experiment 1a-b, $M_{identity}$ = 47 ms in experiment 2). The increase in orthographic length of the suffix also did not seem to increase the effect size of the masked suffix presentation: M_{suffix_short} = 3 ms (experiment 1) and 1 ms (experiment 2); M_{suffix_long} = 4ms (experiment 2). Thus, averaging all the effect sizes of suffix masked priming observed across all experiments here, we obtain an estimate of M_{suffix} = 2 ms, which can be safely discarded as either null or theoretically uninteresting. We take stock of all the results reported above, and discuss the theoretical implications thereby in the next section.

4. General discussion and conclusions

In this paper, we reported two experiments aimed at determining whether putative early and automatic morphological decomposition (Taft 2004; Rastle and Davis 2008) is differentially sensitive to stems and affixes. Previous studies have consistently reported such an asymmetry in masked priming studies: masked priming of stems lead to large effects, whereas masked priming of affixes have either null or small effects. The study reported here attempted to directly address this potential asymmetry by eliciting suffix priming on a larger sample size than ever recruited before (N>100) and directly compare it with other relevant effects, such as identity, stem, semantic and orthographic (rhyme) effects. In order to achieve these large sample sizes in a reasonable time frame, we chose to use *Labvanced*, a GUI-based stimulus delivery program that facilitates experiment implementation and deployment without having to use a specific programming language. In both experiments 1 and 2, we attempted to elicit masked priming effects for different suffix conditions, in addition to other theoretically important conditions: identity, stem, rhyme, and semantic relatedness. The combined results of experiment 1 and 2 reveal uniformly small and statistically non-significant *suffix* priming effects, no matter whether shorter or longer prime durations were used (M_{suffix} = 2 ms). Effects of *stem* priming, on the other hand, were substantially larger and statistically significant, and varied as a function of the prime duration (M_{stem_exp1} = 16 ms vs M_{stem_exp2} = 31 ms), mirroring the dynamics of the *identity* condition ($M_{identity\_exp1}$ = 25 ms vs $M_{identity\_exp2}$ = 47 ms). Effects of semantic priming were null ($M_{semantic}$ = -2 ms), while orthographic similarity effects were non-negligible (M_{rhyme} = 13 ms). As reported in the literature, this pattern of results confirms the idea that stem priming is due to morphological relatedness, thus being comparable to identity priming; while ruling out the possibility that it may just be due to a combination of similarity in form and similarity in meaning, which morphologically related words share by definition. The results here strongly suggest that simple meaning similarity does not give rise to masked priming effects between prime and target words. Orthographic similarity on the other hand seems to give rise to non-negligible effects, but they are still substantially smaller than the stem priming effects, indicating that the latter cannot be reduced to a special case of the former. It should be noted that the orthographic rhyme similarity results observed in our study contrast with previous studies reporting null effects for other types of orthographic similarity (e.g., Rastle et al. 2000: *electrode-ELECT, typhoid-TYPHOON;* Rastle et al. 2004: *against-AGAIN*). Whether this is a function of the particular type of orthographic similarity used in the experiment or some other factor (e.g., online vs. lab setting) merits further exploration.

The contribution of the study reported here is twofold. First, it adds to a growing body of similar studies (e.g., Angele et al. 2022) confirming the feasibility of running masked priming experiments online. Despite current technical limitations preventing equally precise and accurate timing of stimulus presentation in web browsers as can be obtained in the lab, these limitations are seemingly due to the variability of devices used by participants, and may be overcome by applying design-specific rejection criteria. In masked priming experiments, the prime duration is crucial for subliminal prime presentation This can be implemented in a web-based experiment by removing all trials whose prime duration was out of the desired range, or participants in which a substantial proportion of trials had prime durations outside of the desired range. As a consequence, an online experiment may require a substantial increase in sample size (our experience so far indicates at least double the subjects) compared to what would be needed in a regular lab setup. Second, the consistency of null suffix priming effects in our experiments support the affix-stripping model of morphological decomposition in lexical access. At early stages of visual word processing, decomposition occurs on the basis of morpho-orthographic regularities, and eventually triggers lexical activation of the stem, thus leading to priming. However, suffixes do not seem to be activated at all – they are just "stripped off" of the stem, and do not appear to be used in lexical retrieval in the same way stems are, as the latter give rise to priming effects but the former do not. Such a conclusion raises two important questions. (1) Does the same occur for other affixes (prefixes, infixes, etc.) as well? The location of affixes with respect to the stem may indeed impinge on early processing and ultimately priming elicitation. Previous studies on prefixes have however provided mixed results, likely because of lack of statistical power. (2) Why would stems prime more than affixes in the first place, under the assumption that masked priming occurs prior to lexical access, and should therefore be blind to the distinction between stems and affixes? It is crucial to point out that the bulk of the literature of morphological priming is primarily based on English, a stem-based language where words may surface as phonologically identical to the underlying stems. This property of English is not very common cross-linguistically; in many languages (e.g., Romance languages) stems are instead always bound, in the sense that they never occur without at least one grammatical affix. Such an idiosyncratic property may ultimately hinder the more direct comparison between the affix and stem masked priming response, and therefore detection of potential differences between the two. A thorough investigation of stem and affix priming that takes into account such considerations and operationalizes them accordingly is therefore much needed, but must be left for future investigation.

References

Amenta, Simona, and Davide Crepaldi. 2012. "Morphological Processing as We Know It: An Analytical Review of Morphological Effects in Visual Word Identification". *Frontiers in Psychology* 3: 232. DOI: 10.3389/fpsyg.2012.00232.

Andoni Duñabeitia, Jon, Manuel Perea, and Manuel Carreiras. 2008. "Does Darkness Lead to Happiness? Masked Suffixpriming Effects". *Language and Cognitive Processes* 23 (7-8): 1002-1020.

Angele, Bernhard, Ana Baciero, Pablo Gómez, and Manuel Perea. 2023. "Does Online Masked Priming Pass the Test? The Effects of Prime Exposure Duration on Masked Identity Priming". *Behavior Research Methods* 55: 151-167.

Baayen, R. Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. Harald, Ton Dijkstra, and Robert Schreuder. 1997. "Singulars and Plurals in Dutch: Evidence for a Parallel Dual-route Model". *Journal of Memory and Language* 37 (1): 94-117.

Balota, David A., Melvin J. Yap, Keith A. Hutchison, Michael J. Cortese, Brett Kessler, Bjorn Loftis, James H. Neely, Douglas L. Nelson, Greg B. Simpson, and Rebecca Treiman. 2007. "The English Lexicon Project". *Behavior Research Methods* 39 (3): 445-459.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. "Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal". *Journal of Memory and Language* 68 (3): 255-278.

Boudelaa, Sami, and William Marslen-Wilson. 2001. "Morphological Units in the Arabic Mental Lexicon". *Cognition* 81: 65-92.

Chateau, Dan, Erin V. Knudsen, and Debra Jared. 2002. "Masked Priming of Prefixes and the Influence of Spelling–Meaning Consistency". *Brain and Language* 81(1-3): 587-600.

Crepaldi, Davide, Lara Hemsworth, Colin J. Davis, and Kathleen Rastle. 2016. "Masked Suffix Priming and Morpheme Positional Constraints". *The Quarterly Journal of Experimental Psychology* 69 (1): 113-128.

Feldman, Laurie B., and Emily G. Soltano 1999. "Morphological Priming: The Role of Prime Duration, Semantic Transparency, and Affix Position". *Brain and Language* 68 (1-2): 33-39.

Feldman, Laurie B., Emily G. Soltano, Matthew J. Pastizzo, and Sarah E. Francis. 2004. "What Do Graded Effects of Semantic Transparency Reveal about Morphological Processing?" *Brain and Language* 90 (1-3): 17-30.

Finger, Holger, Caspar Goeke, Dorena Diekamp, Kai Standvoß, and Peter König. 2017. "Labvanced: A Unified JavaScript Framework for Online Studies". In *2017 International Conference on Computational Social Science IC2S2*, 10-13 July 2016, Cologne, Germany. <https://www.labvanced.com/static/2017_IC2S2_LabVanced.pdf> (07/2023).

Forster, Kenneth I. 1999. "Microgenesis of Priming Effects in Lexical Access". *Brain and Language* 68: 5-15.

Forster, Kenneth I., and Chris Davis. 1984. "Repetition Priming and Frequency Attenuation in Lexical Access". *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10 (4): 680.

Forster, Kenneth I., Kathleen Mohan, and Jo Hector. 2003. "The Mechanics of Masked Priming". In *Masked Priming: The State of the Art*, ed. by Sachiko Kinoshita, and Stephen J. Lupker, 3-37. New York, NY-Hove: Psychology Press.

Frost, Ram, Avital Deutsch, and Kenneth I. Forster. 2000. "Decomposition Morphologically Complex Words in a Nonlinear Morphology". *Journal of Psychology* 26 (3): 751-765.

Giraudo, Hélène, and Serena Dal Maso. 2016. "Suffix Perceptual Salience in Morphological Masked Priming". *Lingue e Linguaggio*, 85-106. Bologna: Il Mulino.

Giraudo, Hélène, and Jonathan Grainger. 2003. "On the role of derivational affixes in recognizing complex words: Evidence from masked priming." In *Morphological Structure in Language Processing*, ed. by R. Harald Baayen, and Robert Schreuder, 209-232. Berlin-New York: De Gruyter Mouton.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H.B. Christensen. 2017. "lmerTest Package: Tests in Linear Mixed Effects Models". *Journal of Statistical Software* 82 (13): 1-26.

Peirce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. 2019. "PsychoPy2: Experiments in Behavior Made Easy". *Behavioral Research Methods* 51 (1): 195-203.

Petrosino, Roberto. 2020. *More than Islands of Regularity: An Investigation of the Sensitivity of Morphological Decomposition to Higher-level Linguistic Properties*. PhD dissertation, University of Connecticut.

Rastle, Kathleen, and Matthew H. Davis. 2008. "Morphological Decomposition Based on the Analysis of Orthography". *Language and Cognitive Processes* 23 (7-8): 942-971.
Rastle, Kathleen, Matt H. Davis, William D. Marslen-Wilson, and Lorraine K. Tyler. 2000. "Morphological and Semantic Effects in Visual Word Recognition: A Time-course Study". *Language and Cognitive Processes* 15 (4-5): 507-537.
Rastle, Kathleen, Matthew H. Davis, and Boris New. 2004. "The Broth in My Brother's Brothel: Morpho-orthographic Segmentation in Visual Word Recognition". *Psychonomic Bulletin & Review* 11 (6): 1090-1098.
Taft, Marcus. 1994. "Interactive-activation as a Framework for Understanding Morphological Processing". *Language and Cognitive Processes* 9 (3): 271-294.
Taft, Marcus. 2004. "Morphological Decomposition and the Reverse Base Frequency Effect". *The Quarterly Journal of Experimental Psychology Section A* 57 (4): 745-765.
Taft, Marcus, and Kenneth I. Forster. 1975. "Lexical Storage and Retrieval of Prefixed Words". *Journal of Verbal Learning and Verbal Behavior* 14 (6): 638-647.

Appendix

related prime	**unrelated prime**	**TARGET**	**related prime**	**unrelated prime**	**TARGET**
	plural condition			**identity condition**	
sarcasms	canonize	GENDERS	shrimp	enigma	SHRIMP
panoramas	flutist	CHAPTERS	culprit	diploma	CULPRIT
careers	bundled	WAYS	cocoon	fiddle	COCOON
manners	eternity	FELLOWS	tomato	breeze	TOMATO
morals	vaguely	EMPERORS	algebra	villain	ALGEBRA
carpenters	constipated	TAVERNS	ribbon	coffin	RIBBON
goods	mailed	TOWNS	glance	tattoo	GLANCE
valleys	sobbing	DILEMMAS	whistle	persona	WHISTLE
worlds	assumed	HEAVENS	sorrow	coyote	SORROW
councils	inscribed	CEILINGS	fridge	locale	FRIDGE
colonels	possessor	CAMELS	terrace	shackle	TERRACE
kitchens	vengeful	ROADS	toggle	blight	TOGGLE
acorns	rougher	LIZARDS	zodiac	caress	ZODIAC
cancers	booming	PERSONS	acrobat	inferno	ACROBAT
fountains	tormented	DIAMONDS	juggle	saliva	JUGGLE
humans	managed	CREEDS	ghetto	benign	GHETTO
domains	collision	DOGMAS	shuffle	lantern	SHUFFLE
husbands	coastal	MASONS	gorilla	conduit	GORILLA
operas	snowy	PASSIONS	dolphin	sheriff	DOLPHIN
umbrellas	adornment	PROTEINS	tornado	treason	TORNADO

semantic condition			rhyme condition		
cauliflower	allegorical	BROCCOLI	ballad	muffin	SALAD
quench	pidgin	THIRST	beetle	pupil	NEEDLE
fork	ugly	SPOON	casket	noxious	BASKET
niece	swipe	NEPHEW	cuddle	midget	SUBTLE
tutu	patron	BALLET	petal	convey	MEDAL
pint	lair	QUART	deceit	cripple	RECEIPT
bee	mat	STING	fennel	mutton	KENNEL
alto	ammo	SOPRANO	ferry	violin	CHERRY
squid	drool	OCTOPUS	hurdle	apron	TURTLE
sprain	thrall	ANKLE	lactic	garnish	TACTIC
wick	tart	CANDLE	lotion	fissure	NOTION
lizard	oxygen	REPTILE	marrow	gradient	NARROW
circus	walnut	CLOWN	zealous	devour	JEALOUS
chipmunk	dandruff	SQUIRREL	mumble	sundry	HUMBLE
grain	panic	WHEAT	rattle	violet	CATTLE
bacteria	frontier	FUNGUS	ravage	troupe	SAVAGE
poem	dose	RHYME	saloon	purport	BALLOON
convince	terrible	PERSUADE	taper	mimic	VAPOR
entertain	intellect	AMUSE	tumor	satin	RUMOR
volcano	scholar	ERUPT	wallow	torrent	HOLLOW
***er*-condition**			**stem condition**		
scorer	bulging	STALKER	boneless	doctored	BONE
forger	butchered	DANCER	cloudy	sweaty	CLOUD
shipper	outing	BINDER	calmly	dearly	CALM
jogger	excused	FREEZER	dreadful	appealed	DREAD
booker	hostess	HUNTER	egoism	chimed	EGO
grabber	brainless	PREACHER	fondly	eroded	FOND
whiner	sealing	TRAINER	flawless	communal	FLAW
thriller	chopping	CALLER	foggy	renal	FOG
reaper	weaning	MARKER	madden	choked	MAD
cooker	grassy	DRUMMER	waived	choppy	WAIVE
rancher	toxic	BUILDER	acidic	touchy	ACID
golfer	steely	BREEDER	rusted	dreamy	RUST
gunner	likeness	RUNNER	melted	evenly	MELT

stroller	panicked	SENDER	awaken	hinted	AWAKE
drinker	fondness	STICKER	weaned	choral	WEAN
learner	stressing	SCANNER	sadness	agility	SAD
hatter	tickled	FOUNDER	sinful	bridal	SIN
solver	warping	JUMPER	smelly	molded	SMELL
wrangler	sewage	LOVER	vanity	autism	VANE
sniper	bedding	WINNER	wisely	loosen	WISE
	***able*-condition**			***ness*-condition**	
advisable	silencer	PLAYABLE	weakness	container	SICKNESS
honorable	referral	NOTICEABLE	correctness	disposal	FORGIVENESS
washable	theorize	NOTABLE	baldness	childless	KINDNESS
quotable	humorist	ADJUSTABLE	smoothness	distinction	AWARENESS
commendable	vacationing	PAYABLE	blindness	strictest	SOFTNESS
variable	greatest	READABLE	harshness	sparsely	BUSINESS
detectable	timeless	USABLE	fitness	quicker	MADNESS
curable	frontage	PROFITABLE	calmness	vicarious	READINESS
debatable	courageous	DESIRABLE	rudeness	mobilize	GOODNESS
dependable	currencies	PORTABLE	soreness	molesting	UNIQUENESS
traceable	parenthood	VALUABLE	toughness	visionary	ILLNESS
suitable	resistance	ACCEPTABLE	gentleness	voiceless	SWEETNESS
attainable	affliction	RELIABLE	fairness	yearly	SLOWNESS
observable	possessing	REASONABLE	aloofness	flutist	DARKNESS
favorable	adequately	BELIEVABLE	crispness	victimize	HAPPINESS
admirable	descendant	TAXABLE	stubbornness	accentuate	STIFFNESS
pardonable	spiritless	WORKABLE	darkness	coverage	GREATNESS
manageable	imposition	RESPECTABLE	politeness	snapping	HARDNESS
cashable	improviser	EXPANDABLE	awkwardness	subjection	HOLINESS
adaptable	ascension	ALLOWABLE	boldness	fiendish	EMPTINESS