Title
A meta-analytic review of morphological priming in Semitic languages

Journal
The Mental Lexicon, 18(2)

ISSN
1871-1340

Authors
Xu, Lily
Solá-Llonch, Elizabeth
Wang, Huilei
et al.

Publication Date
2023-12-31

DOI
10.1075/ml.00024.xu

A meta-analytic review of morphological priming in Semitic languages

Lily Xu, MA

Elizabeth Solá-Llonch, MA

Huilei Wang, MA

Megha Sundara, Ph.D

UCLA Department of Linguistics

Corresponding author: Lily Xu

Telephone: (310) 825-0634

Fax: (310) 206-5743

Email: lilyokc@ucla.edu

Disclaimer: The views expressed in the submitted article are the authors' own and not an official position of UCLA or the NSF.

Sources of support: This research was partially funded by NSF BCS-2214017 to MS (co-PI Connor Mayer, UCI).

Word count: 10927

Abstract

Two types of discontinuous morphemes are thought to be the basic building blocks of words in Semitic languages: roots and templates. However, the role of these morphemes in lexical access and representation is debated. Priming experiments, where reaction times to target words are predicted to be faster when preceded by morphologically-related primes compared to unrelated control primes, provide conflicting evidence bearing on this debate. We used meta-analysis to synthesise the findings from 229 priming experiments on 4710 unique Semitic speakers. With Bayesian modelling of the aggregate effect sizes, we found credible root and template priming in both Arabic and Hebrew, in nouns as well as verbs. Our results show that root priming effects can be distinguished from the effects of overlap in form and meaning. However, more experiments are needed to determine if template priming effects can be distinguished from overlap in form and morphosyntactic function.

Abstract word count: 146

Keywords: roots, templates, word patterns, Arabic, Hebrew, Maltese, Bayesian modelling

1 Introduction

A central question in linguistics concerns morphology's role in lexical representation and its contribution to lexical access. In some accounts, sublexical units — morphemes — are explicitly represented in the mental lexicon and play an active role in moderating lexical access (e.g., Smolka et al., 2014, 2019; Stockall & Marantz, 2006; Taft, 1988, 2004; Taft & Forster, 1975). In contrast, in whole-word approaches, words are represented as unanalyzed sequences in the mental lexicon, without any internal structure (e.g., Blevins, 2016; Lukatela et al., 1980; Milin et al., 2017; Seidenberg & Gonnerman, 2000). In such models, two lexical representations with shared morphology are typically related by analogy (e.g., Bybee & McClelland, 2005) or overlap in meaning and form (Baayen et al., 2019; Gonnerman et al., 2007; Heitmeier et al., 2022). Thus, in whole-word models, although morphemes are not represented in the lexicon, effects of morphological relatedness emerge due to the fact that words that are morphologically related typically share meaning, and are similar in form.

A third alternative, referred to as dual-route models, have some words stored whole while others are generated from stored morphemes (e.g., Berent & Pinker, 2007; Clahsen et al., 1992). Specifically, only regular and productive processes are thought to be involved in generating words from stored morphemes, whereas words resulting from irregular, unproductive processes are stored whole. While the distinction between productive and unproductive morphological processes is not categorical (see Albright & Hayes, 2003; Nieder et al., 2021 for discussion), a morphemic representation is most often argued for when words involve productive, transparent, and regular processes.

In this paper we evaluate the empirical evidence that bears on the representation of morphemes in Semitic languages, which are unique for their extensive use of nonconcatenative morphology. To address this question, we use a meta-analytic approach to aggregate evidence from 36 sets of studies investigating regular and productive morphological processes in Arabic, Hebrew, and Maltese. In all studies included here, lexical representations were probed using the psycholinguistic technique of priming.

In priming paradigms, prior experience with a stimulus item during the course of the experiment typically results in heightened activation and preferential retrieval of a related target lexical entry. Such heightened activation is reflected in faster times for lexical decision or latency of naming for target words when compared to a baseline where targets are preceded by unrelated primes. Thus, faster reaction times or naming latencies are interpreted as evidence for a shared representation between the prime and the target. The strongest priming effects are found when prime and target are identical (i.e., identity priming), but priming effects are also observed with only partial (e.g., meaning, morphological form) overlap. Therefore, manipulating the extent and type of relationship between the prime and the target in a priming paradigm can provide a window into the structure of the mental lexicon.

In this paper, we present a meta-analysis of morphological priming effects for nonconcatenative morphemes (i.e., roots and templates) in Semitic languages. We used Bayesian modelling to examine the overall effect sizes for data aggregated from all relevant studies, focusing on what these results might contribute towards our understanding of the role of morphology in lexical access and representation. With this approach, we were able to quantify the evidence to evaluate whether morphological priming by root and template morphemes is consistently observed and, more specifically, whether such priming effects are independent of the effects obtained where there is overlap in just form and meaning.

1.1 Priming of abstract morphemes in Semitic languages

Semitic languages present interesting test cases for evaluating morphological priming effects because of their extensive system of nonconcatenative, or nonlinear, morphology. Traditional analyses decompose the majority of Semitic words into two discontinuous morphemes, called roots and templates (Holes, 2004; McCarthy, 1979; *inter alia*). Roots are sequences of two to four (usually three) consonants which carry the main lexical semantic information. Templates, also called word patterns, consist of prosodic information and often include affixes and a fixed sequence of vowels; they are associated with morphosyntactic information such as part of speech, number, and voice, as well as some semantic information. Importantly, both roots and templates are abstract morphemes by default because they cannot be extracted linearly from speech. Table 1 displays the Arabic root *χbz*, related to baking and bread, combined with one verbal and two nominal templates:

Table 1: Arabic root χbz, in combination with various templates

Word	**Meaning**	**Template**	**Morphosyntactic information**
χabaz	‘to bake’	CaCaC	verb-3.M.SG.PAST
χabːaːz	‘baker’	CaCːaːC	M.SG noun relating to a profession
maχbaz	‘bakery’	maCCaC	M.SG noun relating to location

Although the morphological status of Semitic roots and templates is far from uncontroversial (Bat-El, 1994, 2003; Kastner, 2019; Ussishkin, 1999, 2005), there is substantial psycholinguistic evidence from priming experiments that consonantal roots play an important role in lexical access (see Prunet, 2006). Robust priming based on root overlap has been reported in both Arabic and Hebrew for nouns and verbs (e.g., Boudelaa & Marslen-Wilson, 2000; Deutsch et al., 1998; Frost et al., 1997). It has been observed in a variety of methodological paradigms, including masked visual priming (Boudelaa & Marslen-Wilson, 2000, 2005; Deutsch et al., 1998; Frost et al., 1997; cf. Kastner et al., 2018), parafoveal preview (Deutsch et al., 2000 et seq.), and cross-modal priming (Boudelaa & Marslen-Wilson, 2015; Frost et al., 2000). In fact, robust root priming effects are observed even in Arabic learning children as early as second grade (Shalhoub-Awwad & Leikin, 2016).

Strong root priming effects are also found in Maltese. Though a Semitic language, Maltese has had extensive contact with Indo-European languages (Sicilian, Italian, and English) and has developed a lexicon split between native Semitic words and loan words. Root priming has been found in Maltese for verbs (Geary & Ussishkin, 2018; Twist, 2006; Ussishkin et al., 2015) and for nouns (Nieder et al., 2021) using the masked visual, cross-modal, and auditory priming paradigms. For instance, Geary & Ussishkin (2018) compare masked visual root priming for words of both Semitic and non-Semitic origin. They found priming effects for Semitic words when they were preceded by their consonantal root (e.g., *kines* ‘to sweep’ primed by *kns*) but no priming effects for non-Semitic words when they were preceded by similar triconsonantal strings (e.g., *pinġa* ‘to draw’ primed by *pnġ*). These results highlight the special status of roots in Maltese.

While widespread, root priming effects can be moderated by the productivity of the templates they combine with. Productivity typically refers to the type frequency of the templates or the extent to which any template can be used to form novel words. For example, in Hebrew, Farhy et al. (2018) compare root priming effects for two different verbal templates (*Paal* and *Piel*) that have comparable type frequency and semantic and syntactic properties, but which crucially differ in that *Piel* is used far more often in the adaptation of loanwords than *Paal*. Priming was observed only when roots were combined with the *Piel* verbal template, but not the less productive *Paal*.

In contrast, the psycholinguistic evidence for templates in Semitic languages is much more mixed. Compared to root priming, template priming effects are generally reported to be less robust and exhibit greater variation, both within and across languages. In Arabic, for example, facilitatory effects have been observed in visual word identification for both verbal and nominal templates (Boudelaa & Marslen-Wilson, 2005, 2011), but only with specific time intervals between the prime and the target (Stimulus-Onset Asynchrony, SOA). Additionally, nominal template priming has been reported to vary with the type frequency of the root (Boudelaa & Marslen-Wilson, 2011) and to be less robust in children (Shalhoub-Awwad, 2020). Further, not all nominal templates show priming effects: productive nominal templates show template priming effects, but non-productive ones do not (Boudelaa & Marslen-Wilson, 2000, 2015).

In Hebrew, there is a further asymmetry between the template priming effects found for nouns versus verbs. Significant priming effects have been reported for Hebrew verbal templates in lexical decision as well as naming tasks, but generally not for nominal templates (verbal templates: Deutsch et al., 1998; see also Kastner et al., 2018 for MEG data; nominal templates: Deutsch et al., 2005; Frost et al., 1997). The lack of nominal template priming effects in Hebrew has been attributed to a few different factors. First, Deutsch et al. (1998) point out that verbal and nominal templates differ greatly in type frequency: there are only 7 verbal templates in Hebrew and over 100 nominal templates, so each individual verbal template appears more frequently than any individual nominal template. Additionally, they argue that the meaning derived from nominal templates is less transparent and more unpredictable. However, due to the fact that the same holds in other Semitic languages as well (particularly Arabic), it is difficult to reconcile Deutsch et al.'s account of the asymmetry of template priming in Hebrew with its absence in Arabic.

Deutsch et al. (2018) supply a third possible account for the asymmetry of template priming in Hebrew verbs versus nouns. They argue that in past studies, potential facilitative effects of nominal templates might have been obscured by the competition between templates and roots, since roots have been shown to exert a stronger influence on lexical access. The competition with roots is not expected to affect verbal template priming, since verbal template priming effects are overall stronger. Consistent with this account, Deutsch et al. (2018) find nominal template priming effects in a fast-priming paradigm using a "letter-delay" procedure to control for any interference effects from roots. In sum, the findings from Hebrew template priming studies suggest that, minimally, there are differences in the robustness of root priming and templatic priming of verbs versus nouns, with only nominal template priming being sensitive to methodological differences.

Template priming effects are weakest in Maltese. There have been two studies on template priming in Maltese, both of which report no priming effects (Twist, 2006: masked visual priming; Ussishkin et al., 2015: auditory priming). This is despite the fact that both studies evaluated verbal templates, which show consistent priming effects in both Arabic and Hebrew. Ussishkin et al. (2015) remark that these results could be due to templates having much less informational content than roots, since templates generally define much larger classes of words.

This is problematic for any account of a Semitic language because this observation can be made for Arabic and Hebrew as well, and yet verbal template priming is still found in both. The findings from Maltese are further complicated by the fact that many verbs in the language are non-Semitic loans that take suffixes rather than verbal templates (Twist, 2006), so verbal templates in Maltese have comparatively less type frequency than in other Semitic languages. Consistent with these lexical trends, a wug test conducted by Twist (2006) found that Maltese speakers were overall more likely to conjugate nonce verbs using suffixation, though speakers did still favour using verbal templates for nonce words they perceived as Semitic in origin. However, since there are so few studies on Maltese, it is hasty to conclude that there is a definite difference in template priming in Maltese compared to Hebrew and Arabic.

To summarise, while root priming has been robustly observed in both nouns and verbs in Arabic, Hebrew, and Maltese, there is no clear evidence for template priming in Semitic languages. Verbal template priming is typically demonstrated in both Arabic and Hebrew, but not Maltese. In contrast, nominal template priming is generally observed in Arabic, but not Hebrew, possibly due to methodological differences or differences in lexical frequency or productivity. However, there are also many papers that report results contradictory to the overall trends summarised above. The question is thus: how can we isolate the effect of interest (i.e., root and template priming) from the effects of the many different methodologies and cross-linguistic variation in lexical statistics?

In this paper we used meta-analysis to synthesise the large literature on root and template priming effects in Semitic languages in order to tease apart the influence of methodological variables (task and modality of presentation) from lexical factors (e.g., word class) to get at cross-linguistic similarities and differences. To evaluate effect sizes for the meta-analysis, we used Bayesian modelling. Bayesian models are useful for integrating evidence from individual studies, both significant and non-significant, to estimate the strength of the effect size. Additionally, the large sample sizes in a meta-analysis are particularly useful because priming experiments generate reaction time data that are known to be noisy. Finally, because of the variability in the designs of individual experiments, we could make novel comparisons (e.g., across languages) which are not possible in any individual report.

1.2 Is Semitic morphology represented in the mental lexicon?

Priming studies are often used to provide evidence to distinguish between alternate theories of morphological representation. For instance, robust root priming effects are typically used to further an account where morphological structure is independently represented (e.g., Boudelaa & Marslen-Wilson, 2005, 2011; Deutsch et al., 1998; Frost et al., 1997; Geary & Ussishkin, 2018). However, root priming can also be attributed to the overlap in the form of the root and the shared meaning between the prime and the target. Therefore, to distinguish between the two proposals we need to evaluate whether root priming effects are independent of both form and semantic overlap.

In the priming literature, experiments targeting specific relationships between primes and targets are used to isolate the role of morphological overlap. Isolating the role of morphological overlap and semantic overlap typically involves a comparison of facilitation by primes that share the same root as the target, but which either have a transparent semantic relationship (e.g., Hebrew *madrix* 'guide' and *hadraxa* 'guidance') or an opaque semantic relationship (e.g., Hebrew *drixut* 'alertness' and *hadraxa* 'guidance'). In some studies in Arabic and Hebrew, root-related words with an opaque semantic relationship have been reported to prime as much as root-related prime-target pairs with a transparent semantic relationship (Boudelaa & Marslen-Wilson, 2000, 2005; Frost et al., 1997). In other studies in Arabic and Hebrew, root-related prime-target pairs with an opaque semantic relationship have been reported to prime less than pairs with a transparent semantic relationship (Bentin & Feldman, 1990; Boudelaa & Marslen-Wilson, 2013; Frost et al., 2000). Thus the evidence in support of morphological priming in the absence of semantic overlap is mixed.

Similarly, a comparison of priming by root-related words versus those with overlap in 2-3 non-root consonants also presents an ambiguous picture. In some studies root priming effects are larger (Boudelaa & Marslen-Wilson 2005, 2013, 2015; Deutsch et al. 2000 et seq.; Frost et al. 1997, 2000), but in others they are not (Abu-Rabia & Awwad, 2004). So, it is also unclear whether morphological root overlap effects are entirely independent from effects of form overlap.

Since template priming is less robust than root priming, the evidence from template priming experiments in support of one or another account of morphological representation is even less persuasive. Some of these mixed findings from priming studies undoubtedly stem from differences in methodology, lexical factors, or cross-linguistic variation. Therefore, aggregating evidence across a large number of experiments in this area, as we did here with the meta-analysis, can be particularly helpful in determining whether (if at all) morphological priming effects exist independent of semantic and form overlap.

2 Methods

2.1 Paper identification and selection

We initially aggregated a total of 82 papers on psycholinguistic studies of nonconcatenative morphology, including journal publications, conference proceedings, book chapters, and unpublished dissertations. This selection included papers known to the authors (10 papers) and papers identified by systematic searches of databases (60), reference lists (8), and researcher's websites (4). After deciding to focus on only priming and parafoveal preview studies, we excluded papers that used other methodologies (19). One additional paper did not have a retrievable abstract or text, leaving a total of 62 papers.

After screening paper abstracts, we excluded papers that did not evaluate nonconcatenative morphology (7) or those that did not test native speakers of Semitic languages (1). Papers with non-standard priming methodologies were also excluded (3); this included two studies using masked auditory priming, a newer priming methodology with results that are more difficult to interpret. We also decided to discard studies that focused on priming in children (4) or dyslexic participants (1). We then examined the full text of the remaining 46 papers to determine their eligibility for the meta-analysis, resulting in an additional 10 studies being excluded because they focused only on irregular (non-triliteral) roots (2), only used nonce word primes (1), did not report data critical for the meta-analysis (5), or because their control conditions would actually be categorised as template priming by the definition used in this meta-analysis (2). This left a total of 36 papers eligible for the meta-analysis after screening. The PRISMA flowchart and decision spreadsheet for this meta-analysis can be found on this project's OSF page.

2.2 Data entry

Every experiment and control condition pair constituted a row in the meta-analysis spreadsheet, and was coded for a number of dimensions following previous meta-analyses (Sundara et al., 2021; for a full list and explanation of coded variables see Bergmann et al., 2018). For example, Boudelaa & Marslen-Wilson (2011) report two experiments on morphological priming in Arabic. Both experiments had 5 experimental conditions, each compared to a control condition; each experiment provided data for 5 rows of our spreadsheet and we treated this paper as containing ten experimental comparisons in total. Each row coded the specific properties of a single experimental condition.

The relevant dimensions for the present analysis were, (1) background information on the paper, including a unique study ID for each paper, citation, and DOI and whether the paper was peer reviewed; (2) participants' information, including country of origin, native language, dialects, age, and gender; (3) modality of presentation of targets and primes (auditory presentation of both primes and targets: auditory; auditory presentation of primes and visual presentation of targets: cross-modal; visual presentation of primes and targets: visual); (4) type of visual primes (masked or nonmasked); and (5) the mean and standard deviations of the reaction time to experimental and control conditions. When standard deviations were not reported, we estimated them from ANOVA results (see section 2.2.2 below for details).

Additionally, we also coded each experiment and control condition pair for (a) part of speech of stimuli (nouns or verbs); (b) root type of the stimuli (regular (triliteral) vs. irregular); and (c) description of the stimuli, such as the number of total targets, the number of targets per condition, the number of trials, as well as the productivity of the roots and templates when available. Finally, each experimental and control pair was classified for the relationship between primes and targets, using six different areas for potential overlap: root; template; meaning; form; template without morphosyntactic similarity; root without meaning overlap. This was done based on the description of the stimuli or a comparison of the stimulus set, when available.

The final dataset had results from 4710 participants in 103 experiments. These numbers were calculated from the complete experiments as defined by the individual studies. However, we coded each experimental condition as its own experiment for the purposes of our meta-analysis; by this definition, our meta-analysis had observations from 9886 participants in 229 experiments. Hereafter, we will report the number of participants and experiments using the latter definition.

The experiments evaluated root and template priming in three languages (Arabic, Hebrew, and Maltese) using 4 modalities: auditory, cross-modal (auditory-visual), nonmasked visual, and masked visual). The distributions of participants and language of study against the modality are presented in Figure 1.

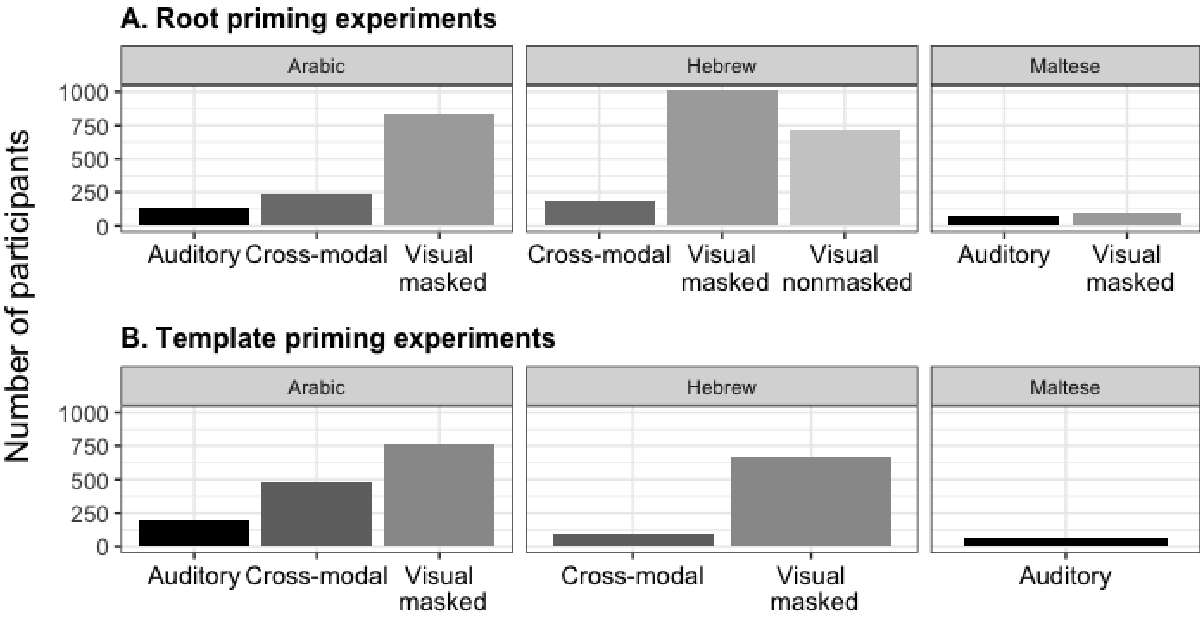

Figure 1. *Distribution of the number of participants across different priming modalities (auditory, cross-modal, masked visual, and nonmasked visual) in Arabic, Hebrew, and Maltese .*

Data from root priming experiments are shown in panel A while data from template priming experiments are in panel B.

2.2 Derived variables

2.2.1 Effect size

A standardised effect size measure (Hedge's *g*) was calculated to index the extent of facilitation in reaction time (i.e., priming effect) in every experimental condition compared to its respective control condition. We used Hedge's *g*, which is calculated by dividing the difference in sample means by the pooled and weighted standard deviation of the two means. Hedge's *g* is interpreted similarly to Cohen's *d*, another common effect size measure. The difference between the two is that Hedge's *g* is weighted by sample sizes; as a result, the meta-analytic estimates are affected more by studies with larger sample sizes (Shadish & Haddock, 2009). Effect size calculations were carried out with the *esc_mean_sd()* function from the *esc* package (Lüdecke, 2019). Hedge's *g* was the dependent variable throughout the paper.

2.2.2 Standard deviation

In about half of the included experiments, we had to derive standard deviations because standard deviation values (or standard errors or 95% confidence intervals) were not reported in either the text or in figures. We did not estimate standard deviations from planned comparisons or *t*-tests because these measures were generally reported only for significant effects, and not all effects within an experiment, and meta-analysis uses both significant and non-significant reported effects. Instead, missing standard deviation values were estimated from ANOVA results (either from group means, sample sizes, and *F*-statistic or MSE, when available). We were able to do so because the *F*-statistic is the variance between samples divided by the variance within samples, which is the pooled standard deviation across conditions. The pooled standard deviation is a reasonable estimate of the standard deviation of each condition, when the samples are homogenous, i.e., have the same variance. Thus, bootstrapping the standard deviation in this way necessarily required an assumption of homogeneity of variance between experimental and control condition pairs.

2.3 Analyses

Analyses were done using the *brms* package (Bürkner, 2017) in R (R Core Team, 2021). To aggregate evidence for priming effects across studies, we used hierarchical Bayesian analyses to model the Hedge's *g* measures derived from all the experimental data. A random intercept of experimental comparison nested within paper was included in all models to capture the residual variance caused by non-uniform factors, such as language, testing method, research team, and population. We also took the uncertainty in the effect size in the original paper into consideration by modelling the effect size derived from a Normal distribution parameterized by the mean and standard error of Hedge's *g*, as described in section 2.2.2. Throughout the paper, we used a Normal (0,1) prior on standard deviations and no prior on coefficients because we did not have prior belief about whether we would find an effect, but also did not expect large effect sizes. A sensitivity analysis (available in the supplemental materials on the OSF page) showed little variation in posterior distributions for a range of prior values.

We report the median value of the posterior distribution for each parameter of interest, along with values denoting the upper and lower limits of the 95% Credible Interval, from which we can make inferences about the likelihood of the values of the parameter. Because priming results in faster reaction times in experiment conditions when compared to the control conditions, priming or facilitation is represented by negative effects sizes in this paper. An aggregate effect size is credible if the interval does not include zero. In cases where the 95% credible interval includes zero, we also report the posterior probability of the effect size; this statistic corresponds to the proportion of credible values less than zero and represents the probability of having any nonzero priming effect.

3 Results and Discussion

3.1 Part I: Evaluating bias in the morphological priming literature

3.1.1 Evidence for selective submission or publication bias

In order to detect publication bias resulting from selective recruitment practices or small-study effects (Sterne & Egger, 2001; Sterne & Harbord, 2004), we first used a funnel plot, in which Hedge's *g* is plotted against 1/Standard Error (Figure 2). Any asymmetries in the funnel plot signal that results which did not confirm *a priori* hypotheses failed to get published. These asymmetries are most likely to be found in the bottom of the funnel plot where power is low, but estimates of effect size are large (Vasishth et al., 2018).

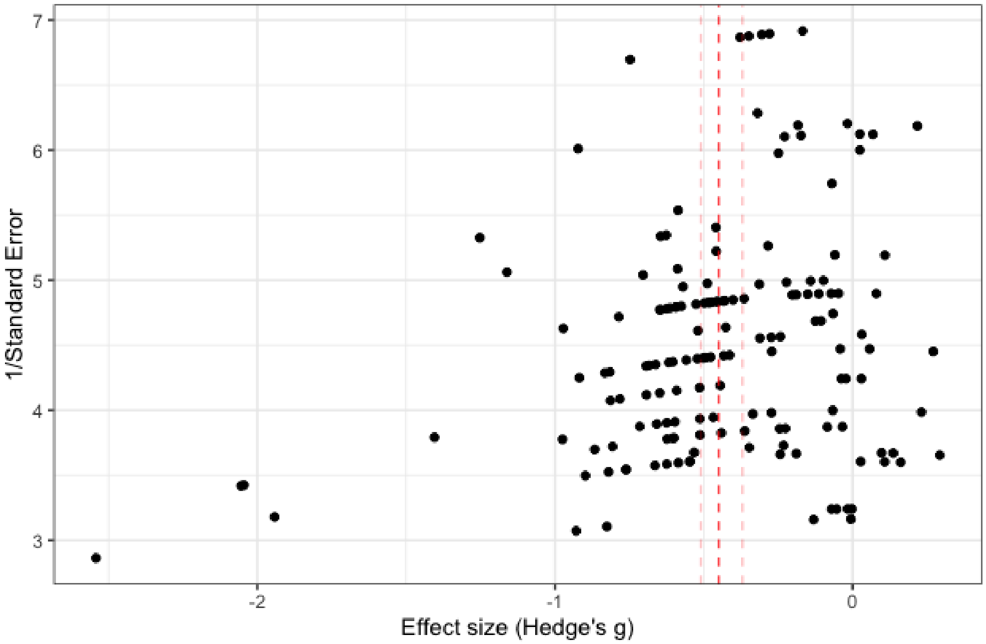

Figure 2. *Funnel plot of the effect size (Hedge's g) against 1/Standard Error. Each dot represents a single experiment in the meta-analysis. The red dotted line indicates the median effect size, while the paler lines indicate the 95% Credible Interval.*

We can see in Figure 2 that outliers contribute substantially to the asymmetry in our funnel plot; low-powered studies (low on the vertical axis) skewed towards the left on the horizontal axis, with large negative effect sizes. This shows a bias where other studies with low power might have been conducted but not published if they resulted in non-significant effects. The left skew of the distribution was confirmed by a Bayesian implementation of Egger's test (Egger et al., 1997), which is a linear regression of effect size estimates normalised using standard error weighted by its inverse variance. The estimate for this model's intercept was negative and excluded zero (β =-2.27, 95% CrI [-3.61, -0.94], $p(\beta<0) = 1$). Furthermore, the model intercept has a 97% probability of being negative even when the three outliers (clusters at and below -2) were excluded from the analysis. In other words, experiments that confirmed the *a* *priori* hypothesis that root and template overlap facilitates target recognition have been favoured in the published literature.

3.1.2 No evidence of p-hacking

Next, we carried out a *p*-curve analysis using the *pcurve()* function from the *dmetar* package (Harrer et al., 2019). This was done to evaluate whether there is evidence of "*p*-hacking", where researchers try out different statistical analyses or modify eligibility specifications and then selectively report those that produce significant results.

In a *p*-curve analysis, the distribution of *p*-values below 0.05 are examined to determine whether they are (a) more likely to have arisen from a series of studies testing a robust underlying effect (a right-skewed *p*-curve), (b) indistinguishable from those which would arise under a null underlying effect (flat *p*-curve), or (c) the likely result of extensive *p*-hacking, and therefore of questionable evidentiary value (left-skewed *p*-curve). We found evidence of right skewness ($p < 0.001$) with a power of 0.71 (CrI 0.61-0.79) indicating that there were enough studies included in the meta-analysis to provide a reasonably-powered estimate of the right skewness of the *p*-value distribution. This confirms the absence of *p*-hacking in the root and template priming experiments aggregated in the meta-analysis. Instead, a robust underlying effect has been documented in the literature.

3.2 Part II: Overall effect size and methodological factors influencing the size of the priming effect

3.2.1 Is there root and template priming?

To assess the extent of root and template priming (if present), we calculated the aggregate meta-analytic effect size for all priming experiments involving either root or template overlap, separately. Note that priming effects have negative effect sizes because experimental conditions with related primes are expected to facilitate lexical access of related items and therefore result in faster reaction times relative to control conditions.

Aggregating across 87 experiments with root overlap on 3930 participants, we found a medium effect size for root priming. The pooled effect size was -0.45, with 95% Confidence Interval (CrI) = [-0.52, -0.39] and $p(\beta<0) = 1$. That is, we can be 100% confident that the effect size is negative, and 95% confident that the size of the effect is between -0.52 and -0.39. The forest plot in Figure 3 shows the effect sizes for individual experiments evaluating root priming, with the pooled effect size at the bottom.

Based on 63 experiments with template overlap on 1604 participants (this set has partial overlap with the root-related experiments), we also found evidence of a medium effect size for template priming, The overall effect size was -0.37 (95% CrI [-0.50, -0.23], $p(\beta<0) = 1$). Thus, the template priming effect was somewhat smaller than the root priming effect and had a lot more variation, but was still robust, as can be seen from the forest plot in Figure 4.

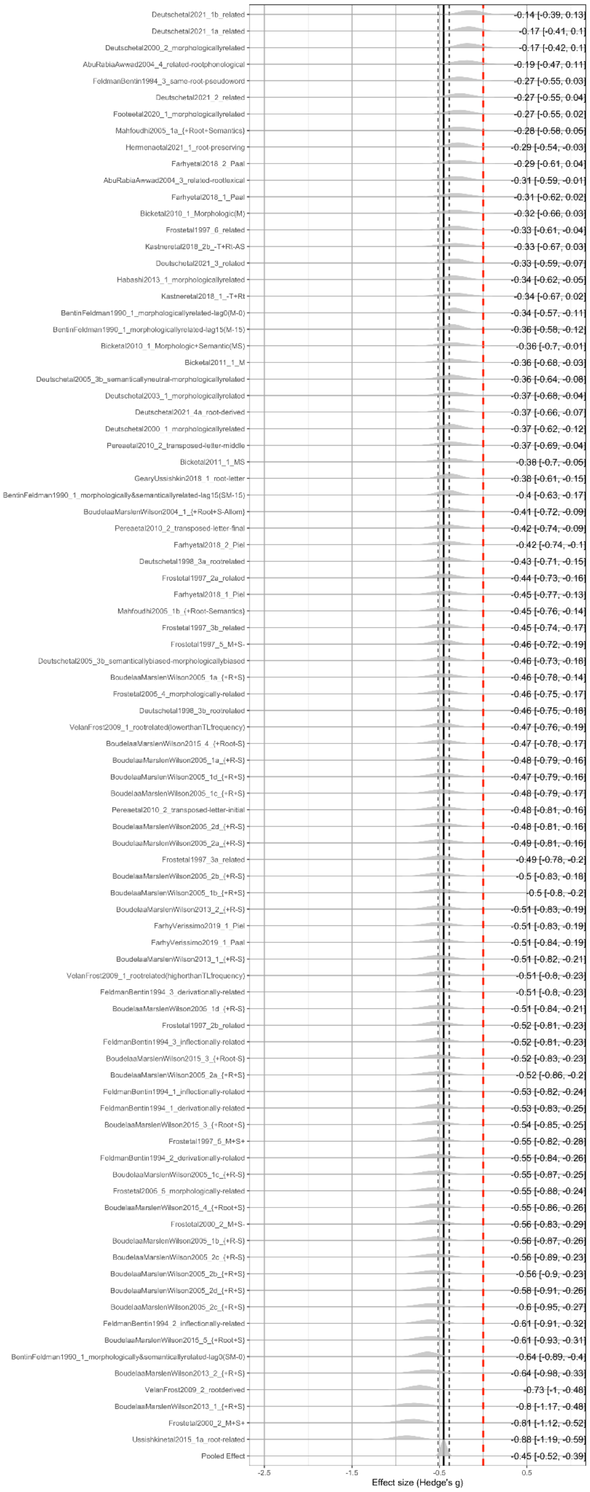

Figure 3. *Forest plot of median and 95% Credible Interval for effect sizes for each experiment evaluating root priming. The pooled effect size and credible interval are plotted on the bottom row and indicated by the black vertical lines. The red dashed line marks an effect size of zero.*

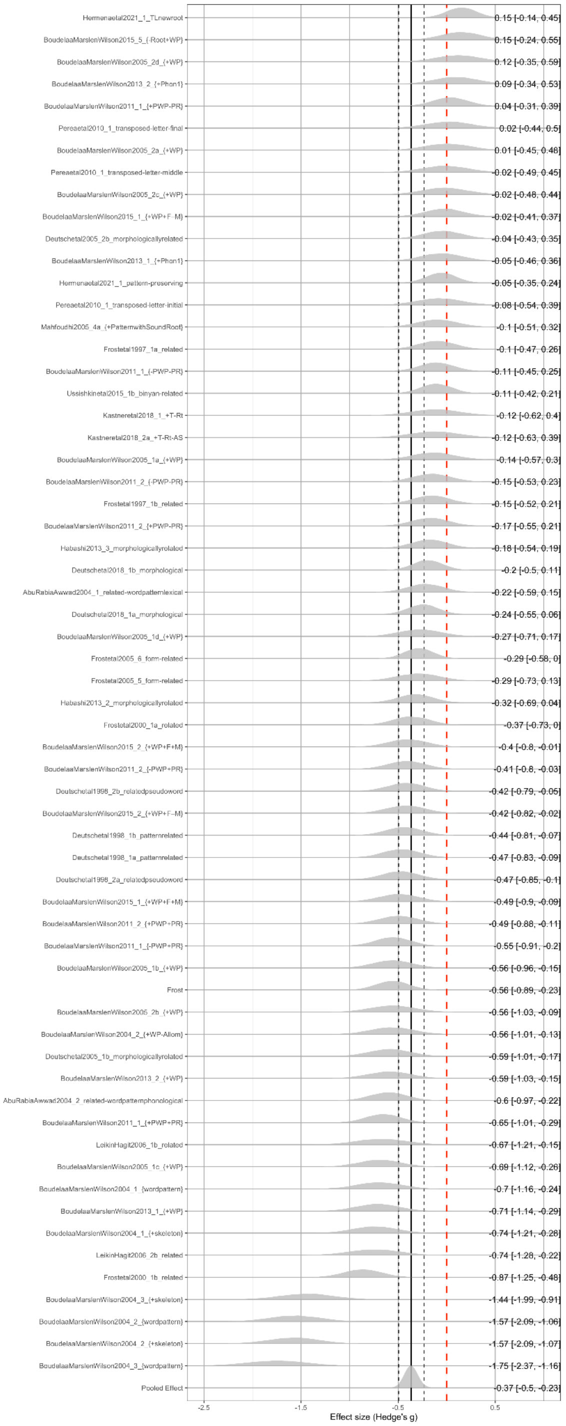

Figure 4. *Median and 95% Credible Interval for effect sizes for each experiment evaluating template priming. The pooled effect size and credible interval are plotted on the bottom row and indicated by the black vertical lines. The red dashed line marks an effect size of zero.*

3.2.2 Methodological factors

3.2.2.1 Root priming effects are smallest with visual primes, do not vary with masking of the visual prime, and are minimally affected by SOA. To determine if root priming effects were modulated by the methodology the priming data were obtained by, we compared effect sizes for the three most widely used paradigms without masked primes: auditory priming, where both primes and targets are presented auditorily (5 experiments, 204 subjects); cross-modal priming, where primes are presented auditorily and targets are visual (10 experiments, 412 subjects); and visual priming, where primes and targets are both presented visually (11 experiments, 720 subjects). We also compared masked visual priming, where both primes and targets are presented visually, but the primes are masked (47 experiments, 1852 subjects). The effect sizes are presented in Figure 5A.

Overall, there was a credible root priming in all paradigms. The largest effect size was found in auditory priming experiments ($\beta = -0.98$, 95% CrI [-1.23, -0.73]), followed by cross-modal priming experiments ($\beta = -0.68$, 95% CrI [-0.85, -0.51]), with the smallest effect size in visual priming experiments, regardless of whether primes were nonmasked ($\beta = -0.49$, 95% CrI [-0.64, -0.35]) or masked ($\beta = -0.44$, 95% CrI [-0.51, -0.37]).

Moreover, the paradigms were credibly different. Root priming was credibly larger in auditory priming compared to cross-modal priming experiments ($\beta = -0.30$, 95% CrI [-0.59, -0.02]), and credibly smaller in visual priming experiments compared to cross-modal priming experiments ($\beta = -0.23$, 95% CrI [-0.06, -0.40]). However, there was no credible difference in the extent of visual priming whether primes were masked or nonmasked ($\beta = -0.05$, 95% CrI [-0.20, 0.09], $p(\beta<0) = 0.77$).

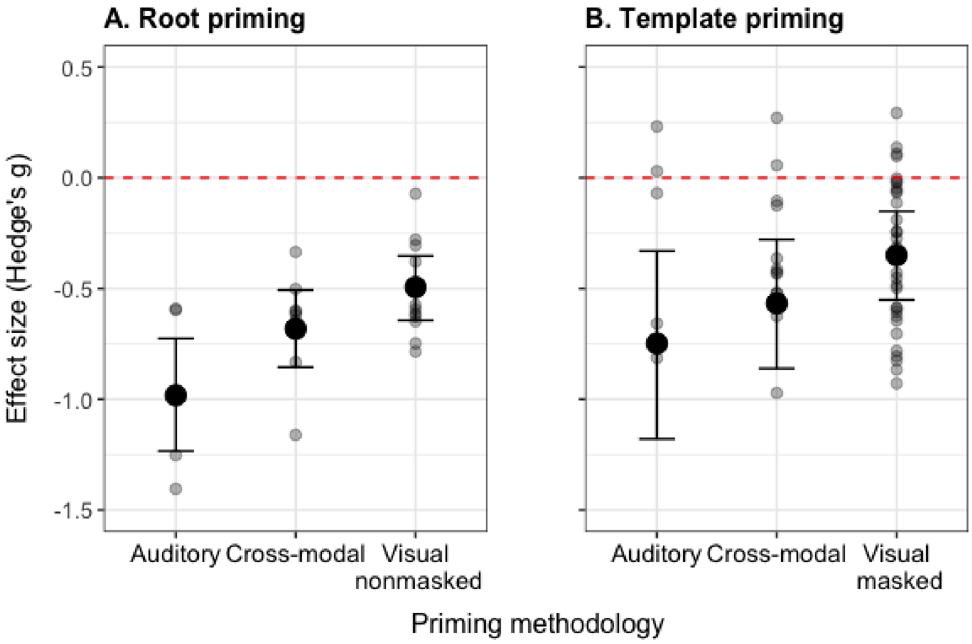

Figure 5. *Effect sizes for root priming (Panel A) and template priming (Panel B) plotted as a function of priming methodology. Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.*

Within masked visual priming paradigms, the timing between the onset of the prime and the onset of the target, referred to as the Stimulus Onset Asynchrony (SOA), has been reported to affect the magnitude and type of priming. To evaluate SOA effects, we first generated a histogram of the number of subjects tested as a function of the SOA (Figure 6). Next, we exploited the natural breaks in the wide range of SOAs included in this meta-analysis to divide the experiments investigating masked priming into five bins: 30-35ms (10 experiments, 254 subjects), 40-45ms (15 experiments, 790 subjects), 45-50ms (13 experiments, 518 subjects), 60-65ms (4 experiments, 126 subjects), and 75-80ms (4 experiments, 114 subjects). Root priming with 100% credibility was observed for all SOA bins (30-35ms: β = -0.31, 95% CrI [-0.50, -0.13]; 40-45ms: β = -0.51, 95% CrI [-0.62, -0.40]; 45-50ms: β = -0.32, 95% CrI [-0.45, -0.19]; 60-65ms: β = -0.70, 95% CrI [-0.97, -0.43]; 75-80ms: β = -0.62, 95% CrI [-0.90, -0.34]). As shown in Figure 7A, the 30-35ms and 45-50ms bins had the smallest magnitude of root priming and were not credibly different from each other. There were larger effects of root priming at an SOA of 40-45ms compared to 30-35ms (β = -0.19, 95% CrI [-0.41, 0.02], *p*(β<0) = 0.96). Effects of priming were also greater at SOAs of 60-65ms and 75-80ms compared to the 30-35ms group, with 99% and 95% confidence, respectively, although this comparison is less reliable because there were very few experiments with the longest SOAs.

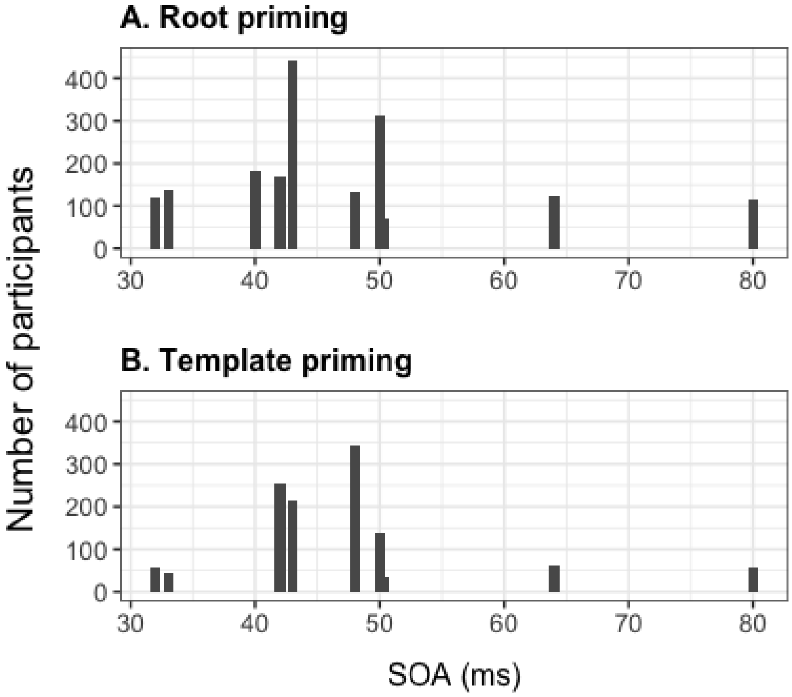

Figure 6. *Number of participants in masked visual priming experiments across the range of Stimulus Onset Asynchrony (SOA) values (ms) investigating roots (Panel A) and templates (Panel B).*

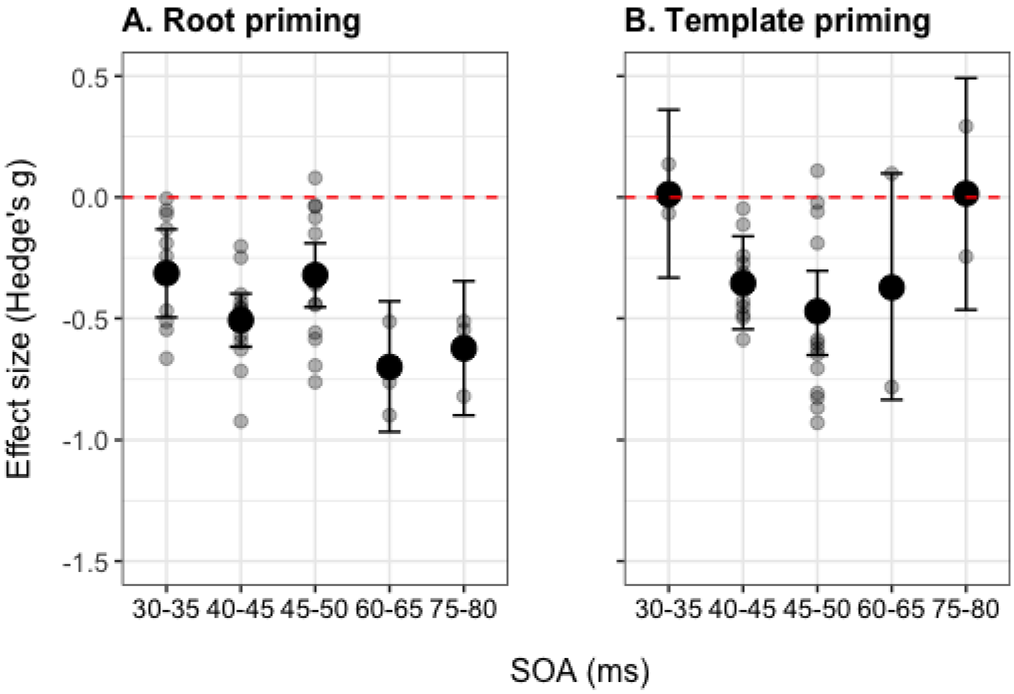

Figure 7. *Effect sizes for different SOA values in root priming (Panel A) and template priming experiments (Panel B). Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.*

3.2.2.2 Template priming is affected by SOA but not prime and target modality. The extent to which template priming is modulated by methodology was also evaluated by comparing effect sizes across the three paradigms: auditory priming (7 experiments, 262 subjects), visual priming (31 experiments, 1206 subjects), and cross-modal priming (14 experiments, 576 subjects). Because all visual priming experiments on template priming used masked primes, we are unable to distinguish the effects of modality of presentation from that of prime masking effects.

Template priming as well was credible in all paradigms, as shown in Figure 5B. The effect size was numerically the largest for auditory priming ($\beta = -0.75$, 95% CrI [-1.18, -0.33]), followed by cross-modal priming ($\beta = -0.57$, 95% CrI [-0.86, -0.28]), with smallest effects in (masked) visual priming ($\beta = -0.35$, 95% CrI [-0.55, -0.15]). However, the effect size was only credibly larger when auditory priming was compared to visual priming ($\beta = -0.40$, 95% CrI [-0.88, 0.06], $p(\beta<0) = 0.96$). The larger effect size in auditory priming, however, needs to be interpreted with caution. This paradigm was used least often, therefore its effect size estimate is the most unstable as indicated by the very large credible interval (see the supplementary materials on the OSF page for Bayesian estimates of more nuanced comparisons between methods).

We also examined the effect of SOA on masked visual template priming experiments. The same five SOA bins (see Figure 6B) were used here as with root priming: 30-35ms (4 experiments, 101 subjects), 40-45ms (10 experiments, 468 subjects), 45-50ms (13 experiments, 517 subjects), 60-65ms (2 experiments, 63 subjects), and 75-80ms (2 experiments, 57 subjects). Unlike for root priming, credible template priming with 100% credibility was only observed in two SOA bins (40-45ms: $\beta = -0.35$, 95% CrI [-0.55, -0.16]; 45-50ms: $\beta = -0.47$, 95% CrI [-0.65, -0.30]), with near credible priming observed for an SOA of 60-65ms ($\beta = -0.37$, 95% CrI [-0.84, 0.10], $p(\beta<0) = 0.94$). There was no credible priming effect for the shortest 30-35ms SOA ($\beta =$

0.01, 95% CrI [-0.34, 0.36], $p(\beta>0) = 0.52$) or the longest SOA of 75-80ms ($\beta = 0.02$, 95% CrI [-0.47, 0.49], $p(\beta<0) = 0.52$). The results are shown in Figure 7B.

3.2.2.3 No difference in root or template priming effects found with lexical decision or naming tasks. Finally, we evaluated if there were any differences in effect size related to the dependent variables measured in these priming experiments. Reaction times measured in a lexical decision task were by far the most common dependent variable in the priming experiments included in this meta-analysis, with naming latencies used in some others. Since all experiments measuring naming latencies involved masked visual priming (4 experiments, 192 subjects), we compared their effect size to reaction times measured in lexical decision tasks that also involved masked visual priming (43 experiments, 1660 subjects). As shown in Figure 8A, we found comparable, credible root priming effects using reaction times from lexical decision tasks ($\beta = -0.44$, 95% CrI [-0.52, -0.37]) and for naming latencies ($\beta = -0.36$, 95% CrI [-0.59, -0.13]). Thus, there are no systematic differences in root priming effects based on the choice of the response mode in masked visual priming experiments.

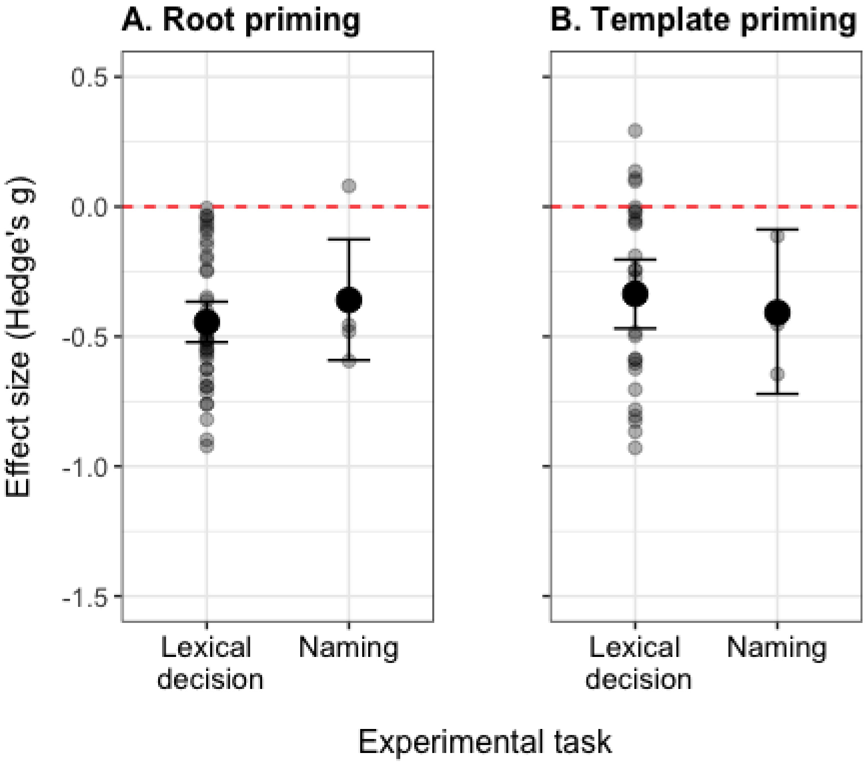

Figure 8. *Effect sizes of root (Panel A) and template priming (Panel B) for lexical decision and naming latency tasks. Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.*

Next, a comparison of template priming effects in experiments using reaction times or naming latencies was also restricted to the masked visual priming methodology. We compared reaction time measures from 27 lexical decision experiments (1014 subjects) and latency measures from 4 naming experiments (192 subjects). As shown in Figure 8B, we again found credible priming for both tasks (lexical decision: $\beta = -0.34$, 95% CrI [-0.47, -0.20]; naming: $\beta = -0.41$, 95% CrI [-0.73, -0.09]), with no credible difference between the two. Thus, there are no systematic differences in template priming effects either based on the choice of the response mode in masked visual priming experiments.

3.3 Part III: Substantive effects on root priming

3.3.1 Root priming does not differ across word class or language

We now turn to a discussion of the effects of language and word class on root priming. Unfortunately, even with a dataset of almost 87 experiments from 3930 participants, the data were too unevenly distributed to evaluate the interaction of method, word class, and language. Because there were too few auditory priming experiments, we did not include them in the analyses in this section. Instead, we first evaluated the interaction of word class and language only within visual priming experiments, where root priming effects were the smallest. We combined experiments with masked and nonmasked visual primes because there were no credible differences in root priming effects regardless of masking. We then conducted this analysis for cross-modal priming experiments as well to see if any effects of language and word class are modulated by modality of prime presentation.

Since there were very few studies investigating root priming in Maltese, we only compared priming effects in Arabic and Hebrew. We compared root priming in Arabic nouns (12 experiments; 454 subjects) and verbs (10 experiments; 288 subjects) with root priming effects in Hebrew nouns (25 experiments; 1577 subjects) and verbs (8 experiments; 258 subjects). These comparisons included experiments using naming and lexical decision tasks because there were no credible differences in effect sizes between experiments using the two tasks.

The effect sizes for root priming in both languages for nouns and verbs are presented in Figure 9. As noted, none of the credible intervals included 0, therefore root priming was credible in nouns and in verbs in both Arabic and Hebrew. Furthermore, there was no credible effect of language (β = 0.07, CrI = [-0.11, 0.26], p(β>0) = 0.78) or word class (β = -0.14, CrI = [-0.39, 0.11], p(β<0) = 0.87) or their interaction (β = 0.24, CrI = [-0.10, 0.58], p(β>0) = 0.92).

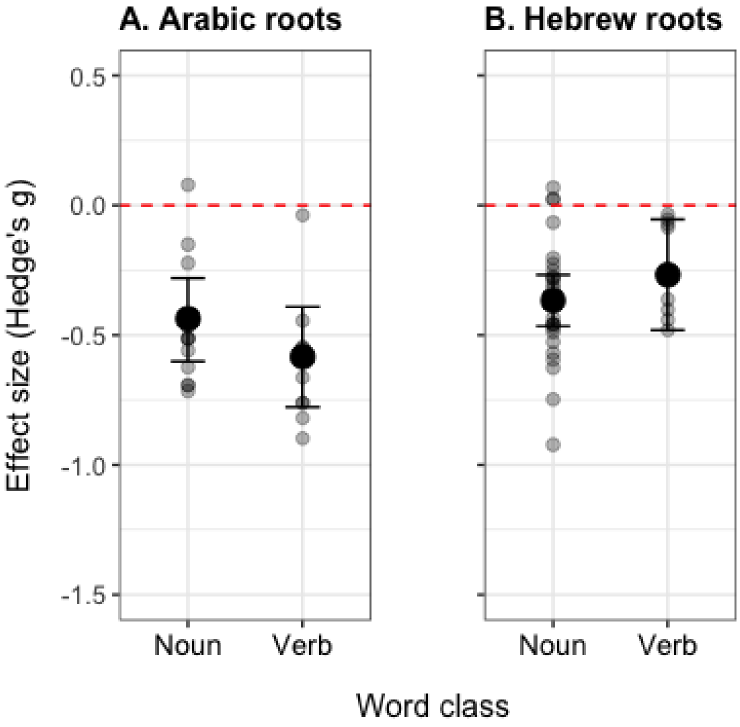

Figure 9. *Root priming effect sizes for nouns and verbs in Arabic (Panel A) and Hebrew (Panel B) for all visual priming experiments. Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.*

Next, we examined whether root priming effects differ across language and word class when tested using the cross-modal priming methodology. The number of experiments available for this comparison was much smaller: three experiments on Arabic nouns (120 subjects), none on Arabic verbs, two experiments on Hebrew nouns (120 subjects), and two experiments on Hebrew verbs (60 subjects). Root priming was credible in all of these groups (Arabic nouns: β = -0.71, CrI = [-1.13, -0.29]; Hebrew nouns: β = -0.90, CrI = [-1.39, -0.43]; Hebrew verbs: β = -0.60, CrI = [-1.13, -0.07]). We found that root priming in Arabic nouns was not credibly different from that in Hebrew nouns (β = 0.19, CrI = [-0.47, 0.83], p(β>0) = 0.75), which, in turn, was not different from root priming in Hebrew verbs (β = 0.29, CrI = [-0.48, 1.03], p(β>0) = 0.82). No such comparison was possible for auditory priming, where all studies were on Arabic nouns. Though limited by the small number of studies using the cross-modal priming methodology, the converging results indicate that root priming effects are similar in visual and cross-modal priming paradigms, with no differences across languages or word class.

3.3.2 Root priming is independent of both semantic and form overlap

Next, we investigated whether root priming is simply a result of shared meaning and form between two root-related words or whether it represents morphological effects that can be disentangled from the two. Independence of root priming from semantic overlap is typically investigated in two different ways, both of which we investigated. Because there were no credible differences across languages and word class, analyses in this section included them all.

In one method, root-related prime-target pairs that have a transparent semantic relationship (e.g., Hebrew *madrix* ‘guide’ and *hadraxa* ‘guidance’) are compared to those with an opaque semantic relationship (e.g., Hebrew *drixut* ‘alertness’ and *hadraxa* ‘guidance’). If priming is observed in root-related pairs with an opaque semantic relation, it cannot be attributed to semantic overlap.

A comparison of priming effects in root-related pairs that have an opaque (8 experiments, 448 subjects) or transparent (13 experiments, 656 subjects) semantic relationship is provided in Figure 10. As expected, root priming effects were credible when there was a transparent semantic relationship between primes and targets ($\beta = -0.75$, CrI = [-0.91, -0.59]). It was also credibly greater when compared to priming of root-related prime-target pairs with an opaque semantic relationship ($\beta = -0.32$, CrI = [-0.56, -0.08]), which might be expected given that prime-target pairs with a transparent semantic relationship additionally overlap in meaning. We also replicated differences in root priming effects across methods. Consistent with the finding in section 3.2.2, the extent of root priming differed across methods, with the smallest effect in visual priming experiments ($\beta = 0.35$, CrI = [-0.09, 0.81], $p(\beta>0) = 0.94$). Further, there was no interaction between the type of semantic overlap and method (auditory vs. cross-modal: $\beta = 0.40$, CrI = [-0.25, 1.06], $p(\beta>0) = 0.89$; auditory vs. visual: $\beta = 0.23$, CrI = [-0.41, 0.86], $p(\beta>0) = 0.76$).

Crucially, however, root priming was still credible when there was an opaque semantic relationship between primes and targets ($\beta = -0.42$, CrI = [-0.62, -0.24]). This shows that root priming does not result solely from overlap in meaning. However, there still remains the question of whether the priming in opaque roots is morphological in nature or merely due to form overlap alone.

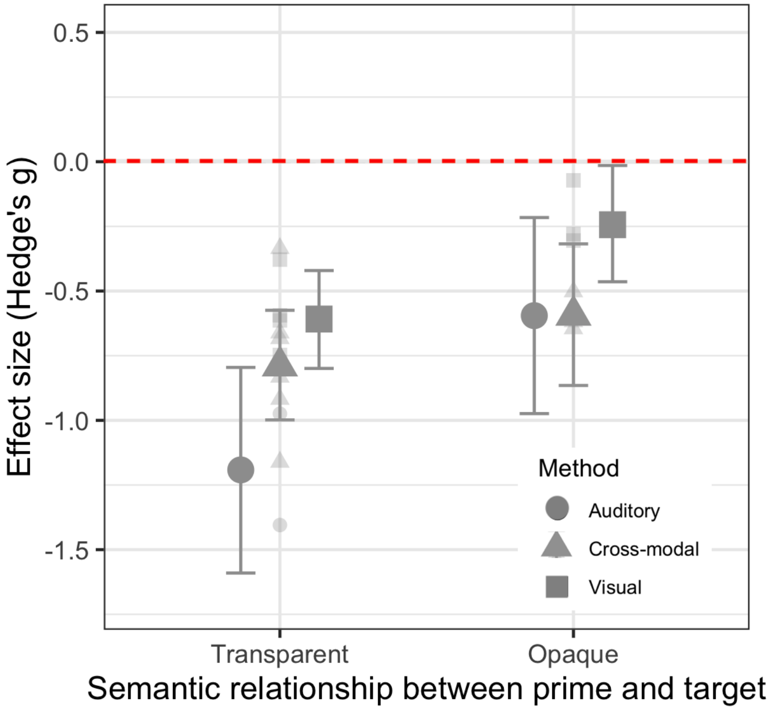

Figure 10. *Root priming effect sizes for experiments with either an opaque or transparent semantic relation between primes and targets, across modalities. Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.*

To rule out the role of form overlap, we need to compare the extent of priming in cases of morphological overlap (which always co-occurs with form overlap) with the extent of priming in cases where there is a comparable degree of form overlap (but no morphological overlap). However, in nearly all studies that were compared in Figure 10, prime-target pairs in control trials with form overlap shared fewer segments than did root-related primes and targets. Therefore, if root priming effects are found to be significantly greater than form priming effects, it could well be attributed to the additional shared segments in root priming conditions. Thus we need a different method to disentangle morphological effects from both form and meaning overlap.

We now turn to the second method used to investigate the independence of root priming from semantics: masked visual priming. In masked visual priming experiments, primes are presented for such a brief period of time that semantic access is interrupted before the presentation of the target, thus minimising semantic priming effects (Frost et al., 1997; Perea et al., 1995). First, to confirm that masking the prime minimises the influence of semantics, we compared the extent of priming in prime-target pairs with opaque or transparent semantic relations compared to unrelated controls (Figure 11A). The extent of masked visual priming was comparable regardless of whether the root relationship was semantically transparent or opaque ($\beta = -0.12$, CrI = [-0.48, 0.24], $p(\beta<0) = 0.74$). In contrast, there was a difference in the extent of priming between root-related prime-target pairs depending on whether or not there was a transparent semantic relationship when nonmasked priming methodologies were used (Figure 10). Therefore, we can confirm that masking the primes does in fact minimise the influence of semantics.

The most conservative test of a dissociation between morphological overlap from form overlap would entail a comparison of the root priming against the priming effects obtained for prime-target pairs that share the same number (i.e., three) of overlapping segments in a masked priming paradigm (thus minimising any effects of meaning overlap). If the extent of priming for morphologically-related pairs is greater than the extent of priming for morphologically-unrelated pairs with an equivalent amount of form overlap, then this would be evidence for a morphological priming effect over and above form and meaning overlap. However, we did not find enough experiments to make such a comparison.

Instead, we exploited the variation in the kinds of control items used across masked visual priming experiments to get at this comparison. Recall that priming is measured as the facilitation in reaction time or naming latency relative to some control condition. In the root priming experiments in our dataset, control stimuli were either completely unrelated to targets or had some degree of form overlap, where the extent of overlap ranged from two or less to three shared consonants, with some additional vowels. This variation allowed us to compare the extent of root priming obtained when controls were completely unrelated vs the extent of root priming when control prime-target pairs had an overlap of three segments, same as the root-related condition.

We first examined the effect of the degree of form overlap in controls (unrelated: 14 experiments, 422 subjects; three-segment overlap: 8 experiments, 375 subjects) for root-related target-prime pairs with a transparent semantic relationship in masked visual experiments. There was no credible difference between the extent of root priming whether they were compared to controls that were completely unrelated or which had three-segment overlap ($\beta = 0.12$, CrI = [-0.12, 0.35], $p(\beta>0) = 0.84$), as shown in Figure 11 (see the left sides of Figures 11A and 11B). Further, based on the fact that for masked visual experiments we found no credible difference in the extent of root priming between prime-target pairs that had a transparent versus opaque semantic relationship (Figure 11A), it can be argued that the priming found for prime-target pairs with an opaque semantic relationship cannot be attributed to form overlap (right side of 11B).

For a more direct test, for root-related prime-target pairs with an opaque semantic relationship, we compared the extent of masked visual priming when control pairs were completely unrelated (10 experiments, 302 subjects) to the extent of root priming when control pairs had an overlap of three segments (2 experiments, 87 subjects), shown on the right sides of

Figure 11A and 11B. Again, there was no credible difference between the extent of root priming whether control pairs were completely unrelated or shared three segments ($\beta = -0.12$, CrI = [-0.50, 0.26], $p(\beta<0) = 0.73$). Crucially, as shown in Figure 11B (right), when compared to control pairs with three-segment overlap, root-related pairs with opaque semantic relations still primed credibly ($\beta = -0.37$, CrI = [-0.71, -0.02]), although the credible interval was large due to the small sample size. This credible priming effect is not expected if root priming is simply an additive effect of form and semantic priming.

We can be sure that these priming effects cannot be attributed to overlap in meaning because these are masked visual priming experiments for prime-target pairs with an opaque semantic relationship. If morphological priming is obtained solely because root-related pairs overlap in form and meaning, in contexts where meaning overlap is irrelevant (i.e., masked priming of pairs with opaque semantic relations), then there should be no difference in the extent of facilitation between root-related pairs and morphologically-unrelated pairs with the same amount of form overlap; as a result, root-related pairs should not show any priming effect when compared to controls with three-segment overlap. Thus, the fact that credible priming is obtained in this context is evidence that morphological relation, especially when this relation is in the form of root morphemes, plays a special role in lexical representation, distinct from effects of shared form and meaning.

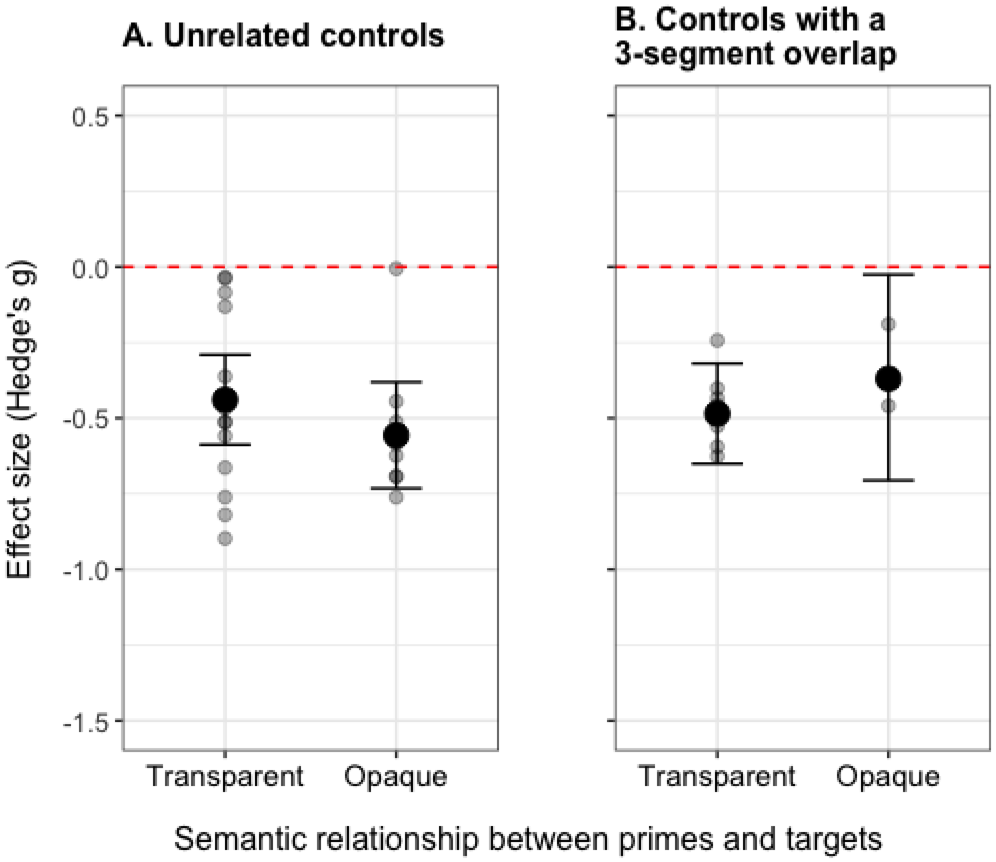

Figure 11. *Root priming effects for experiments where the semantic relationship between primes and targets was transparent vs. opaque compared to unrelated controls (Panel A) and controls with 3-segment overlap (Panel B). Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.*

3.4 Part IV Substantive effects on template priming

3.4.1 Verbal templates prime more than nominal templates in both Arabic and Hebrew

Next, we examined the effects of language and word class on template priming. As discussed in section 1.1, template priming has been reported to be typically more robust for verbs compared to nouns. These effects are often attributed to differences in type frequency, since verbal templates have a higher overall type frequency than nominal ones (e.g., Deutsch et al. 1998).

Based on the findings on methodological effects (section 3.2.2.2), we excluded auditory priming experiments because template priming was credibly different for auditory priming compared to visual priming. We also excluded experiments with masked primes where the SOA was 30-35 ms, 60-65 ms, or 75-80 ms because there was no credible template priming at these SOA values. Because there were too few studies on Maltese, we were only able to compare template priming effects in Hebrew and Arabic. The final set included 16 template priming experiments on Arabic nouns (730 subjects) and 10 on Arabic verbs (357 subjects), 8 on Hebrew nouns (420 subjects) and 9 on Hebrew verbs (382 subjects). The effect sizes are plotted in Figure 12.

As expected, we found a credible effect of word class such that there was more priming for verbs than nouns ($\beta = -0.49$, CrI = [-0.80, -0.19]). Surprisingly, the effect of language was not credible ($\beta = 0.11$, CrI = [-0.21, 0.43], $p(\beta>0) = 0.75$), nor was the interaction of language with word class ($\beta = 0.03$, CrI = [-0.44, 0.51], $p(\beta>0) = 0.56$). In fact, we found a small but credible effect of template priming in Hebrew nouns ($\beta = -0.18$, CrI = [-0.33, -0.03]), despite the fact that the findings of individual studies on template priming in Hebrew nouns are inconsistent.

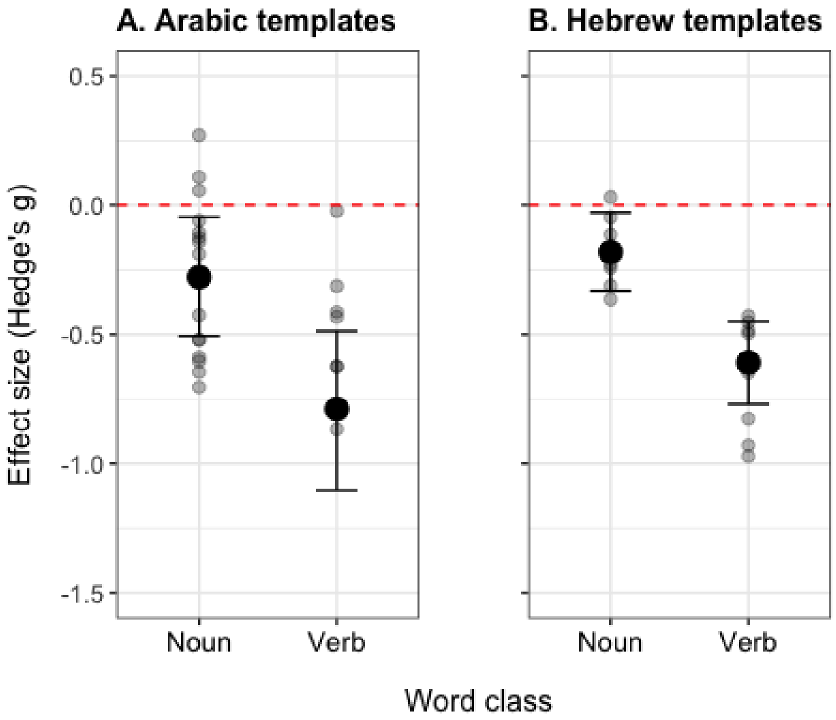

Figure 12. *Template priming effect sizes as a function of word class (noun or verb) in Arabic (Panel A) and Hebrew (Panel B). Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.*

3.4.2 Mixed evidence that template priming is independent of morphosyntactic overlap

Analogous to semantic overlap between root-related pairs, template priming effects could potentially be attributed to shared morphosyntactic properties of primes and targets. Similar to transparent and opaque semantic relations in root-related pairs, template-related pairs can also differ in whether they share morphosyntactic properties. For example, in Arabic, *χudˤuuʕun* ‘submission’ and *ħuduuθun* ‘happening’ share the template CuCuuCun as well as the morphosyntactic property of being verbal nouns. However, *suʒuunun* ‘prisons’ also has the same template (defined here as overlap in prosodic structure) but denotes the plural form of a noun instead of a gerund, and thus the same template can form words with different morphosyntactic information.

To disambiguate the effect of morphosyntactic overlap on template priming effects, we compared the extent of priming for template-related pairs with (30 experiments, 1287 subjects) and without (2 experiments, 80 subjects) shared morphosyntactic information. Because there were differences in the extent of template priming across word class, we further subdivided template-related pairs with morphosyntactic overlap into those involving nouns (18 experiments, 836 subjects) versus verbs (12 experiments, 451 subjects). For this analysis as well, we excluded all auditory priming experiments, given the credible difference in the effect size (section 3.2.2.2). Additionally, we also excluded all experiments where masked visual primes were presented with an SOA of 30-35 ms, 60-65 ms, or 75-80 ms because there was no credible template priming at these SOA values (section 3.2.2.2).

As shown in Figure 13, there was no credible priming when template-related pairs did not share morphosyntactic information (β = -0.19, 95% CrI [-0.77, 0.38], $p(\beta<0)$ = 0.75), but we found credible effects for morphosyntactically-related pairs regardless of whether they involve shared nominal templates (β = -0.29, 95% CrI [-0.46, -0.12]) or verbal templates (β = -0.70, 95% CrI [-0.92, -0.48]). Morphosyntactically-related verb pairs primed credibly more than both related noun pairs (β = -0.41, 95% CrI [-0.69, -0.14]) and morphosyntactically-unrelated pairs (β = -0.51, 95% CrI [-1.07, 0.04], $p(\beta<0)$ = 0.97). Additionally, there was no credible difference in the extent of priming between related noun pairs and morphosyntactically-unrelated pairs (β =

-0.10, 95% CrI [-0.64, 0.44], $p(\beta<0) = 0.64$), despite the fact that the former primes credibly and the latter does not. This is due to the small sample size of the latter group, causing a very large credible interval which subsumes the credible interval of the former group (this can be seen in Figure 13). While we currently do not find credible priming for template-related pairs without shared morphosyntactic information, we cannot dismiss the potential that, with more experiments, this effect may indeed end up credible, though with a small effect size, similar to that for the morphosyntactically-related noun group.

The effect of word class on template priming (i.e., that verbal templates prime more than nominal ones) is likely due to type frequency, which is higher for verbal templates. This may also explain the absence of priming in template-related pairs without morphosyntactic overlap. In the two experiments that make up this group (from the same paper: Boudelaa & Marslen-Wilson, 2015) that do explore this relationship, the templates used were not highly productive. Since one of the experiments was on nouns and the other on verbs (experiments 1 and 2, respectively), we could not assess the affect of morphosyntactic overlap and word class simultaneously. Therefore, we need experiments where word class and, more importantly, type frequency is carefully controlled before we can argue one way or the other about whether template priming effects arise independent of overlap in morphosyntactic information.

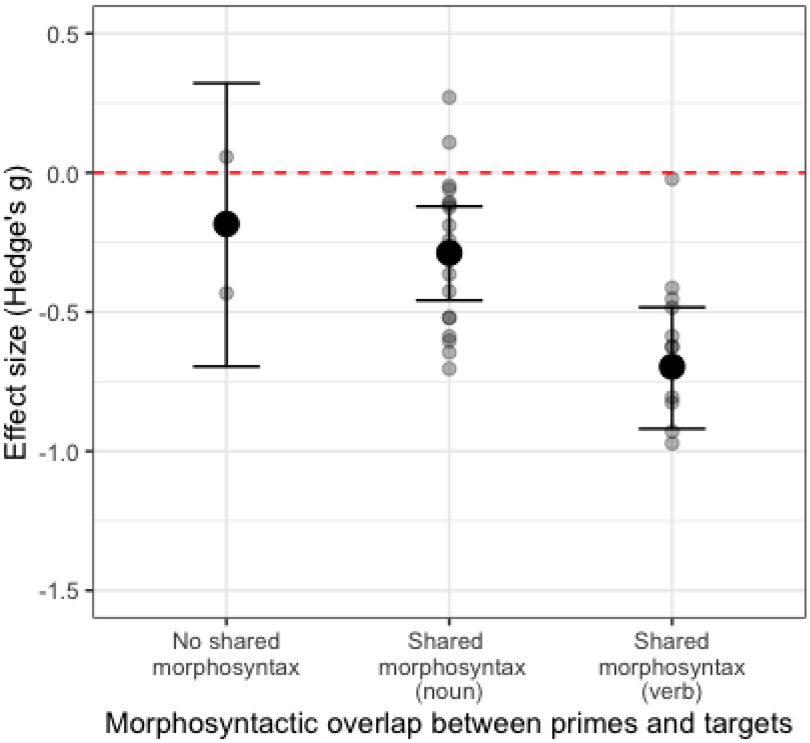

Figure 13. *Template priming effect sizes plotted as a function of whether there was morphosyntactic overlap between prime-target pairs. Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.*

3.4.3 Template priming is independent of form overlap

Next, we present analyses to dissociate template priming effects from the effect of form overlap. The most conservative test would entail a comparison of the extent of template priming (33 experiments, 1407 subjects) with the extent of priming between pairs of words with a near complete overlap in prosodic structure and vowels. There were only three such experiments (99 subjects).

Again, we excluded experiments with SOAs of 30-35 ms, 64 ms, and 80 ms because no credible template priming was observed for these SOAs, and we excluded auditory priming experiments because of credible differences based on the modality of stimulus presentation. Since vowels and prosodic information (e.g., vowel length) are typically not represented in the orthography of either Arabic or Hebrew, one might expect distinct form priming results from visual priming experiments versus those from cross-modal priming experiments. Recall that in the former, primes are presented visually (using orthography) whereas in the latter primes are auditory. However, since we had only three experiments with near complete form overlap between morphologically-unrelated prime-target pairs (two visual and one cross-modal) we were not able to make this distinction.

As shown in Figure 14, no credible priming was found for prime-target pairs with solely form overlap ($\beta = 0.14$, 95% CrI [-0.33, 0.62], $p(\beta>0) = 0.73$), whereas priming for template-related pairs was credible ($\beta = -0.46$, 95% CrI [-0.61, -0.31]) and was additionally credibly greater than the priming for pairs with just form overlap ($\beta = -0.60$, 95% CrI [-1.09, -0.11]). These results show that template priming effects cannot be reduced to effects of form overlap, because form overlap, when it occurred in the form of shared prosodic structure and vowels, did not prime when it did not constitute a template.

The general finding from many different languages is that form overlap alone between primes and targets typically does not result in a facilitation (i.e., priming) effect and often even result in an inhibition effect (e.g., Slowiaczek & Hamburger, 1992; see Dufour, 2008 for a review). We find that form overlap of vowels and prosodic information caused neither facilitation nor inhibition (Figure 14). Similarly, we indirectly confirmed that the three-consonant overlap in morphologically-unrelated control pairs also caused neither facilitation nor inhibition by the fact that there was no credible difference in the root priming effect when root-related pairs were compared to unrelated versus form-related control pairs (Figure 11).

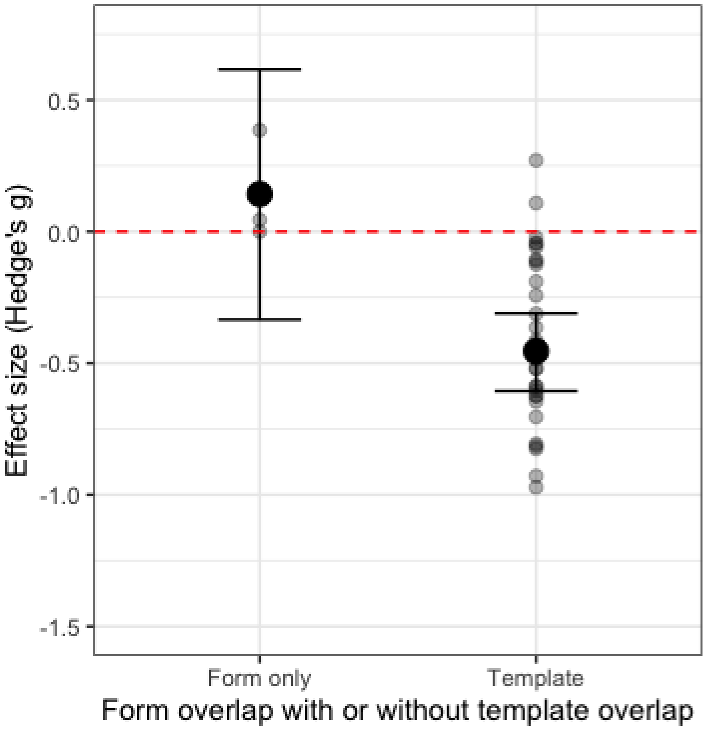

Figure 14. *Template priming effects compared to the effects when prime-target pairs overlap in form alone. Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.*

While we were able to distinguish template priming from form priming and showed that there is no reliable way to distinguish template priming from priming due to overlap in morphosyntactic information in our dataset, we were not able to examine whether template priming is credible when both morphosyntactic and form overlap are controlled for. More experiments are required to make this comparison, specifically those investigating priming in template-related pairs without morphosyntactic overlap with respect to form-related pairs where the extent of overlap is comparable to that in templates, in order to dissociate the effects of template priming from effects obtained from shared form and morphosyntactic properties.

Additionally, it may be useful to use priming experiments to investigate the potentially different contributions that various subparts of templates may make for morphological priming. For example, the Arabic template taCaaCaC can be further decomposed into a combination of the syllabic shape of its stem CVVCVC, the vocalic melody [-aa-a-], and the prefix *ta-*. In our coding of template overlap, we used a very general definition of template which often included all three subparts. More experiments controlling for various types of template overlap are needed in order to investigate which subpart or combination of subparts best represents the morphological aspect of the template.

4 General Discussion

In this paper, we aggregated data from 4710 unique participants in 229 experiments on Semitic languages to assess the evidence for priming by nonconcatenative morphemes, namely roots and templates. With Bayesian modelling of the meta-analytic effect size we were able to draw inferences about experimental design for future studies, cross-linguistic differences in root and template priming, and implications for their representations. We discuss each in turn.

First, the findings from the meta-analysis provided useful guidance for future priming experiments. Nonconcatenative morphological priming effects were largest when using the auditory priming methodology, followed by cross-modal priming, and were smallest when using visual priming. However, priming effects were insensitive to task, with similar effect sizes for naming and lexical decision. Additionally, only template, not root priming was sensitive to SOA. Specifically, in masked visual priming, template priming was credible only when the SOA was between 40- and 50ms.

Second, we found credible root priming in both Arabic and Hebrew, for nouns and verbs, with no consistent differences in the effect sizes across languages or word class. We also found converging evidence for a dissociation of root priming from priming due to overlap in form and meaning. The most compelling evidence in support of the dissociation came from credible priming in two experiments where root-related pairs had an opaque semantic relationship and control conditions had a 3-segment overlap. In such cases, the opaque semantic relationship precludes overlap in meaning, and the 3-segment overlap in the control conditions was comparable to the root-related conditions. Thus, credible priming in such cases cannot be attributed to the effects of overlap in either form or meaning. The findings from these two experiments provide empirical support for the independent representation of roots.

Recent evidence from computational modelling, however, calls into question whether it is necessary for morphemes to be explicitly represented to account for priming effects in pairs of words with an opaque semantic relationship. Specifically, using the Linear Discriminative Learner (LDL; Baayen et al., 2018; Chuang et al. 2021), Baayen and colleagues have shown that morphological priming for pairs of words with an opaque semantic relationship in Dutch (Creemers et al., 2020) can be modelled even without representing morphemes. Chuang et al., obtained equal facilitation regardless of whether the meaning relation between pairs of morphologically-related words was transparent (e.g., *afwerpen* ‘throw off’ paired with the target *werpen* ‘throw’) or opaque (e.g., *ontwerpen* ‘design’). Input representations to the LDL included only vectors for form and meaning. They modelled priming effects as the correlation of the semantic vector of the target and the predicted semantic vector of the prime calculated from vector transformation, where closer correlation represents smaller RTs. In their model, primes that share form overlap that did not constitute a morpheme (e.g., *aanscherpen* 'sharpen') had no facilitation effect. As we can see from the example, the extent of form overlap in the form-related control condition in Creemers et al., (as in the majority of experiments in this meta-analysis), is smaller than the form overlap when words share morphemes. Thus, whether similar findings would emerge if the extent of form overlap was comparable between the form-related controls and the morphologically related words, is unclear.

Despite its limitations, Baayen et al.,'s modelling demonstrates the possibility of accounting for priming effects without representation of morphemes. Whether such an approach can be extended to account for the robust root priming effects described in this meta-analysis remains to be determined. There are several reasons why this might not be straightforward. Because the Linear Discriminant Learner relies on *n*-grams to encode information about form overlap, one obvious challenge is the fact that roots are made up of non-adjacent consonants. More generally, this highlights the challenge of encoding form overlap in cases of nonconcatenative root and template overlap (cf. Nieder et al., 2022 for an attempt to model plural templates in Maltese). When primes and targets are presented in the visual modality, at least for roots, this particular problem is alleviated. Because roots are often represented contiguously in the orthography, at least in Arabic and Hebrew, one could model form overlap effects based on letter *n*-grams. Such a solution cannot account for results from either cross-modal priming or auditory priming experiments. This is problematic because we have shown in the meta-analysis that there are no qualitative differences in root priming effects that relate to modality of stimulus presentation. Thus, not only would it be unparsimonious to posit different representations to account for visual and cross-modal compared to auditory priming, there is no empirical support for such a distinction.

A further challenge for both modelling efforts as well as empirical investigations of root priming is that with almost no exceptions the extent of form overlap in form-related conditions is less than the extent of form overlap because of shared roots. In our meta-analysis, we were able to circumvent this limitation by carefully matching the extent of form overlap among the controls to the extent of form overlap due to shared roots. But there were only 2 experiments in this category, with a resulting imprecise estimate of the effect size. To carefully tease apart the role of form overlap from root overlap we need more experiments where the form-related condition has a comparable amount of form overlap to the root-related condition.

Like root priming, template priming was also credible in both Arabic and Hebrew and for both nouns and verbs, though the effect was smaller than that of root priming. The finding that nominal templates prime credibly in Hebrew, is perhaps surprising given that there are numerous studies which do not find any evidence for nominal template priming (Deutsch et al., 2005; Frost et al., 1997; *inter alia*). It is possible that template priming effects are not consistently detected in individual experiments given their smaller effect size. However, as a result of pooling results from multiple experiments, the priming effect emerged as credible due to the increased statistical power of a meta-analytic approach.

Despite being credible, priming effects for nominal templates were still smaller than priming effects for verbal templates. This is in contrast to the finding on root priming, where there are no differences in effect size between nouns and verbs. It has been proposed that nominal template priming, at least in Hebrew, is relatively weak due to the lower productivity of nominal templates (e.g., Deutsch et al., 1998). We were unable to systematically evaluate this claim across experiments because there are few empirical investigations of the effects of productivity on template priming (for exceptions see Boudelaa & Marslen-Wilson, 2011, 2015). Further, because stimulus lists were unavailable for about half the experiments included here, and we had no access to trial-level data, we were also unable to assess productivity effects on priming within any given experiment. To determine whether differences in productivity are at the core of the template priming difference between nouns and verbs, we need both additional empirical studies as well as shared access to raw data at the level of individuals and trials.

Unlike for root priming, our meta-analysis did not provide evidence in support of a clear dissociation between template priming and priming due to overlap in form and morphosyntactic function. There was evidence that template priming effects could not be reduced to effects of form overlap only, since overlap in prosodic structure and vowels alone did not prime unless it constituted a template. However, there was no credible priming in the 2 experiments where template overlap without overlapping morphosyntactic functions was evaluated. Thus, there is no clear evidence in support of the independent representation of templates. Further, given the large error in the estimate, we cannot be sure that template priming can be obtained in the absence of morphosyntactic overlap. Rather, our findings are compatible with shared morphosyntactic function being the major contributor to template priming effects (e.g., Kastner 2019). To better understand the relationship between shared morphosyntactic function and form and distinguish it from templates, we need more empirical investigations. More generally, this raises the question of how (if at all) shared morphosyntactic function is represented in the lexicon.

Finally, even with the very large sample size in the meta-analysis, we were unable to make any general conclusions about Maltese. There were the fewest experiments in Maltese, namely three: two investigating root priming and one on template priming. Given limited data points, we cannot draw any conclusions about root and template priming, though we did see that root priming in Maltese was credible (see supplemental materials on the OSF page for details). Whether or not template priming in Maltese is really different from the other Semitic languages needs to be investigated in future research.

5 Conclusion

In this meta-analysis, we aggregated data from 4710 unique participants in 229 experiments on morphological priming effects in Semitic languages. Using Bayesian modelling, we established that robust priming effects are present for both roots and templates in Semitic languages, albeit smaller for templates than roots. In terms of methodological factors, priming effects were credible in all methodologies for both roots and templates, with the largest effects in auditory priming. More substantively, root priming did not differ across word class or language, while template priming was modulated by word class. We observed more template priming effects in verbs than in nouns in both Arabic and Hebrew. Furthermore, we found robust root priming effects which were demonstrably independent from effects of both form and meaning overlap. In contrast, while we found credible template priming that was distinct from the effects of form overlap, more empirical work, especially that on the relationship between template and morphosyntactic representation, is needed. Overall, the meta-analysis provided compellingly strong evidence for the psychological reality of nonconcatenative morphemes, particularly roots, which is not an epiphenomenon resulting from mere overlap in form and meaning. Broadly, this underscores the importance that existing implementations of morphological processing need to be able to account for the abstract nature of nonlinear morphemes.

Acknowledgements: We thank Hannah Chu, Jean-Max Fawzi, Nicole Fuches, Ethan Kahn, Rowan Konstanzer, Joyce Lin, and Eden Moyal for their help with data entry; Canaan Breiss for all of his help and advice regarding the Bayesian modelling; Hironori Katsuda and Z.L. Zhou for their aid with various aspects of the meta-analysis process; and Stefan Keine and the audience at the UCLA Linguistics Department for their feedback and discussion. This research was partially funded by NSF BCS-2214017 to MS (co-PI Connor Mayer, UCI).

Disclosure of interest: The authors report there are no competing interests to declare.

Data availability statement: The data that support the findings of this study and all supplemental material are openly available on OSF at http://doi.org/10.17605/OSF.IO/2WR5T

References

Abu-Rabia, S. & Awwad, J. (2004). Morphological structures in visual word recognition: The case of Arabic. *Journal of Research in Reading, 27*(3), 321–336.

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition, 90*(2), 119-161.

Baayen, R. H., Chuang, Y. Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de) composition but in linear discriminative learning. *Complexity*, vol. 2019, Article 4895891.

Bat-El, O. (1994). Stem modification and cluster transfer in Modern Hebrew. *Natural Language and Linguistics Theory, 12*, 571-596.

Bat-El, O. (2003). The fate of the consonantal root and the binyan in Optimality Theory. *Recherches linguistiques de Vincennes*, (32), 31-60.

Bentin, S. & Feldman, L. B. (1990). The contribution of morphological and semantic relatedness to repetition priming at short and long lags: Evidence from hebrew. *The Quarterly Journal of Experimental Psychology, 42*(4), 693-711.

Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C. & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development, 89,* 1996-2009.

Berent, I. & Pinker, S. (2007). The dislike of regular plurals in compounds: Phonological or morphological? *The Mental Lexicon, 2*(2), 129-181.

Blevins, J. P. (2016). *Word and Paradigm Morphology*. Oxford University Press.

Boudelaa, S. & Marslen-Wilson, W. D. (2000). Non-concatenative morphemes in language processing: Evidence from modern standard Arabic. In A. Cutler, J. McQueen, & R. Zondervan (Eds.), *Proceedings of SWAP (Workshop on Spoken Word Access Processes)* (vol. 1, pp. 23-26). Max-Planck Institute for Psycholinguistics.

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

Boudelaa, S. & Marslen-Wilson, W. D. (2011). Productivity and priming: Morphemic decomposition in Arabic. *Language and Cognitive Processes*, *26*(4-6), 624-652.

Boudelaa, S. & Marslen-Wilson, W. D. (2013). Morphological structure in the Arabic mental lexicon: Parallels between standard and dialectal Arabic. *Language and Cognitive Processes, 28*(10), 1453-1473.

Boudelaa, S. & Marslen-Wilson, W. D. (2015). Structure, form, and meaning in the mental lexicon: Evidence from Arabic. *Language, Cognition and Neuroscience*, *30*(8), 955–992.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software, 80*, 1-28.

Bybee, J. & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review, 22*(2-4), 381-410.

Chuang, Y. Y., Kang, M., Xuefeng, L., & Harald, B. R. (2021). Vector space morphology with linear discriminative learning. In D. Crepaldi (Ed.), *Linguistic morphology in the mind and brain* (pp. 167-183). Routledge.

Clahsen, H., Rothweiler, M., Woest, A. & Marcus, G. (1992). Regular and irregular inflection in the acquisition of German noun plurals. *Cognition, 45*(3), 225-255.

Creemers, A., Davies, A. G., Wilder, R. J., Tamminga, M., & Embick, D. (2020). Opacity, transparency, and morphological priming: A study of prefixed verbs in Dutch. *Journal of Memory and Language, 110,* Article 104055.

Cristia, A. (2018). Can infants learn phonology in the lab? A meta-analytic answer. *Cognition, 170*, 312-327.

Deutsch, A., Frost, R. & Forster, K. I. (1998). Verbs and nouns are organized and accessed differently in the mental lexicon: Evidence from Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(5), 1238-1255.

Deutsch, A., Frost, R., Pollatsek, A. & Rayner, K. (2000). Early morphological effects in word recognition in Hebrew: Evidence from parafoveal preview benefit. *Language and Cognitive Processes, 15*(4-5), 487-506.

Dufour S. (2008). Phonological priming in auditory word recognition: when both controlled and automatic processes are responsible for the effects. *Canadian journal of experimental psychology = Revue canadienne de psychologie expérimentale, 62*(1), 33–41.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629-634.

Farhy, Y. & Verissimo, J. (2019). Semantic effects in morphological priming: The case of Hebrew stems. *Language and Speech, 62*(4), 737-750.

Farhy, Y., Verissimo, J., & Clahsen, H. (2018). Universal and particular in morphological processing: Evidence from Hebrew. *Quarterly Journal of Experimental Psychology*, *71*(5), 1125-1133.

Frost, R., Deutsch, A., Gilboa, O., Tannenbaum, M., & Marslen-Wilson, W. (2000). Morphological priming: Dissociation of phonological, semantic, and morphological factors. *Memory & Cognition, 28*(8), 1277-1288.

Frost, R., Forster, K. I., & Deutsch, A. (1997). What can we learn from the morphology of Hebrew? A masked-priming investigation of morphological representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(4), 829-856.

Geary, J. A. & Ussishkin, A. (2018). Root-letter priming in Maltese visual word recognition. *The Mental Lexicon, 13*(1), 1-25.

Gonnerman, L. M., Seidenberg, M. S., & Andersen, E. S. (2007). Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology: General*, *136*(2), 323-345.

Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. D. (2019). *dmetar: Companion R Package For The Guide 'Doing Meta-Analysis in R'* (Version 0.0. 9000) [Computer software]. http://dmetar.protectlab.org/

Heitmeier, M., Chuang, Y. Y., & Baayen, R. H. (2022). How trial-to-trial learning shapes mappings in the mental lexicon: Modelling Lexical Decision with Linear Discriminative Learning. *arXiv preprint arXiv:2207.00430*. https://doi.org/10.48550/arXiv.2207.00430

Hermana, E. W., Juma, E. J., & AlJassmi, M. (2021). Parafoveal processing of orthographic, morphological, and semantic information during reading Arabic: A boundary paradigm investigation. *PLOS ONE, 16*(8), Article e0254745.

Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties*. Georgetown University Press.

Kastner, I., Pylkkänen, L., & Marantz, A. (2018). The form of morphemes: MEG evidence from masked priming of two Hebrew templates. *Frontiers in Psychology*, *9*.

Kastner, I. (2019). Inchoatives in causative clothing: Change of state in Modern Hebrew heXYiZ. *The Linguistic Review*, *36*(3), 437-451.

Lüdecke, D. (2019). *esc: effect size computation for meta analysis* (Version 0.5.1) [Computer software]. https://CRAN.R-project.org/package=esc

Lukatela, V., Gligorijević, B., Kostić, A., & Turvey, M. T. (1980). Representation of inflected nouns in the internal lexicon. *Memory & Cognition, 8*(5), 415-423.

McCarthy, J. J., (1979). *Formal Problems in Semitic Phonology and Morphology* [Unpublished doctoral dissertation]. Massachusetts Institute of Technology.

Milin, P., Feldman, L. B., Ramscar, M., Hendrix, P., & Baayen, R. H. (2017). Discrimination in lexical decision. *PLOS ONE, 12*(2), Article e0171935.

Nieder, J., Tomaschek, F., Cohrs, E., & de Vijver, R. V. (2022). Modelling Maltese noun plural classes without morphemes. *Language, Cognition and Neuroscience, 37*(3), 381-402.

Nieder, J., van de Vijver, R., & Mitterer, H. (2021). Priming Maltese plurals: Representation of sound and broken plurals in the mental lexicon. *The Mental Lexicon, 16*(1), 69-97.

Perea, M., Gotor, A., Rosa, E., & Algarabel, S. (1995). *Time course of semantic activation for different prime-target pairs in the lexical decision task* [Poster presentation]. The 36th Annual Meeting of the Psychonomic Society, Los Angeles, USA.

Prunet, J. (2006). External evidence and the Semitic root. *Morphology, 16*(1), 41-67.

R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Seidenberg, M. S. & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences, 4*(9), 353-361.

Shadish, W. R., & Haddock, C. K. (2009). Combining estimates of effect size. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 257-277). Russell Sage Foundation.

Shalhoub-Awwad, Y. (2020). The role of nominal word pattern in Arabic reading acquisition: Insights from cross-modal priming. *Scientific Studies of Reading, 24*(4), 307-320.

Shalhoub-Awwad, Y., & Leikin, M. (2016). The lexical status of the root in processing morphologically complex words in Arabic. *Scientific Studies of Reading, 20*(4), 296-310.

Slowiaczek, L. M., & Hamburger, M. B. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(6), 1239 –1250.

Smolka, E., Libben, G., & Dressler, W. U. (2019). When morphological structure overrides meaning: Evidence from German prefix and particle verbs. *Language, Cognition and Neuroscience, 34*(5), 599-614.

Smolka, E., Preller, K. H., & Eulitz, C. (2014). 'Verstehen' ('understand') primes 'stehen' ('stand'): Morphological structure overrides semantic compositionality in the lexical representation of German complex verbs. *Journal of Memory and Language, 72*, 16-36.

Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *Journal of clinical epidemiology, 54*(10), 1046-1055.

Sterne, J. A. C., & Harbord, R. M. (2004). Funnel plots in meta-analysis. *The Stata Journal, 4*(2), 127-141.

Stockall, L., & Marantz, A. (2006). A single route, full decompositional model of morphological complexity. *The Mental Lexicon, 1*(1), 85-123.

Sundara, M., Zhou, Z. L., Breiss, C., Katsuda, H., & Steffman, J. (2021). Infants' developing sensitivity to native language phonotactics: A meta-analysis. *Cognition, 221*, 104993.

Taft, M. (1988). A morphological-decomposition model of lexical representation. *Linguistics, 26*(4), 657-667.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology Section A, 47*(4), 745-765.

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

Twist, A. E. (2006). *A psycholinguistic investigation of the verbal morphology of Maltese* [Unpublished doctoral dissertation]. University of Arizona.

Ussishkin, A. (1999). The inadequacy of the consonantal root: Modern Hebrew denominal verbs and output–output correspondence. *Phonology, 16*(3), 401-442.

Ussishkin, A. (2005). A fixed prosodic theory of nonconcatenative templatic morphology. *Natural Language and Linguistic Theory, 23*(1), 169-218.

Ussishkin, A., Dawson, C. R., Wedel, A., & Schluter, K. (2015). Auditory masked priming in Maltese spoken word recognition. *Language, Cognition and Neuroscience*, *30*(9), 1096-1115.

Vasishth, S., Mertzen, D., Jäger, L. A., & Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language, 103,* 151-175.

Table 1: Arabic root χbz, in combination with various templates

Word	**Meaning**	**Template**	**Morphosyntactic information**
χabaz	‘to bake’	CaCaC	verb-3.M.SG.PAST
χabːaːz	‘baker’	CaCːaːC	M.SG noun relating to a profession
maχbaz	‘bakery’	maCCaC	M.SG noun relating to location

Figures captions:

- **Figure 1.** Distribution of the number of participants across different priming modalities (auditory, cross-modal, masked visual, and nonmasked visual) in Arabic, Hebrew, and Maltese . Data from root priming experiments are shown in panel A while data from template priming experiments are in panel B.
- **Figure 2.** Funnel plot of the effect size (Hedge's g) against 1/Standard Error. Each dot represents a single experiment in the meta-analysis. The red dotted line indicates the median effect size, while the paler lines indicate the 95% Credible Interval.
- **Figure 3.** Forest plot of median and 95% Credible Interval for effect sizes for each experiment evaluating root priming. The pooled effect size and credible interval are plotted on the bottom row and indicated by the black vertical lines. The red dashed line marks an effect size of zero.
- **Figure 4.** Median and 95% Credible Interval for effect sizes for each experiment evaluating template priming. The pooled effect size and credible interval are plotted on the bottom row and indicated by the black vertical lines. The red dashed line marks an effect size of zero.
- **Figure 5.** Effect sizes for root priming (Panel A) and template priming (Panel B) plotted as a function of priming methodology. Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.

- **Figure 6.** Number of participants in masked visual priming experiments across the range of Stimulus Onset Asynchrony (SOA) values (ms) investigating roots (Panel A) and templates (Panel B).
- **Figure 7.** Effect sizes for different SOA values in root priming (Panel A) and template priming experiments (Panel B). Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.
- **Figure 8.** Effect sizes of root (Panel A) and template priming (Panel B) for lexical decision and naming latency tasks. Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.
- **Figure 9.** Root priming effect sizes for nouns and verbs in Arabic (Panel A) and Hebrew (Panel B) for all visual priming experiments. Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.
- **Figure 10.** Root priming effect sizes for experiments with either an opaque or transparent semantic relation between primes and targets, across modalities. Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.

- **Figure 11.** Root priming effects for experiments where the semantic relationship between primes and targets was transparent vs. opaque compared to unrelated controls (Panel A) and controls with 3-segment overlap (Panel B). Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.
- **Figure 12.** Template priming effect sizes as a function of word class (noun or verb) in Arabic (Panel A) and Hebrew (Panel B). Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.
- **Figure 13.** Template priming effect sizes plotted as a function of whether there was morphosyntactic overlap between prime-target pairs. Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.
- **Figure 14.** Template priming effects compared to the effects when prime-target pairs overlap in form alone. Point estimates are the medians of the posterior distribution for each group, with error bars marking the boundaries of the 95% Credible Interval. Each dot represents a single experiment's effect size in the model; the horizontal red dashed line marks an effect size of zero.