Journal of Memory and Language 146 (2026) 104709

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Novel dynamics in semantic ambiguity processing: insights from a Semitic language

Sami Boudelaa[a,*], Noha Fathi[b], Sameh Al Ansary[b]

[a] Department of Cognitive Sciences, UAEU, the United Arab Emirates
[b] Department of Phonetics, University of Alexandria, Egypt

ARTICLE INFO

Keywords:
Semantic ambiguity
Morphological decomposition
Lexical Decision
Semantic Categorization
Arabic lexicon

ABSTRACT

This study investigates the effects of semantic ambiguity in Arabic, a Semitic language with a unique root–pattern morphological structure, using lexical decision and semantic categorization tasks. Across five experiments, the study contrasted the processing of polysemous and homonymous words with that of unambiguous controls, examining how task demands modulate ambiguity effects. These findings reveal an early-stage processing advantage for both polysemes and homonyms during morphological decomposition, facilitated by shared root activation. In tasks emphasizing deeper semantic processing or the recombination of morphological components, polysemes demonstrate a robust advantage owing to cooperative dynamics among related senses, whereas homonyms show a processing disadvantage. Task specificity further shapes outcomes: smaller, well-defined categories minimize homonym competition, whereas broader categories amplify it. These results highlight the interplay between morphological decomposition and recombination in shaping lexical access and clarifies how cross-linguistic structure affects ambiguity processing.

Introduction

A persistent question in psycholinguistic research is how semantic factors influence the reading process. In this respect, one prominent line of investigation has come to focus on semantic ambiguity, examining how multiple interpretations of ambiguous words affect response times. Many early studies on semantic ambiguity did not distinguish between polysemous words, which have multiple related senses (e.g., *twist*: “twist one’s ankle” vs. “a twist in a story”), and homonymous words, which have distinct and unrelated meanings (e.g., *bank*: “financial institution” vs. “side of a river”) (e.g., Gottlob et al., 1999; Hino & Lupker, 1996; Hino et al., 1998; Jastrzembski, 1981; Kellas et al., 1988; Lichacz et al., 1999; Millis & Button, 1989; Rubenstein et al., 1970). The mainstay of those studies is the “ambiguity advantage”, the observation that ambiguous words, including polysemes and homonyms, were processed faster than words with fewer interpretations (e.g., *food, tent*) across tasks, such as lexical decision and naming. Subsequent research began to recognize and explore the distinction between polysemes and homonyms, revealing more nuanced processing patterns for the two word types. In lexical decision research, polysemous words generally show a processing advantage (i.e., faster reaction times, sometimes accompanied by lower error rates) compared to unambiguous controls, whereas homonyms are either processed like unambiguous controls or show disadvantages (higher latencies and error rates) (Beretta et al., 2005; Klepousniotou & Baum, 2007; Klepousniotou et al., 2008; Locker et al., 2003; Rodd et al., 2002; 2004).

In semantic categorization tasks, the results differ slightly. No polysemy advantage is observed; however, homonyms show a processing time disadvantage when decisions are relatively difficult, such as when participants have to determine whether a given word (e.g., *bank, bark*) belongs to a large semantic category (e.g., *man-made objects*). When semantic categorization involves a narrow and well-defined

* Corresponding author.
E-mail address: s.boudelaa@uaeu.ac.ae (S. Boudelaa).

https://doi.org/10.1016/j.jml.2025.104709
Received 8 March 2025; Received in revised form 15 October 2025; Accepted 20 October 2025
Available online 28 November 2025

category (e.g., *vegetables* or *animals*), no processing disadvantage is observed (Forster, 2004; Hino et al., 2006).

Several models have been proposed to explain the effects of semantic ambiguity, following two notable approaches. The first is a connectionist approach with different instantiations. For example, the Attractor Network Model (Rodd et al., 2004), explains how the mental lexicon processes semantic ambiguity by structuring representations based on relatedness. For polysemous words with related senses (e.g., twist), the model posits broader, unified attractor basins that enable shallow activation states and facilitate quick recognition. Homonymous words (e.g., bank, bark) with unrelated meanings occupy deeper and narrower basins, leading to competition and slower processing. Another connectionist model, the Semantic Settling Dynamics (SSD) model (Armstrong & Plaut, 2008, 2016), extends this approach by focusing on temporal dynamics. The SSD highlights how polysemous word senses cooperate early in processing, offering an advantage, whereas homonymous meanings compete, and show a disadvantage, particularly at later processing stages. This model also links these effects to task-specific factors, such as difficulty in distinguishing nonwords or degraded stimuli, showing how such conditions can amplify the processing dynamics of ambiguity.

The second approach is theory-driven and outlined in the Response Configuration Model (Hino et al., 2006). This model challenges the connectionist perspective by attributing ambiguity effects to task-dependent response selection rather than differences in semantic representations. On this account, all semantic meanings of a word are activated; however, task demands influence how the response system resolves these activations. For example, categorizing a polysemous word such as *paper* into a broad category (e.g., *non-living thing*) is easier because its related senses align, simplifying response selection. In contrast, homonyms such as *bank*, with unrelated meanings, create conflicting activations, complicate categorization tasks, and lead to a processing disadvantage. This model underscores how the cognitive system adapts its response strategies to semantic and task contexts, offering a distinct perspective on the effects of ambiguity.

A key shared feature of these models is the assumption that meaning resolution occurs after lexical access to whole-word forms, a view well suited to morphologically simple words (e.g., *bank, bark, twist*) in Indo-European languages. In contrast, in Arabic, words are systematically decomposed into roots and word patterns before lexical access (Boudelaa & Marslen-Wilson, 2005; Boudelaa et al., 2023; Boudelaa et al., 2024; Boudelaa et al., 2025; Perea et al., 2010; Prunet and Béland, 2000; Saiegh-Haddad, 2003). This decomposition stage may shape the ambiguity resolution process itself: polysemous words may benefit from shared root activation that facilitates integration, whereas homonyms may experience an early advantage followed by a late disadvantage owing to the divergent meanings conveyed by the same root. Investigating ambiguity in Arabic offers a unique opportunity to test how morphological structure modulates the dynamics of semantic processing and whether current ambiguity resolution models can be generalized to morphologically complex languages.

In the remainder of this study, we first introduce semantic ambiguity in Arabic, highlighting the interaction between morphological and semantic knowledge. Then, we outline the key experimental considerations that motivated our use of lexical decision and semantic categorization tasks. We follow by presenting five experiments that systematically examine how Arabic morphological structure modulates the processing of polysemous and homonymous words. Three lexical decision experiments assessed early-stage decomposition effects, whereas two semantic categorization experiments examined meaning integration. We conclude with a discussion of the theoretical implications of our findings for ambiguity resolution models.

Semantic ambiguity in Arabic

Arabic lexical items are formed through the intricate interplay of consonantal roots, which carry core semantic content, and word patterns, which provide phonological and grammatical structure (Boudelaa & Marslen-Wilson, 2015; Boudelaa et al., 2024). For example, the word jalas"[1] consists of the root {jls}, meaning to "*sit down,*" and the" جلس word pattern {CaCaC}, which defines its phonological structure as consonant–vowel sequence stressed on the first syllable, with the morpho-syntactic functions of *perfective, active, singular verb* (Boudelaa & Marslen-Wilson, 2005; Boudelaa et al., 2024; Saiegh-Haddad & Henkin-Roitfarb, 2014). Roots combine with different patterns to produce numerous surface forms, which may be unambiguous, polysemous, or homonymous. For instance, the root {jls} can combine with the pattern {CaCyC} to form the polysemous word "jalys," which has multiple related senses: "*associate,*" "*companion,*" "*comrade,*" "*friend,*" and "*person with whom someone sits*." The root {dqq} can also be combined with the same pattern {CaCyC} to produce the homonymous word "daqyq," which has multiple unrelated meanings: "*critical,*" "*punctual,*" "*delicate,*" "*subtle,*" "*small,*" "*flour,*" and "*careful*.".

This taxonomy of Arabic words as consisting of root morphemes interleaved with word pattern morphemes, is supported by evidence from behavioral, imaging, and neuropsychological research (Boudelaa & Marslen-Wilson, 2004, 2005, 2011, 2015; Boudelaa et al., 2019; Frost, 2012; Frost et al., 1997; Seghier & Boudelaa, 2024; Boudelaa et al., 2025). Overt priming (e.g., cross-modal, and auditory-auditory) and covert priming (masked priming) techniques provide ample evidence that word pairs sharing a root morpheme significantly facilitate each other. When two words share a root, significant priming is observed, regardless of whether they share a transparent semantic relationship (e.g., "maktab"- "kitAb" *office-book*) or an opaque semantic relationship (e.g., "maktab"- "katybah" *office-squadron*) (Boudelaa & Marslen-Wilson, 2005, 2015; see also Frost et al., for similar results in Hebrew e.g., Frost et al., 1997; Frost et al., 2000). Similarly, words sharing a pattern (e.g., "xaraja"- "jalasa" *go out-sit down*) significantly prime each other in overt and covert priming tasks (Boudelaa & Marslen-Wilson, 2004, 2005, 2011). Neuroimaging and neuropsychological research corroborate the prominent role of morphological units in Semitic languages. For example, Gwilliam and Marantz (2015) used magnetoencephalography to show that root predictability correlates with neural activity in superior temporal regions, and that pseudo-words violating the Obligatory Contour Principle[2] constraint (e.g., *qaqara*) were easier to reject as valid words than frequency-matched counterparts (e.g., *vadaha*). Similarly, neuropsychological research suggests that speakers of Semitic languages can show selective impairment of either roots (Boudelaa et al., 2023; Prunet and Béland, 2000), or word patterns (Barkai, 1980).

Beyond their prominent role in word recognition, morphological units, especially roots, affect processing early. Boudelaa and Marslen-Wilson (2005) examined the timing of morphological effects in Arabic word recognition using masked priming across multiple stimulus-onset asynchronies (SOAs). They found that root morpheme effects emerged as early as SOA 32 ms and persisted across all tested SOAs (up to 80 ms). In contrast, word pattern effects appeared later, were transient and detectable only at SOAs of 48 and 64 ms. On the other hand, orthographic and semantic priming effects occurred only at the longest SOA

[1] Throughout this work, we will primarily use the Buckwalter transliteration system in quotes to represent Arabic words and morphemes in Latin script for clarity and consistency. Braces will be used to indicate morphemes (i.e., root and word pattern), and within word patterns, we will use the letter 'C' as a placeholder for root consonants. Buckwalter's scheme is available here: https://en.wikipedia.org/wiki/Buckwalter_transliteration. In cases where precise phonetic detail is necessary, we will provide International Phonetic Alphabet (IPA) transcriptions in /slashes/. However, in order to minimize typesetting complications, IPA will be used only when the Buckwalter transliteration is insufficient for disambiguation.

[2] The Obligatory Contour Principle (OCP) prohibits adjacent identical elements in phonological or morphological structures, notably preventing consonant repetition in Semitic roots (McCarthy, 1981).

(80 ms). These results underscore the early and sustained influence of roots on Arabic visual processing, separating them from the later, transient effects of form or meaning relationships. Boudelaa et al., (2010) reached a similar conclusion for the auditory modality, extending evidence for the early prominence of root morphemes. Using mismatch negativity (MMN) responses, they found that root effects emerged 160 ms after the deviation point, marked by early and bilateral fronto-central activation. In contrast, the word pattern effects appeared later, at 250 ms, with a left-lateralized topography. These findings confirm the early and dominant role of the root in the auditory domain, underscoring its central function in Arabic language processing across modalities.

Based on this, Boudelaa (2014) and Boudelaa and Marslen-Wilson (2015) proposed the Obligatory Morphological Decomposition (OMD) model, in which the comprehension of Arabic words, both written and spoken, is mediated by processes that segment roots and word patterns to access the lexicon. Importantly, roots and patterns are segmented both in fully transparent, semantically combinatorial words such as "katab" *write,* where the overall meaning revolves around the root's general sense of *writing,* and in semantically non-combinatorial words like "katybah" *squadron*, where surface meaning departs significantly from the core meaning of the root. To recover the idiosyncratic meaning of "katybah"-type words, the OMD posits a *recombination stage* after initial obligatory decomposition (cf., Rastle & Davis, 2008; Taft, 2004).

Extensive evidence for obligatory morphological decomposition in Arabic, particularly the early and sustained roles of roots and word patterns in lexical access, raises important questions regarding how these processes interact with semantic ambiguity. Because roots and patterns are automatically segmented even in semantically opaque words, morphological complexity likely shapes the processing of polysemy and homonymy. This form-meaning interplay is absent in Indo-European languages, where ambiguity resolution operates on morphologically simple words. Thus, while prior models of semantic ambiguity focused on how multiple meanings compete or cooperate during lexical access, Arabic presents a unique case where ambiguity resolution occurs alongside (and may be constrained by) morphological decomposition.

To address these issues, we investigated whether Arabic polysemes and homonyms diverge in their reliance on morphological decomposition during meaning resolution. Specifically, do shared roots unify polysemous senses to facilitate processing, while homonyms, despite their shared morphology, trigger competition among unrelated meanings? Task demands are equally critical: Does semantic categorization, which hinges on recombination, amplify the homonym disadvantage compared with lexical decision that may tap onto early root extraction processes? Our experiments dissect these dynamics by manipulating (1) ambiguity type (polysemy vs. homonymy) and (2) task requirements (lexical decision vs. semantic categorization), offering the first systematic test of how morphological and semantic forces interact in Semitic language ambiguity resolution.

If root extraction facilitates lexical access, as predicted by OMD, we expect both polysemous and homonymous words to show a processing advantage in tasks emphasizing early decomposition, such as lexical decision. Both word types activate shared root morphemes, providing a rapid route to the lexicon. However, in tasks requiring semantic integration, such as semantic categorization, diverging patterns may be expected: polysemous words should maintain an advantage through cooperative activation of related senses, whereas homonymous words may be disadvantaged by competition among unrelated meanings during recombination.

Experimental considerations

To investigate how Arabic morphological structure modulates semantic ambiguity, we used two complementary experimental paradigms: lexical decisions and semantic categorization. These tasks are methodologically well established in psycholinguistics and theoretically aligned with the distinct stages of lexical access and semantic integration central to ambiguity resolution (Balota & Yap, 2011; Forster, 2004; Goh et al., 2016; Hino et al., 2002; Kinoshita & Norris, 2012a,b; Rastle & Coltheart, 2010; Rodd et al., 2004). Lexical decision tasks are methodological cornerstones for establishing morphological decomposition in Semitic languages, such as Hebrew and Arabic, with a particular strength in detecting the early stages of word recognition processes (Boudelaa et al., 2023, Boudelaa et al., 2025; Perea et al., 2010; Frost et al., 1997). As seen above, masked priming and neurophysiological studies demonstrate that Arabic root morphemes are accessed rapidly and obligatorily, with effects observable as early as 32 ms after prime word offset. This temporal resolution makes lexical decisions ideal for examining how the polysemy-homonymy distinction interacts with morphological decomposition. Thus the lexical decision task allows us to test whether polysemous words benefit from early root activation, facilitating lexical access through a shared semantic core, while homonyms trigger competition due to their divergent meanings despite shared morphology. These questions connect directly to connectionist models, such as the Semantic Settling Dynamics model, which links early activation patterns to semantic cohesion among word senses. We further leverage this task by systematically varying nonword difficulty to test whether ambiguity effects persist when morphological parsing is strained by increasingly word-like distractors.

While lexical decision captures early morpho-lexical processing, semantic categorization offers an equally crucial view of later meaning integration, where task demands guide interpretation. Semantic categorization requires full semantic resolution, making it particularly useful for assessing leading models of ambiguity resolution such as the Response Configuration Model (Hino et al., 2006). Research shows that semantic categorization reveals distinct processing patterns for homonyms versus polysemes, depending on category characteristics. Broad categories exacerbate the homonym-disadvantage by activating competing senses, whereas narrow categories constrain interpretations and reduce this effect (Forster, 2006; Hino et al., 2006). In the Arabic context, semantic categorization is valuable for testing how the recombination processes proposed by the Obligatory Morphological Decomposition model operate under semantic pressure. This task tests whether successful semantic integration depends on the alignment between decomposed morphological structures and category membership.

In summary, the lexical decision task isolates early decomposition effects and allows the examination of how semantic ambiguity interacts with morphological processing, whereas semantic categorization reveals how ambiguity resolution is shaped by interpretive constraints during meaning integration. This dual approach systematically tests whether Arabic polysemes and homonyms diverge in their initial activation dynamics and final semantic outcomes. By capturing the full trajectory from morphological decomposition to semantic decision making, this methodological combination extends our understanding of ambiguity resolution beyond Indo-European languages and enriches cross-linguistic models of lexical access through the unique morphological architecture of Arabic.

Data availability

The raw data and analysis scripts from this study are openly available via the Open Science Framework at https://osf.io/t497p/files/osfstorage. Supplementary materials, including detailed statistical outputs for all experiments, are also provided. These outputs adhere to the recommended reporting standards for mixed-effects models as outlined by Brysbaert and Stevens (2018), and Meteyard and Davies (2020).

Experiment 1: Lexical decision with nonwords based on non-existent roots

The first experiment used a lexical decision task with nonwords created by altering one or two root letters of real Arabic words (e.g., *maktab* becomes *majtab* and *safynah* becomes *jatynah*). These nonwords lack valid roots and are straightforward to reject, thus the task primarily taps onto the early decomposition stage of word recognition. Accordingly, two processing consequences are expected: First, root extraction processes should apply to polysemous and homonymous words, yielding a semantic ambiguity advantage for both word types compared with unambiguous words. Second, the recombination stage, during which specific semantic interpretations are identified, may not have had enough time to begin; therefore, the multiple unrelated meanings of homonyms (e.g., "daqyq" *flour* and *precise*) would not yet compete. Conversely, if the recombination is immediate and unrelated meanings begin competing early, we may observe a relatedness advantage with polysemes showing faster latencies than homonyms.

Method

Participants

Fifty undergraduate students from Alexandria University, Egypt participated in this experiment voluntarily or for course credit. This yielded 1,700 observations per condition, consistent with the recommendation of Brysbaert and Stevens (2018) for lexical-decision experiments. All participants were adults with normal (or corrected-to-normal) vision and reported no reading or learning disorders. All participants provided informed consent.

Stimuli

One hundred and two non-homographic Arabic words (82 nouns and 20 adjectives) were selected from the Aralex database (Boudelaa & Marslen-Wilson, 2010) to create three conditions of 34 words each. The first condition comprised polysemous words with multiple related senses (e.g., جليس *associate, a person one sits with, friend; companion; comrade, sitter*); the second consisted of homonyms with multiple unrelated meanings (e.g., دقيق *correct, difficult, excellent, flimsy, flour*). The third condition was made up of unambiguous words without multiple meanings or senses (e.g., غداء *lunch*). These words were then pre-tested on the following variables:

Number of Meanings/Senses: We determined the number of meanings and senses for each word in three complementary ways. First, we relied on authors' judgment. Second, we consulted the Modern Arabic Dictionary (almaany.com/). Although this dictionary lists entries under root morphemes, it distinguishes between multiple senses and multiple meanings via subheadings. Third, we conducted a semantic pretest. We compiled 200 items (the 102 experimental words and 98 foils) and asked 30 participants, who did not take part in any of the five experiments, to write down all the meanings and/or senses each word evoked. A meaning or sense was counted if at least five participants produced it.

Relatedness Among Meanings/Senses: We assessed relatedness for polysemes and homonyms using two complementary methods. First, we administered a similarity-judgment task in which 30 new participants rated the relatedness of the experimental word pairs intermixed with 24 non-experimental fillers. Participants used a 7-point scale, where 1 indicated "*completely unrelated*" and 7 indicated "*highly related.*" This task captured subjective human judgments of semantic similarity between word pairs. Second, we calculated the semantic distances among meanings and senses using word embeddings trained on the entire Arabic Wikipedia (approximately 350 million tokens). These embeddings provide a high-dimensional vector representation, enabling a quantitative measure of the semantic proximity. Together, the two methods yielded a reliable convergent distinction between polysemes and homonyms.

Meaning/Sense Dominance: To establish the dominant meaning/sense of each ambiguous word, we classified and tallied responses from the elicitation pretest. For each ambiguous word, any meaning/sense produced by more than 10% of participants was considered dominant. Table 1 lists the stimulus characteristics. Orthographic frequency counts were obtained from ARALEX (Boudelaa & Marslen-Wilson, 2010). The OLD20 (Orthographic Levenshtein Distance 20) quantifies orthographic similarity as the average Levenshtein distance between a target word and its 20 nearest orthographic neighbors (Yarkoni et al., 2008).

Statistical comparisons of the variables in Table 1 using Tukey's HSD tests with FDR correction showed significant differences among ambiguity-related measures (NMD, NSD, CDoMWE, RoMHR), but not among control variables such as NoI, NoS, and OLD20 (all FDR corrected p-values > 0.05; see Supplementary Table S1 for details).

Familiarity of Senses and Meanings: To assess familiarity with the meanings of homonyms and the senses of polysemes, we conducted a rating pretest on 340-item list. This list included the two most frequently suggested homonym meanings, polyseme senses, and unambiguous words, along with an additional 170 foils. Fifteen participants took part in the pre-test. They rated the familiarity of word meanings and senses on a 7-point scale, where 1 indicated *very unfamiliar* and 7 indicated *very familiar*. Semantically ambiguous words were presented once for each interpretation, with a gloss to clarify the intended meaning. For example, the Arabic word دقيق ("daqyq") was shown twice: once labeled as *flour* and once as *precise*, and participants rated each version separately. Unambiguous words and foils were presented once without additional context. The results of this test are summarized in Table 2.

Imageability of Senses and Meanings: We used the same approach as in the familiarity pretest. All word types (i.e., polysemes, homonyms, and unambiguous) and 170 foils were pre-tested for imageability by asking 30 new participants to evaluate each meaning/sense on a 7-point scale, where 1 indicated *not at all easy to visualize* and 7 indicating *very easy to visualize*. This approach quantified the concreteness of each meaning/sense, measuring how readily participants formed mental images of the experimental items. As shown in Table 2, which presents familiarity and imageability results, the experimental words were well matched on these characteristics. Statistical comparisons of the lexical characteristics in Table 2 using Tukey's HSD tests with FDR correction revealed no significant differences in familiarity or imageability between any of the compared senses or meanings (all FDR-corrected p > 0.05; see Supplementary Table S2 for details).

For the purposes of the lexical decision task, 102 nonwords were created by altering one or two root letters of existing Arabic words. This follows standard practices in Semitic languages research, where nonwords are designed to maintain realistic morphological patterns while ensuring that their roots do not exist in the language. For example, the word واحدة ("wAHdp," *one*) was transformed into the nonword باحدة ("bAHdp") by changing one root letter, and the word رفاهة ("rfAhp," *comfort*) was modified into the nonword رناءة ("rnAvp") by changing two root letters. These nonwords conform to the phonotactic and orthotactic rules of the language, and participants can quickly identify them as such because of the absence of meaningful roots, making them a good starting

Table 1
Number of Meanings in Dictionary (NMD); Number of Senses in Dictionary (NSD); Number of Meanings or Senses based on Subjects (NoMoS/S); Imageability (Img); Relatedness of Meaning from Human Ratings (RoMHR); Cosine Distance of Meaning based on Word Embeddings (CDoMWE); Number of Letters (NoL); Number of Syllables (NoS); Orthographic Frequency (OF) based on ARALEX; OLD20; and Bigram Frequency as a percentage of total bigrams (BiF).

	polysemes	homonyms	unambiguous	nonwords
NMD	1.00(0.00)	2.88(1.39)	1.00(0.00)	n/a
NSD	5.65(4.23)	4.09(3.59)	1.15(0.36)	n/a
NoMoS/S	2.00(0.00)	2.00(0.00)	1.00(0.00)	n/a
Img	3.52(1.03)	4.11(1.15)	4.14(1.17)	n/a
CDoMWE	0.58(0.11)	0.82(0.10)	n/a	n/a
RoMHR	5.97(0.87)	1.77(0.66)	n/a	n/a
NoL	4.24(0.61)	4.15(0.56)	4.53(0.86)	4.28(0.63)
NoS	2.21(0.59)	2.21(0.48)	2.21(0.48)	2.21(0.48)
OF	28.60(40.34)	28.62(40.14)	28.78(49.64)	n/a
OLD20	1.85(0.23)	1.85(0.20)	1.88(0.33)	1.92(0.08)
BIF	1.17(1.46)	1.03 (0.46)(3383)	1.14(1.49)	0.99(1.15)

Table 2
Lexical characteristics of the different senses/meanings of the stimuli.

		Familiarity	Imageability
Polysemes	Meaning 1	3.41 (0.9)	2.9 (1.1)
	Meaning 2	3.34 (1.0)	2.8 (1.0)
Homonyms	Sense 1	3.48 (0.8)	2.5 (0.9)
	Sense 2	3.26 (0.9)	2.4 (1.1)
Unambiguous	Meaning	3.41 (0.10)	2.7 (1.2)

point for assessing semantic ambiguity in a Semitic language.

Procedure

Participants were tested either individually or in groups of up to nine in a normally lit, quiet room. Each participant completed 204 trials (102 words and 102 nonwords) presented in two blocks, with a self-paced break between blocks. They performed a word–nonword decision task using a keyboard with two predefined response keys: one for “Word” and one for “Nonword.” The “Word” response was always assigned to the dominant hand, typically the right. Stimulus presentation and data collection were controlled using the DMDX software (Forster & Forster, 2003), with the stimulus display synchronized to the refresh rate of the screen (16.7 ms). Participants were seated approximately 60 cm away from the standard computer monitor on which the stimuli were presented and instructed to respond as quickly and accurately as possible. Before the experimental trials, they completed 20 practice trials consisting of 10 words and 10 nonwords to familiarize themselves with the task. The order of stimulus presentation was randomized for each participant. Each trial began with a fixation point “+” at the center of the screen for 500 ms, followed by the target for 2000 ms or until response. The entire experiment lasted for 15–20 min.

Results

We adopted the following general statistical approach for this and the subsequent four experiments: reaction time (RT) and accuracy data were analyzed using linear mixed-effects (LME) models, treating subjects and stimuli as crossed random factors (Baayen, 2008). Analyses were conducted in R (Version 4.1.3; R Core Team, 2022) using the lme4 package (Version 1.1.35.2; Bates et al., 2018) for model estimation and the lmerTest (Version 3.1.3; Kuznetsova, Brockhoff, & Christensen, 2017) for hypothesis testing. The fixed effect factor Word Type consisted of three levels (polysemes, homonyms, and unambiguous) and was referenced to the unambiguous level. We used R’s default dummy coding scheme for a number of reasons. First, it offers clear interpretability since coefficients directly represent differences from the reference category. Second, it aligns with our research questions and hypothesis-testing needs, and avoids the unnecessary complexity of alternative coding schemes in a non-factorial context. Thus, we created two contrasts addressing the effects: “Homonym vs. Unambiguous” and “Polyseme vs. Unambiguous” with the reference category “Unambiguous” coded as 0 in both. For the final contrast “Polyseme vs. Homonym,” which addresses *relatedness* effects and is not provided directly by the model, we used the linear Hypothesis function from the *car* package (Version 3.1.2; Fox & Weisberg, 2019). For each contrast, the coefficient equaled the difference between the beta coefficients of the “Homonym vs. Unambiguous” and the “Polyseme vs. Unambiguous” contrasts. This provided chi-square statistics, from which p-values were derived to determine statistical significance. Standard errors, calculated from the contrast coefficient and chi-square statistic, were used to compute 95 % confidence intervals.

In the latency analyses, error trials were excluded, and the RTs were log-transformed to meet the distributional assumptions of the LME. Cut-off points were set at 100 ms and 2000 ms to minimize data loss. In each experiment, for both latencies and accuracies, we tested linear mixed-effects models with random slopes and intercepts, simplifying the structure if they failed to converge or if added complexity did not improve fit. We began by removing correlations between random slopes and intercepts, eliminating random slopes for items, and finally retaining only random intercepts for participants and items (Barr et al., 2013).

The mean correct reaction times (RTs), error rates, ambiguity effects, and relatedness effects for Experiment 1 are summarized in Table 3a.

Latencies. Our pruning procedure eliminated 250 data points (4.94 %), of which 4.90 % were errors and 0.04 % outliers, leaving 4848 data points. We then fitted a linear mixed-effects model (estimated using REML and the optimx optimizer) to predict log-transformed reaction times (logRT) based on Word Type, with random intercepts for Target and Subject. The main effect of Word Type was highly reliable ($F(2, 99.00) = 29.74$, $p < .001$). Planned contrasts addressing the semantic ambiguity effect showed both polysemous words ($t(4842) = -7.70$, $p < .001$, beta = -0.07, 95 % CI [-0.09, −0.05]) and homonymous words ($t(4842) = -3.58$, $p < .001$, beta = -0.03, 95 % CI [-0.05, −0.02]) to be significantly faster than the unambiguous baseline. Finally, the contrast between polysemous words against homonymous words, testing semantic relatedness, also revealed statistically reliable differences ($t(4845) = 4.12$, $p < .001$, beta = -0.04, 95 % CI [-0.06, −0.02]).

Accuracies. A logistic mixed-effects model (estimated using ML and the Nelder–Mead optimizer) was fitted to predict errors, with Word Type as a fixed effect and Target and Subject as random effects. Polysemous words had a significantly lower error rate than the unambiguous condition ($z = -3.45$, $p < .001$; beta = -0.62, 95 % CI [-0.98, −0.27]), whereas the homonymous words did not ($z = -0.783$, $p = .434$). The difference between polysemes and homonyms was significant ($z = 2.69$, $p < 0.007$, beta = -0.50, 95 % CI [-0.86, −0.14]).

Discussion

Compared to unambiguous baseline words, polysemes showed a processing advantage of 41 ms, whereas homonyms showed an advantage of 21 ms (Hino et al., 2006, Experiment 1). There was also a 20 ms processing advantage for polysemes over homonyms. The processing advantages for both polysemes and homonyms, though counterintuitive in non-Semitic contexts (Rodd et al., 2004; Armstrong & Plaut, 2016), can be explained within the linguistic context of Semitic languages, where root morphemes are *the* primary units of lexical access. Boudelaa (2014) and Boudelaa and Marslen-Wilson (2005, 2015) suggested that lexical access processes in Arabic involve the systematic extraction of roots, serving as semantic cores, followed by the extraction of word pattern morphemes that convey phonological and morphosyntactic information and help converge on the specific target interpretation. On this view, polysemes and their related senses benefit because root activation reinforces overlapping conceptual features, leading to efficient processing. Similarly, despite their unrelated meanings, homonyms display processing advantages owing to the same root extraction process, which prioritizes morphological over semantic disambiguation

Table 3
A. Mean response latencies in milliseconds and percent error rate for each stimulus group in Experiment 1 (lexical decision with nonwords based on non-existent roots). For full statistical models (linear and logistic mixed-effects analyses), see Supplementary Tables S3b and S3c.

Word type	RT	% Error	Ambiguity effect		Relatedness effect	
			RT	%Error	RT	%Error
Polyseme	555 (4.49)	3.35 (0.0043)	41*	2.65*	20*	2.0*
Homonym	575 (3.68)	5.35 (0.0054)	21*	0.65		
Unambiguous	596 (5.83)	6.00 (0.0057)				
Nonword	769 (3.78)	5.08 (0.0030)				

Note. RT = mean reaction time; standard error of the mean is shown in parentheses. Asterisks indicate reliable effects.

during the early stages of word recognition (Boudelaa, 2014; Boudelaa & Marslen-Wilson, 2015).

The relatedness effect (polysemes being recognized faster than homonyms) suggests that the recombination stage, in which the root and word pattern are merged to determine the meaning of idiosyncratic and ambiguous words, begins immediately after initial morphological parsing. For homonyms like "daqyq," where the same root {dqq} maps onto multiple unrelated meanings (e.g., *flour*, *precise* etc.), competition among meanings seem to begin but does not develop into a full-fledged process, hence the advantage compared to unambiguous forms and the disadvantage compared to the polysemes (Boudelaa, 2014; Taft, 2004). Priming studies support this two-stage approach, showing that morphological relations activate shared roots regardless of semantic transparency, while semantic interpretation influences later decision-making processes (Boudelaa & Marslen-Wilson, 2005).

Finally, the use of nonwords based on non-existent roots, created by altering one or two root letters of existing words, promoted rapid responses and minimized opportunities for the recombination or integration stage to fully influence processing. This implies that the early stages of morphological processing mask potential homonym disadvantages that might otherwise emerge during extended semantic or contextual evaluation. We tested this possibility in Experiment 2 using pseudohomophones and in Experiment 3 with semantically interpretable nonwords created by illegally combining existing roots and existing word patterns.

Experiment 2: Lexical decision with pseudohomophones

Experiment 2 tested whether, with sufficient processing time, lexical dynamics in Arabic would allow semantic recombination processes to fully unfold, revealing cooperation among related senses and competition among unrelated word meanings (cf., Armstrong & Plaut, 2016). To this end, this experiment used pseudohomophones as nonwords (e.g., *brane* for *brain*). Past research shows that such nonwords slow RT's and increase error rates because they create a conflict between the orthographic representation that does not match any real word and a phonological representation consistent with an existing word (Briesemeister et al., 2009; Martin, 1982).

However, a key challenge in adapting the pseudohomophone paradigm to Arabic is its predominantly consonantal orthography, where each grapheme is consistently mapped onto a single phoneme. This strict correspondence makes the alternative spellings used in languages like English (e.g., *brane* for *brain* or *peeple* for *people*) impossible. To overcome this, we exploited the systematic phonological variations between Modern Standard Arabic (MSA) and the Egyptian Arabic dialect spoken by our participants to create the pseudohomophones. Shifts in phoneme realization in this dialect provide a basis for designing nonwords that are phonologically identical to real MSA words but are spelled with dialect-specific orthographic deviations.

For example, the MSA voiceless uvular plosive /q/ (as in مقفول /maqfu:l/, 'closed') is pronounced as a voiceless glottal stop /ʔ/ (as in /maʔfu:l/) in Egyptian Arabic. Similarly, the MSA voiceless dental fricative /θ/ is realized as either a voiceless alveolar stop /t/ (e.g., تلاتة /tala:ta/, 'three') or a voiceless alveolar fricative /s/ (e.g., سانية /sa:nya/, 'second'). We used these predictable shifts to create pseudohomophones like مأفول /maʔfu:l/ (from MSA مقفول /maqfu:l/) and تمسيل /tamsi:l/ (from MSA تمثيل /tamθi:l/, 'acting').

In the lexical decision task, this manipulation creates a direct conflict. Participants see the dialectal spelling (e.g., مأفول) but access a phonological form that matches a real MSA word (/maqfu:l/). To correctly reject the item as a nonword, they must suppress this activated MSA lexical representation. This process is akin to bilingual lexical decision, where words from a non-target language slow responses due to cross-linguistic competition (Dijkstra, Van Jaarsveld & Ten Brinke, 1998).

In standard lexical decision tasks, rapid root extraction from visually presented words facilitates quick and efficient word recognition, leaving little opportunity for extended semantic processes. However, when stimuli include pseudohomophones that differ only minimally in orthography from real words, participants must engage additional cognitive processes to distinguish real words from plausible nonwords. As previous research shows, this orthographic decision-making process leads to longer response times and more errors as the cognitive system evaluates and resolves orthographic discrepancies (Briesemeister et al., 2009; Taha & Khateb, 2013; Martin, 1982). The key question, then, is whether the anticipated increase in processing time will enable semantic processing to fully develop, engaging the recombination stage and revealing an ambiguity advantage for polysemes and an ambiguity disadvantage for homonyms.

Method

Participants

Fifty-three new undergraduate students from the same population as in Experiment 1 participated on a voluntary basis or for course credit. All participants provided informed consent and agreed to participate freely.

Stimuli

The word stimuli were the same 102 items used in Experiment 1: 34 polysemes, 34 homonyms, and 34 unambiguous words. Additionally, 102 pseudohomophones were created from phonological variations between Standard and Egyptian Arabic. For instance, the Standard Arabic words تمثيل /tamθi:l/ *acting*, مقفول /maqfu:l/ *closed*, and سطر /satˤr / *line* were modified to their dialectal counterparts as تمسيل /tamsi:l/, مأفول /maʔfu:l/, and صطر / sˤatˤr / respectively. The pseudohomophones were matched to the word stimuli for number of letters (Mean = 4.28, SD = 0.87), syllable count (Mean = 2.07, SD = 0.63), and orthographic neighborhood density (OLD20; Mean = 1.90, SD = 0.26), ensuring comparability across the word-nonword stimulus types.

Procedure

This procedure was identical to Experiment 1, except that participants were explicitly instructed to classify dialectal spellings as nonwords, accepting only correctly spelled standard Arabic items as valid words.

Results

In total, 13.13 % of the data were excluded (12.69 % errors and 0.044 % outliers), leaving 4,720 valid data points. The mean response time in this experiment (Mean = 677.33 ms, SD = 6.77 ms) was considerably longer than in Experiment 1 (Mean = 575.33 ms, SD = 4.66 ms). The overall error rate (Mean = 12.69 %, SD = 0.32 %) was also significantly higher than in Experiment 1 (Mean = 4.94 %, SD = 0.21 %). These findings indicate that the use of pseudohomophones as nonwords effectively slowed participants' responses and elicited a higher rate of errors. Mean RTs, accuracy rates, ambiguity and relatedness effects are summarized in Table 4a.

Latencies. A linear mixed-effects model predicting log-transformed reaction times (logRT) by Word Type, with random intercepts for Target and Subject, revealed a significant main effect of Word Type ($F(2,89.99) = 3.92$, $p < .023$). Planned comparisons assessing the semantic ambiguity effect revealed significant differences between polysemes and unambiguous words (t (4690) = -2.29, p = 0.022; beta =

Table 4
A. Mean response latencies in milliseconds and percent error rate for each stimulus group in Experiment 2 (lexical decision with pseudohomophones). Detailed model outputs are in Supplementary Tables S4b and S4c.

Word type	RT	% Error	Ambiguity effect		Relatedness effect	
			RT	%Error	RT	%Error
Polyseme	659 (6.16)	8.77 (0.006)	21*	8.49*	34*	3.27
Homonym	693 (7.03)	12.04 (0.007)	−13	5.22		
Unambiguous	680 (7.11)	17.26 (0.008)				
Nonword	902 (5.64)	36.76 (0.006)				

Note. RT = mean reaction time; standard error of the mean is shown in parentheses. Asterisks indicate reliable effects.

-0.05, 95 % CI [-0.09, −0.0069]). Differences between homonyms and unambiguous words were non-significant (t (4690) = 0.23, p = 0.816; beta = 0.004, 95 % CI [-0.04, 0.05]). The semantic relatedness effect, contrasting polysemes with homonyms, was significant (t (4693) = 2.54, p = 0.0128; beta = -0.05, 95 % CI [-0.09, −0.01]).

Accuracies. A logistic mixed-effects model predicting errors, with Word Type as a fixed effect and Target and Subject as random effects, revealed a significant negative effect for polysemous words compared to the unambiguous words (z = -3.51, p < .001; beta = -0.94, 95 % CI [-1.47, −0.41]). Homonyms did not differ significantly from either the unambiguous baseline (z = -1.79, p = .072; beta = -0.47, 95 % CI [-0.98, 0.04]) or from polysemes (z = 1.73, p = .082; beta = -0.47, 95 % CI [-0.81, −0.13]).

Discussion

Experiment 2 shows that the use of pseudohomophones slows reaction times and increases error rates, confirming their role in introducing competition between dialectal and MSA representations at a lexical level. Participants must suppress the dialectal interpretation to reject these items as MSA nonwords, akin to bilingual lexical decision tasks where non-target language words slow responses due to cross-linguistic competition (Dijkstra, Van Jaarsveld & Ten Brinke, 1998; Vanlangendonck et al., 2020). This decision-stage conflict prolongs response times, increases error rates, and requires more extensive processing to resolve the competition, thus providing an opportunity for semantic processes to unfold following the recombination stage.

Polysemes were processed faster than homonyms and unambiguous words, supporting the idea that the related senses of polysemes cooperate to facilitate quicker and more accurate semantic resolution (Rodd et al., 2002; 2004; Armstrong & Plaut, 2016). In contrast, homonyms showed no reliable disadvantage compared to unambiguous words, challenging the idea that the unrelated meanings of homonyms create competition that delays processing and increases errors. These findings suggest that while polysemes enjoy a clear processing advantage, the anticipated disadvantage for homonyms does not reliably emerge, even under conditions designed to prolong processing time.

Although the error rate for unambiguous words reached 17.26 %, this likely reflects a genuine processing difficulty, specifically phonological-orthographic conflict, rather than data degradation. Comparable error rates (>20 %) have been observed in German lexical decision tasks with pseudo-derived words (e.g., Bölte et al., 2025), where high error rates signify increased cognitive demand rather than noise. Critically, in our study, the elevated error rates did not compromise latency data, as sufficient observations remained for the RT analyses. The higher error rate for unambiguous words likely stems from limited access to semantic or morphological cues that could otherwise resolve decision conflicts.

The absence of a homonym disadvantage could have different causes. One possibility is the sustained influence of bottom-up activation at the root level, which continues to drive processing even at later stages. In Arabic, the root morpheme, which is central to lexical access, carries core semantic and structural information that appears to remain active throughout the word-recognition process (Boudelaa & Marslen-Wilson, 2005; 2015; Prior & Markus, 2014). Unlike in languages where homonymy introduces competition at the lexical–semantic level, the Arabic root system seems to provide a stabilizing context that reduces competition among words. Thus, the absence of a homonym disadvantage may reflect the unique role of morphological knowledge in prioritizing structural coherence over purely semantic resolution.

Alternatively, the absence of a reliable disadvantage may stem from the artificial nature of the pseudohomophones used in this experiment. Our pseudohomophones were Standard Arabic words written with Egyptian Arabic sounds (e.g., ممسل /mumassil/ for ممثل /mumaθil/ *actor*). These stimuli may have engaged additional processes, such as heightened attention to orthographic verification or language control mechanisms, which could have masked the expected effects of semantic competition.[3]

To address this limitation and further investigate the role of the root as a unit that mitigates lexical competition, Experiment 3 uses nonwords formed by the illegal combination of existing roots (e.g., {ktm} meaning *hide*) with existing word patterns (e.g., {CaCyC} *singular masculine, noun expressing intensity, permanence, or a descriptive attribute*), producing semantically interpretable nonwords such as *katym,* which can be understood as referring to *someone or something that hides a lot*. These nonwords are expected to be particularly challenging to reject due to their familiar roots and word patterns and their semantic plausibility. Functionally analogous nonwords in English, such as “mirths”—created by combining the existing stem “mirth” with the suffix ~ *s*—have been shown to pose a great challenge, requiring more detailed analysis to determine that the stem and suffix combination does not form a valid word (Taft, 2004; see also Longtin & Meunier, 2005 for similar data in French). In this study, these semantically plausible nonwords were expected to slow response latencies and increase error rates while minimizing interference from language control mechanisms, thereby allowing semantic competition to emerge.

Experiment 3: Lexical decision with semantically interpretable nonwords

The aim of this experiment was to determine whether the use of semantically interpretable nonwords can engage in deeper semantic processing, such that bottom-up activation from the root level to the word level is counteracted by competition within the word level (i.e., lateral inhibition). Specifically, this study investigates whether homonymic meanings derived from the same root exhibit measurable competition during recognition. The nonwords were created by illegally pairing valid roots (e.g., {ktm}, *hiding*, {ktb} *writing*) with valid word patterns (e.g., {CaCEyC}, {CaCwC}), producing semantically interpretable but invalid constructions such as “katym” and “katwb,” interpretable as *secretive* and *prolific writer* respectively. This approach parallels the morphologically legal but lexically invalid combinations used in Hebrew by Norman, Degani, and Peleg (2017). Such stimuli were expected to challenge participants due to their morphological and semantic plausibility, making them difficult to reject and requiring prolonged cognitive effort to disambiguate their lexical status. This extended processing should allow competition among unrelated semantic meanings to emerge while maintaining the obligatory decomposition processes characteristic of Semitic languages.

[3] See for example Frisson (2009) for a detailed discussion on the homonym disadvantage, attributed to strategic under specification during meaning resolution.

Table 5
A. Mean response latencies in milliseconds and percent error rates for each stimulus condition in Experiment 3 (lexical decision with semantically interpretable nonwords). Detailed model outputs are in Supplementary Tables S5b and S5c.

Word type	RT	% Error	Ambiguity effect		Relatedness effect	
			RT	%Error	RT	%Error
Polyseme	666 (6.48)	15.77 (0.009)	23*	8.13*	68*	1.94
Homonym	734 (6.26)	17.71 (0.009)	−45*	6.19*		
Unambiguous	689 (7.18)	23.90 (0.010)				
Nonword	914 (5.96)	45 (0.007)				

Note. RT = mean reaction time; Standard Error of the mean is shown in parentheses (). The asterisks indicate reliable effects.

Method

Participants

Forty-seven participants from the same population as in the previous two experiments participated voluntarily or for course credit. All participants were adults with normal (corrected-to-normal) vision and no reading or learning disorders. All participants provided informed consent before the experiment and agreed to participate freely.

Stimuli

The word stimuli comprised the same 102 items, with 34 polysemes, 34 homonyms, and 34 unambiguous words. Additionally, 102 orthotactically and phonotactically legal nonwords (e.g., “munfakir,” “snyd”) were created by illegally combining existing roots (e.g., {fkr} *thinking*, {snd} *supporting/relying on*) with existing word patterns (e.g., {munCaCiC} = *passive, reflexive*, {CaCyC} = *intensive*). These nonwords are semantically interpretable. For example, “munfakir” can mean “*someone who is deeply reflective, or engaged in serious thought*” and “sanyd” can mean “*someone who provides support, help, or acts as a reliable companion*.” These were matched to the word stimuli on number of letters (Mean = 4.78, SD = 0.57), syllable count (Mean = 2.39, SD = 0.55), and orthographic neighborhood density (OLD20; Mean = 1.90, SD = 0.17), ensuring comparability across the word-nonword types. These nonwords were expected to be difficult to reject, leading to slower response latencies and higher error rates. More importantly, they were expected to systematically engage the recombination stage to determine whether a combined word is valid or meaningful. As a result, competition among the multiple unrelated meanings of homonyms will have time to develop, while cooperation among the related meanings of polysemes will have time to peak or diminish.

Procedure

The procedure was identical to Experiments 1 and 2, except participants were explicitly informed that the experiment contained items that appeared very word-like because they involved existing roots and word patterns (e.g., “glyb,” consisting of the root {glb} and the word pattern {CaCyC}). Participants were instructed that while such items might seem interpretable, they were not actual words and should therefore be rejected as nonwords.

Results

A total of 19.19 % of the data were excluded from the analysis (19.17 % errors and 0.062 % outliers), leaving 3874 valid data points. The average response time in this experiment (Mean = 687.67 ms, SD = 6.96 ms) was considerably higher than that in Experiment 1 (Mean = 575.33 ms, SD = 4.66 ms). Similarly, the overall error rate in this experiment (Mean = 19.19 %, SD = 0.39 %) was significantly higher than that in Experiment 1 (mean = 4.94 %, SD = 0.21 %). These findings indicate that the nonwords used in this experiment behaved as expected effectively slowing down response times and eliciting higher error rates. Average RT’s, accuracy rates, ambiguity and relatedness effects are summarized in Table 5a.

Latencies. A linear mixed-effects model predicting log-transformed reaction times (logRT) from Word Type, with random intercepts for Target and Subject, revealed the main effect of Word Type to be significant ($F(2,95.56) = 4.55$, $p < .013$). Planned comparisons assessing ambiguity effects yielded a reliable difference between polysemous and unambiguous words ($t(3868) = -2.14$, $p < 0.032$, beta = -0.04, 95 % CI [-0.09, −3.78e-03]), and homonym and unambiguous words ($t(3868) = 2.56$, $p < 0.010$, beta = 0.05, 95 % CI [0.01, 0.09]. The effects of meaning relatedness, as revealed by the contrast between polysemes and homonyms, was statistically significant ($t(3871) = 4.74$, $p < 0.000$, beta = -0.10, 95 % CI [-0.14, −0.06]).

Accuracies. A logistic mixed-effects model predicting Error as a function of Word Type, with Target and Subject as random effects revealed the effect of polysemous words to be significant compared to unambiguous words ($z = -3.33$, $p < .001$; beta = -0.95, 95 % CI [-1.50, −0.39]). Surprisingly, homonymy showed a significant decrease in error rate compared to the unambiguous words ($z = -2.21$, $p < .027$; beta = -0.62, 95 % CI [-1.17, −0.07]). The difference between polysemes and homonyms was not significant ($z = 1.138$, $p = 0.255$; beta = -0.33, 95 % CI [-0.89, 0.24]).

Discussion

When participants had to recombine roots and word patterns to determine the specific semantic interpretation of lexical forms, the dynamics of semantic cooperation among related meanings and competition among unrelated meanings emerged in full swing. Specifically, the results reveal a clear semantic advantage for polysemes, with related meanings recognized faster and more accurately. This suggests that facilitation from a common root morpheme persists into later processing stages of polysemes. In contrast, the unrelated meanings of homonyms show a processing disadvantage compared to polysemes and unambiguous words.

Semantically interpretable nonwords highlight the impact of semantic plausibility on decision-making. Participants found these stimuli particularly challenging to reject because their morphological and semantic structures mimicked real words. This plausibility forced deeper semantic processing as participants analyzed the validity of each recombined root and word pattern. The extended processing time allowed competitive dynamics among unrelated meanings and cooperative effects among related meanings to emerge. In Experiment 4, we seek converging evidence for the cooperation and competition processes in polysemes and homonyms using a semantic categorization task with a narrow and well-defined category, namely the “Food category.”.

Experiment 4: Semantic categorization with food terms

In the previous three experiments, we used a lexical decision task and progressively increased nonword difficulty, causing participants to shift from relying on orthographic familiarity to deeper semantic processing. This shift revealed full semantic processing dynamics at play in Arabic, ranging from cooperation among the related senses of polysemes to competition among the multiple unrelated meanings of homonyms. Experiments 4 and 5 extend Experiments 1–3 by using a semantic categorization paradigm. This tested whether the previously observed effects of homonymy and polysemy persisted at deeper levels of semantic integration. As Arabic requires obligatory morphological decomposition, particularly of root morphemes, these experiments offer a unique opportunity to observe how morphological activation continues to influence meaning resolution when lexical access is followed by semantic categorization. Beginning with Experiment 4, we used a semantic categorization task that involved deciding whether a word fits within a narrow semantic category (e.g., Living Things, "Vegetables," and so on). Specifically, we use a well-defined small category, a "Food category," to establish whether cooperative and competitive processes can be observed in Arabic in a task that directly engages semantic processing. Two outcomes are anticipated here. First, it is conceivable that we replicate the results of Experiment 3 with a semantic advantage for polysemes and a semantic disadvantage for homonyms. This is because semantic categorization cannot be based on the orthographic familiarity of letter strings or the root as an orthographic/phonological unit. Instead, deep semantic processing based on the recombined words is required to determine whether a given target belongs to a given category. Alternatively, however, and as suggested by Forster (2004) and Hino et al., (2006), participants may resort to a "candidate search strategy" whereby they generate potential exemplars from the small "Food category" and compare these directly to the target. This strategy should reduce the reliance on full lexical/semantic processing and limit the scope for the emergence of full semantic processing dynamics.

Method

Participants

Fifty participants from the same population voluntarily participated in the study or received course credit. All the participants were adults with normal or corrected-to-normal vision and reported no reading or learning disorders. Informed consent was obtained from all participants.

Stimuli

Of the 102 stimulus items used in the previous experiments, we discarded two homonymic and two unambiguous words because of their close association with the Food Category. Accordingly, the stimuli consisted of 98 items, with 34 polysemes, 32 homonyms, and 32 unambiguous words. Additionally, 98 items falling within the category of food were selected including names of fruits (e.g., عنب *grapes,* رمّان *pomegranate*), vegetable (e.g., بصل *onion,* سبانخ *spinach*), grains and staples (e.g., رغيف *loaf of bread,* أرز *rice*). To establish how typical the different examples were in the category, we administered a subjective judgment pretest to 30 new participants who were asked to rate the typicality of 294 items, including 98 food words, 98 experimental items (polysemes, homonyms, and unambiguous), and 98 foils. These items were intermixed and presented in a randomized order to minimize potential order effects. Participants rated each item on a 1–7 scale, with 1 indicating *not typical of food at all* and 7 indicating *very typical of food* and were instructed to base their judgments on their intuitive sense of how representative the item was of the Food category. The results of this pretest indicated that the "food" words were all highly typical members of the category (Mean = 6.60, SD = 0.40), while polysemes, homonyms and unambiguous words received much lower typicality ratings (Mean = 1.92, SD = 0.45; Mean = 1.88, SD = 0.48; Mean = 1.89, SD = 0.53 respectively). Besides, the food words closely matched the experimental items overall in terms of the number of letters (Mean = 4.11, SD = 1.00), number of syllables (Mean = 2.09, SD = 0.70), and OLD20 (Mean = 1.79, SD = 0.33).

Procedure

Participants were asked to decide whether a word appearing on a computer monitor is the name of a "Food" by pressing either a "Yes" or a "No" key. The "Yes" response was made using the participant's dominant hand. The response accuracy and response latency from the onset of the stimulus to the participant's key-press were automatically recorded in each trial. Twenty practice trials were conducted prior to the experimental trials. The practice trials mimicked the experimental trials, consisting of 10 food-word and 10 nonfood-word trials. In all other respects, the procedure was identical to that used in the previous experiments.

Results

Performance was highly accurate, with only 1.06 % of the data excluded from the analysis (0.92 % errors and 0.12 % outliers), leaving 4848 valid data points. The average overall response latencies in this experiment (Mean = 761 ms, SE = 6.02) are much higher than even the latencies of Experiment 3 (Mean = 687.67 ms, SE = 6.96), because the responses of interest here are the "NO" responses. However, the overall error rate in this experiment (Mean = 1.06 %, SD = 0.09) was significantly lower than that observed in previous experiments. Mean RTs, accuracy rates, ambiguity and relatedness effects for this experiment are summarized in Table 6a.

Latencies. A similar linear mixed-effects model predicting log-transformed reaction times (logRT) based on Word Type, with random intercepts for Target and Subject as in the previous experiments, revealed the main effect of Word Type to be unreliable ($F(2,93.18) = 2.26$, $p = .10$). Planned comparisons showed that the contrast between Polysemes and Unambiguous words was reliable ($t(4842) = -2.13$, $p = 0.033$, beta = -0.05, 95 % CI [-0.09, −3.83e-03]). However, neither the Homonyms vs. Unambiguous contrast ($t(4842) = -1.04$, $p = 0.300$, beta = -0.02, 95 % CI [-0.07, 0.02]) nor the Polysemes vs. Homonyms contrast ($t(4842) = 1.08$, $p = .285$, beta = -0.02, 95 % CI [-0.07, 0.02]) reached significance.

Accuracies. A logistic mixed-effects model as in prior experiments revealed no significant differences either between Polysemes and Unambiguous ($z = -0.290$, $p = .772$; beta = −0.12, 95 % CI [−0.93, 0.69]), Homonyms and Unambiguous ($z = -1.915$, $p = .0555$; beta = −0.95, 95 % CI [−1.92, 0.02]), or Polysemes and Homonyms ($z = 1.66$, $p = .096$; beta = 1.62, 95 % CI [−0.29, 3.54]).

Discussion

The results of Experiment 4 contribute to our understanding of semantic categorization processes in Arabic. The high accuracy rates and relatively slower response latencies compared to the lexical decision tasks in the previous three experiments suggest that participants engaged in deeper semantic processing when determining whether a word fits within the Food category. The use of a small, well-defined semantic category likely required a recombination semantic processing stage, as participants assessed the meaning of the target word to establish category membership.

Table 6

A. Mean response latencies in milliseconds and percent error rate for each stimulus conditions in Experiment 4 (semantic categorization with food words). Comprehensive model results are in Supplementary Tables S6b and S6c.

Word type	RT	% Error	Ambiguity effect		Relatedness effect	
			RT	%Error	RT	%Error
Polyseme	744 (5.60)	1.06 (0.002)	33*	0.19	18	−0.56
Homonym	762 (6.28)	0.50 (0.001)	15	0.75		
Unambiguous	777 (6.19)	1.25 (0.002)				
Food	712 (3.48)	1.24 (0.002)				

Note. RT = mean reaction time; standard error of the mean is shown in parentheses. Asterisks indicate reliable effects.

A key finding here is the semantic ambiguity advantage observed for polysemous words, which elicit significantly faster reaction times than both homonyms and unambiguous words, highlighting the cooperative dynamics among related meanings. However, we did not observe any evidence of a processing disadvantage for homonyms in this experiment. As Forster (2004) and Hino et al., (2006) suggested, this may reflect the adoption of a "candidate search strategy," whereby participants generate exemplars from a small category, compare them directly to the target, and output their responses. This strategy likely reduces reliance on full-lexical or semantic processing, thereby diminishing competitive effects among unrelated meanings. Consequently, the ambiguity disadvantage, which is typically associated with homonyms in tasks requiring comprehensive semantic processing, may not be observed in this context. In the final experiment, we use a semantic categorization task with a broad category that should preempt the use of a "candidate search strategy" and should in principle allow us to observe a semantic ambiguity advantage for polysemes and a disadvantage for homonyms in Arabic.

Experiment 5: Semantic categorization with living thing words

In Experiment 4, we used a well-defined, small semantic category (food) which minimized the processing disadvantage for homonyms, suggesting that participants likely generated potential exemplars from the Food category and directly compared them to the target, thereby reducing reliance on full lexical/semantic processing (Forster, 2004; Hino et al., 2006). In Experiment 5, we used a large and semantically diverse category, the *Living Thing*. Unlike smaller categories, larger categories demand extensive re-combinatory processes, that is, recombining the root and word pattern to evaluate the fit of a target word within a large category. This shift in processing is expected to reveal the full extent of ambiguity effects, including advantages for polysemes and disadvantages for homonyms, thereby offering a more comprehensive understanding of semantic ambiguity in a Semitic context.

Method

Participants

Another 50 participants from the same population voluntarily participated in the experiment. All participants were adults with normal or corrected-to-normal vision and reported no reading or learning disorders. Informed consent was obtained from all participants.

Stimuli

The same set of stimuli were used as in Experiments 1–3. These consisted of 34 polysemes, 34 homonyms, and 34 unambiguous words. Additionally, 102 Living Thing items were selected including names of animals (e.g., سمكة *fish*, حمامة *pigeon*), plants (e.g., زهرة *flower*, نخلة *palm tree*), insects and small creatures (e.g., برغوث *flea*, حشرة *insect*). To determine how typical the Living Thing category was, we administered a subjective judgment pretest to 30 new participants who were asked to rate the typicality of 306 items, including 102 Living Thing words, 102 experimental items (polysemes, homonyms, and unambiguous), and 102 foils. All items were intermixed and presented in a randomized order. Participants rated each item on a 1–7 scale, where 1 indicated *not typical of Living Things at all* and 7 indicated *very typical of Living Thing*. The participants were instructed to base their judgments on their intuitive sense of how typical the item was of the Living-Thing category. The results of this pre-test indicated that Living Thing words were all highly typical members of the category (Mean = 6.79, SD = 0.46), while polysemes, homonyms, and unambiguous words received much lower typicality ratings (Mean = 1.90, SD = 0.40; Mean = 1.91, SD = 0.42; Mean = 1.88, SD = 0.45, respectively). Living Thing words were also closely matched to the experimental items in number of letters (Mean = 4.31, SD = 1.02), number of syllables (Mean = 2.17, SD = 0.66), and OLD20 (Mean = 1.85, SD = 0.38).

Procedure

This was the same as Experiment 4, except that the "Yes" response corresponded to the Living Thing item.

Results

Of the total 9,498 data points, 7.39 % were errors and 0.06 % were outliers, both excluded from the analysis. The average response latency (Mean = 823.75 ms, SE = 6.38) were significantly higher than in Experiment 4, as was the average error rate (Mean = 7.10 %, SE = 0.005). Mean RTs, accuracy rates, ambiguity and relatedness effects are summarized in Table 7a.

Latencies

A similar linear mixed-effects model predicting log-transformed reaction times (logRT) based on Word Type, with random intercepts for Target and Subject, as in the previous experiments, revealed a reliable main effect of Word Type ($F(2, 98.756) = 6.588$, $p < .002$). Planned comparison revealed the Polysemes to be indistinguishable from the Unambiguous words ($t(4716) = -0.69$, $p < 0.493$, beta = -0.01, 95 % CI [-0.06, 0.03]). Conversely, the Homonyms were statistically different from both the Unambiguous ($t(4716) = 2.74$, $p = .006$, beta = 0.06, 95 % CI [0.02, 0.10]) and the Polysemes ($t(4719) = 3.43$, $p = .00088$, beta = -0.07, 95 % CI [-0.11, −0.03]), thus revealing a semantic ambiguity disadvantage for Homonyms but no processing advantage for Polysemes.

Accuracies

A logistic mixed-effects model like the ones fitted for the previous experiments revealed no significant difference either between Polysemous and Unambiguous words ($z = -1.375$, $p = .169$; beta = -0.40, 95 % CI [-0.98, 0.17]), or between Polysemous and Homonymous words ($z = 0.877$, $p = .380$; beta = 0.27, 95 % CI [-0.33, 0.86]). The only contrast that reached significance was between the Homonymous and Unambiguous words ($z = -2.241$, $p < .025$, (beta = -0.67, 95 % CI [-1.26, −0.08]).

Discussion

Experiment 5 investigated the cognitive processing demands of semantic categorization using a large, diverse Living Things category, compared to the smaller, well-defined category used in Experiment 4. This diverse category necessitated more elaborate semantic processing to determine category membership.

The results shed light on the processing dynamics of semantic ambiguity. RTs were significantly slower than in the previous four experiments, and error rates were higher than those in Experiments 1 and 4, suggesting increased cognitive effort in processing the semantically diverse Living Things category. Specifically, while homonyms exhibited a clear disadvantage in response latencies relative to both unambiguous and polysemous words, they also yielded significantly lower error rates

Table 7

A. Mean response latencies in milliseconds and percent error rate for each stimulus group in Experiment 5 (semantic categorization with Living Things). The complete model results are in Supplementary Tables S7b and S7c.

Word type	RT	% Error	Ambiguity effect		Relatedness effect	
			RT	%Error	RT	%Error
Polyseme	830 (7.03)	7.00 (0.006)	12	1.65	35*	−0.65
Homonym	865 (7.22)	6.35 (0.005)	−23*	2.30		
Unambiguous.	842 (7.29)	8.65 (0.006)				
Living Thing	758 (3.96)	6.43 (0.003)				

Note. RT = mean reaction time; standard error of the mean is shown in parentheses. Asterisks indicate reliable effects.

than unambiguous words, a facilitation in accuracy that was statistically reliable (see Table 7a). This dual pattern—slower but more accurate responses for homonyms, and faster but more error-prone responses for unambiguous words—resembles a speed-accuracy tradeoff. However, this does not undermine the interpretation of a processing cost for homonyms. Rather than a simple strategic shift, the increased RTs for homonyms likely reflect greater semantic competition, not just a general caution or slowed response tendency. Conversely, the faster but less accurate performance on unambiguous items could reflect shallower processing under increased task demands. Thus, both measures—latency and accuracy—point to distinct cognitive dynamics associated with semantic ambiguity, and we interpret their divergence as complementary, not contradictory.

These finding underscore the cognitive cost of the semantic competition inherent in homonyms, where multiple meanings compete for recognition. Polysemes on the other hand showed no processing advantage over unambiguous words, suggesting that the ambiguity advantage diminishes over time as deeper lexical processing occurs (Armstrong & Plaut, 2016). Overall, these findings highlight the role of category scope in modulating semantic ambiguity effects. While small, well-defined categories may obscure such effects, broader, diverse categories, such as the Living Thing category, reveal the full cognitive cost of homonym processing. In addition, the absence of polyseme advantage suggests that the coactivation of related senses occurs in early processing.

General discussion

Across five experiments, this study examined how the unique morphological structure of Arabic shapes semantic ambiguity. Experiment 1 used a lexical decision task with nonwords based on non-existent roots to demonstrate a processing advantage for both polysemes and homonyms compared to unambiguous words, with polysemes showing a larger advantage. These results highlight the early stage of Arabic word recognition—morphological decomposition—where roots and word patterns are extracted, and the root serves as a primary access unit to the lexicon. This mechanism facilitates efficient lexical access for both polysemes and homonyms because the shared morphological unit supports activation regardless of semantic relatedness (Boudelaa, 2014; Boudelaa & Marslen-Wilson, 2015; Boudelaa et al., 2023, 2024; see also Frost et al., 1997).

Experiment 2 introduced pseudohomophones, increasing task complexity, and slowing down RTs. Here polysemes retained their advantage, but homonyms showed no disadvantage, suggesting that root-based decomposition continued to stabilize their processing. Experiment 3, used semantically interpretable nonwords that required recombination to verify lexical and semantic validity. Under such conditions, polysemes displayed a robust advantage due to cooperative dynamics among related meanings, whereas homonyms exhibited a marked disadvantage, as competition among unrelated meanings had sufficient time to build and emerge following recombination (Boudelaa, 2015; Hino et al., 2002; Rodd et al., 2004; Taft, 2004).

Experiment 4, using a small, well-defined Food category minimized homonym competition by enabling exemplar matching. This reduced reliance on morphological recombination processes, and allowed polysemes to retain their advantage. Conversely, Experiment 5, with its broad Living Things category, required full recombination, revealing a homonym disadvantage owing to semantic competition. The polyseme advantage diminished in the context of the broad semantic category, suggesting their cooperative dynamics lost momentum over time (Forster, 2004; Hino et al., 2002, 2006).

The use of both lexical decision and semantic categorization tasks enabled us to examine the extent to which the morphological structure of Arabic influences different levels of processing. While Experiments 1–3 demonstrated effects during lexical access, Experiments 4 and 5 revealed that morphological activation, specifically the activation of root morphemes, continued to shape performance, even during later semantic decision stages. This highlights the uniquely persistent role of morphology in Arabic and supports the idea that ambiguity resolution is not only semantic but also morphologically constrained in Semitic languages. Taken together, these findings highlight the interplay among task demands, morphological decomposition, and recombination. The lexical decision task as used here taps onto the early operations of the morphological disassembly of Arabic words into roots and word patterns. Once the root is parsed, it is used to access the lexicon; since polysemes and homonyms have one root associated with multiple related senses in one case and multiple unrelated meanings in another, it provides a basis for a semantic ambiguity advantage in both cases. Tasks that require deeper semantic processing (e.g., Experiment 5) or more stringent verification of the nonwords (e.g., Experiment 3) tap, not onto the initial decomposition process, but onto the recombination stage, where the root and the word pattern are reassembled to form a coherent interpretation of the word (Boudelaa, 2014, Boudelaa & Marslen-Wilson, 2015; cf., Taft, 2004). In the case of non-transparent words, where meaning is not combinatorial, as in the case of polysemes and homonyms, this stage is particularly crucial for establishing a fine-grained interpretation of the word. At this stage, the related senses of the polysemes are not sufficiently related to cooperate, while the unrelated meanings of the homonyms are divergent enough to compete; hence, the homonym disadvantage and the polyseme lack of advantage.

We have yet to discuss the behavior of unambiguous words, which, with the exception of Experiment 1, consistently occupied an intermediate position in reaction times across the experiments. Indeed, these words were systematically slower than polysemes but faster than homonyms in the more demanding tasks (Experiments 3 and 5). To account for this, a feedback mechanism involving the interactions between semantic and lexical representations may be helpful (Kawamoto, 1993; Kawamoto, Farrar & Kello, 1994; Pexman & Lupker, 1999; Pexman, Lupker & Hino, 2002). Under this scenario, unambiguous words may initially receive less recurrent feedback because of their unitary meanings, which limits the reinforcement of activation pathways and leads to slower recognition compared to polysemes in early processing stages (Experiment 1). In contrast, polysemous and homonymous words activate broader, more interactive networks. As task difficulty increases and recombination processes become necessary (Experiments 3 and 5), the lack of overlapping or competing representations for unambiguous words may hinder their adaptability, preventing them from benefiting from the cooperative dynamics that aid polysemes, yet allowing them to avoid the full competitive cost that hinders homonyms. This results in their stable, intermediate level of processing.

Theoretical implications

In this section, we consider the implications of our findings for the three models we mentioned in the introduction, namely the Attractor Network Model (Rodd et al., 2004), the Semantic Settling Dynamics (SSD) model (Armstrong & Plaut, 2008, 2016), and the Response Configuration Model (Hino et al., 2006).

The Attractor Network Model (Rodd et al., 2004) conceptualizes meanings as stable states (attractors) within a high-dimensional semantic space. Specifically, polysemes (e.g., *twist*) share a single, broad, and shallow attractor basin, enabling faster recognition, as the network can more easily settle into these overlapping representations, whereas homonyms (e.g., *bark*) form distinct, shallow attractor basins, causing interference and slower recognition owing to competition between meanings. Our findings are consistent with this model. The polyseme advantage in the early processing stages is predicted by this account. In contrast, the behavior of homonyms that showed an advantage over unambiguous words in Experiment 1 is clearly at variance with this model, which does not leave any room for co-activation among the multiple unrelated meanings of a homonym. Incorporating a morphological pre-processing layer to capture root activation may address this shortcoming by directly acknowledging root-based decomposition, particularly in the early processing stages of Arabic.

The second model, the Semantic Settling Dynamics, or (SSD) (Armstrong & Plaut, 2008, 2016), explains the effects of semantic ambiguity as arising from the temporal dynamics of semantic processing. Here, polysemous words with overlapping semantic features benefit from cooperative dynamics early, leading to processing advantages. In contrast, homonymous words with distinct and competing meanings experience growing interference over time, resulting in processing disadvantages. This model accounts for the emergence of a homonym disadvantage as the task becomes more difficult (e.g., Experiments 3 and 5). However, it fails to capture the task-dependent effects evident in homonym stability (Experiment 2) or the diminished polyseme advantage in broad semantic categorization tasks (Experiment 5). The most challenging result for this model though is the processing advantage for Homonyms in Experiment 1. Like the Attractor Network model, however, the SSD could, in principle, be extended to incorporate a morphological layer where Arabic words are disassembled into roots and word patterns to capture the root-based homonym advantage.

The final model, the Response Configuration Model (Hino et al., 2006), proposes that semantic ambiguity effects stem not from differences in the semantic representations themselves, but from how these representations are mapped onto the response system across tasks. Hino et al., (2006) argued that the relatedness of meanings or senses influences task-specific configurations: homonyms with distinct meanings may introduce competition, whereas polysemes with related senses can simplify response decisions. The configuration of the response system, modulated by task demands, determines whether ambiguity yields processing advantages or disadvantages. Our results align with this perspective, particularly when explaining the disadvantages of homonyms in larger semantic categories (Experiment 5). However, the consistent polyseme advantage across tasks suggests that morphological factors contribute substantially beyond response strategies. As this model focuses on response configuration, it is difficult to incorporate a pre-semantic morphological stage into the framework to accommodate the full scope of the current findings.

In summary, our findings illuminate the limitations and strengths of existing models in explaining semantic ambiguity effects in Arabic, a language with a rich morphological structure. The Attractor Network and SSD models capture the interplay of competition and cooperation but lack the necessary morphological component to explain early-stage morphology-driven effects. Conversely, the Response Configuration Model provides valuable insights into task-dependent variability but lacks information about linguistic structure needed to address cross-task consistency. Collectively, our findings underscore the need to adapt cognitive models to the linguistic typology of diverse languages, ensuring that they are sufficiently general to account for the universal aspects of semantic processing while remaining specific enough to capture the idiosyncratic properties of different languages. Finally, we note that the tasks in this study were temporally coarse, providing a single data point per word per participant. Future research should use more fine-grained experimental techniques, such as EEG or MEG, to capture the dynamic neural processes involved in semantic ambiguity resolution in Arabic, enabling a more precise understanding of its temporal progression and spatial localization.

Declarations

This research was supported by the United Arab Emirates University (UAEU) through two grants: UPAR Grant No. G00003452 and CHSS Grant No. G00003887, both awarded to Sami Boudelaa. Additional research support was provided by the Dean's Office of the College of Humanities and Social Sciences (CHSS) at UAEU.

Consent to participate

All participants provided informed consent before taking part in the study. The nature, purpose, and procedures of the research were explained in detail, ensuring participants understood their rights, including the right to withdraw at any time without penalty. Participants were assured of the confidentiality of their data and its use solely for research purposes.

Consent for publication

All participants provided explicit consent for the publication of anonymized data and findings derived from this study.

Code availability

The code used for analysis and supporting the findings of this study is openly available at: https://osf.io/t497p/files/osfstorage.

Open access

This manuscript is being submitted to the Journal of Memory and Language to be considered for publication. If accepted, the authors intend to make the final published version available in accordance with the journal's open-access policies.

CRediT authorship contribution statement

Sami Boudelaa: Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Noha Fathi:** Writing – original draft, Resources, Project administration, Investigation, Data curation. **Sameh Al Ansary:** Writing – original draft, Supervision, Software, Resources, Project administration, Investigation.

Ethics approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2025.104709.

Data availability

The data and materials supporting the findings of this study are openly available at (https://osf.io/t497p/files/osfstorage

References

Armstrong, B. C., & Plaut, D. C. (2008). Settling dynamics in distributed networks explain task differences in semantic ambiguity effects: Computational and behavioral evidence. In Proceedings of the annual meeting of the cognitive science society (Vol. 30, No. 30).

Armstrong, B. C., & Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language,*

Cognition and Neuroscience, 31(7), 940–966. https://doi.org/10.1080/23273798.2016.1171366
Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge: Cambridge University Press.
Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science, 20*(3), 160–166. https://doi.org/10.1177/0963721411408885
Barkai, M. (1980). Aphasic evidence for lexical and phonological representations. *Undena.* Publications.
Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001
Bates, D., Mächler, M., Bolker, B., & Walker, S. (2018). *lme4: Linear mixed-effects models using Eigen and S4* (Version 1.1-17).
Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: An MEG study. *Cognitive Brain Research, 24*(1), 57–65. https://doi.org/10.1016/j.cogbrainres.2004.12.006
Bölte, J., Ravenschlag, A. N., Rehbein, M., Roesmann, K., Junghöfer, M., & Zwitserlood, P. (2025). Neural correlates of masked priming: Only morphologically derived words facilitate lexical decisions. *Language, Cognition and Neuroscience, 40*(2), 221–238. https://doi.org/10.1080/23273798.2024.2419414
Boudelaa, S. (2014). Is the Arabic Mental Lexicon Morpheme-Based or Stem-Based? Implications for Spoken and Written Word Recognition. In: Saiegh-Haddad, E., Joshi, R. (eds) Handbook of Arabic Literacy. Literacy Studies, vol 9. Springer, Dordrecht. doi: 10.1007/978-94-017-8545-7_2.
Boudelaa, S. (2015). The differential time course for consonant and vowel processing in Arabic: Implications for language learning and rehabilitation. *Frontiers in Psychology, 5*, 1557. https://doi.org/10.3389/fpsyg.2014.01557
Boudelaa, S., & Marslen-Wilson, W. D. (2004). Allomorphic variation in Arabic: Implications for lexical processing and representation. *Brain and Language, 90*(1–3), 106–116. https://doi.org/10.1016/S0093-934X(03)00424-3
Boudelaa, S., & Marslen-Wilson, W. D. (2005). Discontinuous morphology in time: Incremental masked priming in Arabic. *Language and Cognitive Processes, 20*(1), 207–260. https://doi.org/10.1080/01690960444000106
Boudelaa, S., & Marslen-Wilson, W. D. (2011). Productivity and priming: Morphemic decomposition in Arabic. *Language and Cognitive Processes, 26*(4–6), 624–652. https://doi.org/10.1080/01690965.2010.521022
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. https://doi.org/10.1080/23273798.2015.1048258
Boudelaa, S., Boujraf, S., Belahcen, F., Benzagmout, M., & Farooqui, A. (2023). Impaired morphological processing: Insights from multiple sclerosis. *Language, Cognition and Neuroscience, 38*(9), 1237–1250. https://doi.org/10.1080/23273798.2023.2226267
Boudelaa, S., Carreiras, M., Jariya, N., et al. (2025). SUBTLEX-AR: Arabic word distributional characteristics based on movie subtitles. *Behavior Research Methods, 57*(1), 104. https://doi.org/10.3758/s13428-024-02560-8
Boudelaa, S., & Marslen-Wilson, W. D. (2010). Aralex: A lexical database for Modern Standard Arabic. *Behavior Research Methods, 42*(2), 481–487. https://doi.org/10.3758/BRM.42.2.481
Boudelaa, S., Norris, D., & Kinoshita, S. (2024). The differential effects of consonant and vowel diacritics in Arabic. *Journal of Memory and Language, 138*, Article 104533. https://doi.org/10.1016/j.jml.2024.104533
Boudelaa, S., Norris, D., Mahfoudhi, A., & Kinoshita, S. (2019). Transposed letter priming effects and allographic variation in Arabic: Insights from lexical decision and the same–different task. *Journal of Experimental Psychology: Human Perception and Performance, 45*(6), 729–757. https://doi.org/10.1037/xhp0000621
Boudelaa, S., Pulvermüller, F., Hauk, O., Shtyrov, Y., & Marslen-Wilson, W. (2010). Arabic morphology in the neural language system. *Journal of Cognitive Neuroscience, 22*(5), 998–1010. https://doi.org/10.1162/jocn.2009.21273
Briesemeister, B. B., Hofmann, M. J., Tamm, S., Kuchinke, L., Braun, M., & Jacobs, A. M. (2009). The pseudohomophone effect: Evidence for an orthography–phonology-conflict. *Neuroscience letters, 455*(2), 124–128. https://doi.org/10.1016/j.neulet.2009.03.010
Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition, 1*(1). https://doi.org/10.5334/joc.10. Article 9.
Dijkstra, T., Van Jaarsveld, H., & Ten Brinke, S. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition, 1*(1), 51–66. https://doi.org/10.1017/S1366728998000121
Forster, K. I. (2004). Category size effects revisited: Frequency and masked priming effects in semantic categorization. *Brain and Language, 90*(1–3), 276–286. https://doi.org/10.1016/S0093-934X(03)00440-1
Forster, K. I. (2006). Early activation of category information in visual word recognition: More on the turple effect. *The Mental Lexicon, 1*(1), 35–58. https://doi.org/10.1075/ml.1.1.05for
Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers, 35*(1), 116–124. https://doi.org/10.3758/BF03195503
Fox J, Weisberg S (2019). An R Companion to Applied Regression, Third edition. Sage, Thousand Oaks CA. https://www.john-fox.ca/Companion/.
Frisson, S. (2009). Semantic underspecification in language processing. *Language and Linguistics Compass, 3*(1), 111–127. https://doi.org/10.1111/j.1749-818X.2008.00104.x
Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences, 35*(5), 263–279. https://doi.org/10.1017/S0140525X11001841
Frost, R., Deutsch, A., Gilboa, O., et al. (2000). Morphological priming: Dissociation of phonological, semantic, and morphological factors. *Memory & Cognition, 28*, 1277–1288. https://doi.org/10.3758/BF03211828
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. https://doi.org/10.1037/0278-7393.23.4.829
Goh, W. D., Yap, M. J., Lau, M. C., Ng, M. M. R., & Tan, L.-C. (2016). Semantic richness effects in spoken word recognition: A lexical decision and semantic categorization megastudy. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.00976. Article 976.
Gottlob, L. R., Goldinger, S. D., Stone, G. O., & Van Orden, G. C. (1999). Reading homographs: Orthographic, phonologic, and semantic dynamics. *Journal of Experimental Psychology: Human Perception and Performance, 25*(2), 561–574. https://doi.org/10.1037/0096-1523.25.2.561
Gwilliams, L., & Marantz, A. (2015). Non-linear processing of a linear speech stream: The influence of morphological structure on the recognition of spoken Arabic words. *Brain and Language, 147*, 1–13. https://doi.org/10.1016/j.bandl.2015.04.006
Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance, 22*(6), 1331–1356. https://doi.org/10.1037/0096-1523.22.6.1331
Hino, Y., Lupker, S. J., & Pexman, P. M. (2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interactions between orthography, phonology, and semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 686–713. https://doi.org/10.1037/0278-7393.28.4.686
Hino, Y., Lupker, S. J., Sears, C. R., & Ogawa, T. (1998). The effects of polysemy for Japanese katakana words. *Reading and Writing: An Interdisciplinary Journal, 10*(3–5), 395–424. https://doi.org/10.1023/A:1008060924384
Hino, Y., Pexman, P. M., & Lupker, S. J. (2006). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interactions between orthography, phonology, and semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(1), 2–24. https://doi.org/10.1037/0278-7393.32.1.2
Jastrzembski, J. E. (1981). Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology, 13*(2), 278–305. https://doi.org/10.1016/0010-0285(81)90011-6
Prunet, J.-F., & Béland, R. (2000). Ali Idrissi; the Mental Representation of Semitic Words. *Linguistic Inquiry, 31*(4), 609–648. https://doi.org/10.1162/002438900554497
Kawamoto, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language, 32*, 474–516.
Kawamoto, A. H., Farrar, W. T., & Kello, C. (1994). When two meanings are better than one: Modeling the ambiguity advantage using a recurrent distributed network. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1233–1247.
Kellas, G., Ferraro, F. R., & Simpson, G. B. (1988). Lexical ambiguity and the time course of attentional allocation in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 14*(4), 601–609. https://doi.org/10.1037/0096-1523.14.4.601
Kinoshita, S., & Norris, D. (2012a). Task-dependent masked priming effects in visual word recognition. *Frontiers in Psychology, 3*, 178. https://doi.org/10.3389/fpsyg.2012.00178
Kinoshita, S., & Norris, D. (2012b). Pseudohomophone priming in lexical decision is not fragile in a sparse lexical neighborhood. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(3), 764–775. https://doi.org/10.1037/a0026782
Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous words. *Journal of Neurolinguistics, 20*(1), 1–24. https://doi.org/10.1016/j.jneuroling.2006.02.001
Klepousniotou, E., Titone, D., & Romero, C. (2008). Making sense of word senses: The comprehension of polysemy depends on sense overlap. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1534–1543. https://doi.org/10.1037/a0013012
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13
Lichacz, F. M., Herdman, C. M., Lefevre, J.-A., & Baird, B. (1999). Polysemy effects in word naming. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 53*(2), 189–193. https://doi.org/10.1037/h0087309
Locker, L., Jr., Simpson, G. B., & Yates, M. (2003). Semantic neighborhood effects on the recognition of ambiguous words. *Memory & Cognition, 31*(4), 505–515. https://doi.org/10.3758/BF03196092
Longtin, C.-M., & Meunier, F. (2005). Morphological decomposition in early visual word processing. *Journal of Memory and Language, 53*(1), 26–41. https://doi.org/10.1016/j.jml.2005.02.008
Martin, R. C. (1982). The Pseudohomophone effect: The Role of Visual Similarity in Non-Word Decisions. *The Quarterly Journal of Experimental Psychology Section A, 34*(3), 395–409. https://doi.org/10.1080/14640748208400851
McCarthy, J. J. (1981). A Prosodic Theory of Nonconcatenative Morphology. *Linguistic Inquiry, 12*(3), 373–418.
Meteyard, L., & Davies, R. A. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language, 112*, Article 104092. https://doi.org/10.1016/j.jml.2020.104092

Millis, M. L., & Button, S. B. (1989). The effect of polysemy on lexical decision time: Now you see it, now you don't. *Memory & Cognition, 17*(2), 141–147. https://doi.org/10.3758/BF03197064

Norman, T., Degani, T., & Peleg, O. (2017). Morphological processing during visual word recognition in Hebrew as a first and a second language. *Reading and Writing, 30*(6), 1259–1286. https://doi.org/10.1007/s11145-016-9663-7

Perea, M., Abu Mallouh, R., & Carreiras, M. (2010). The search for an input-coding scheme: Transposed-letter priming in Arabic. *Psychonomic Bulletin & Review, 17*(3), 375–380. https://doi.org/10.3758/PBR.17.3.375

Pexman, P. M., & Lupker, S. J. (1999). Ambiguity and visual word recognition: Can feedback explain both homophone and polysemy effects? *Canadian Journal of Experimental Psychology, 53*, 323–334.

Pexman, P. M., Lupker, S. J., & Hino, Y. (2002). The impact of feedback semantics in visual word recognition: Number of features effects in lexical decision and naming tasks. *Psychonomic Bulletin & Review, 9*, 542–549.

Prior, A., & Markus, E. (2014). Morphological activation in sentence context: When the root prevails over the meaning. *Language, Cognition and Neuroscience, 29*(3), 334–350. https://doi.org/10.1080/23273798.2014.920511

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org.

Rastle, K., & Coltheart, M. (2010). *Is there serial processing in the reading system; and are there local representations? In from Inkmarks to ideas* (pp. 33–54). Psychology Press.

Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes, 23*, 942–971.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language, 46*(2), 245–266. https://doi.org/10.1006/jmla.2001.2810

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science, 28*(1), 89–104. https://doi.org/10.1207/s15516709cog2801_4

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 9*(5), 487–494. https://doi.org/10.1016/S0022-5371(70)80091-3

Saiegh-Haddad, E. (2003). Linguistic distance and initial reading acquisition: The case of Arabic diglossia. *Applied Psycholinguistics, 24*(3), 431–451.

Saiegh-Haddad, E., & Henkin-Roitfarb, R. (2014). The structure of Arabic language and orthography. In E. Saiegh-Haddad, & M. Joshi (Eds.), *Handbook of Arabic literacy: Insights and perspectives* (pp. 3–28). Springer.

Seghier, M. L., & Boudelaa, S. (2024). Constraining current neuroanatomical models of reading: The view from Arabic. *Brain Structure & Function, 229*, 2167–2185. https://doi.org/10.1007/s00429-024-02827-y

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 57A*(4), 745–765. https://doi.org/10.1080/02724980343000477

Taha, H., & Khateb, A. (2013). Resolving the orthographic ambiguity during visual word recognition in Arabic: An event-related potential investigation. *Frontiers in Human Neuroscience, 7*. https://doi.org/10.3389/fnhum.2013.00821. Article 821.

Vanlangendonck, F., Peeters, D., Rueschemeyer, S. A., & Dijkstra, T. (2020). Mixing the stimulus list in bilingual lexical decision turns cognate facilitation effects into mirrored inhibition effects. *Bilingualism: Language and Cognition, 23*(4), 836–844. https://doi.org/10.1017/S1366728919000531

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*, 971–979. https://doi.org/10.3758/PBR.15.5.971