

Similarity-induced interference or facilitation in language production reflects representation, not selection

Gary M. Oppenheim^{1,2}, Nazbanou Nozari^{3,4}

¹ Department of Psychology, Bangor University, Bangor, Wales, UK

² Department of Psychology, The University of Texas at Austin, USA

³ Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

⁴ Cognitive Science Program, Indiana University, Bloomington, IN, USA

Manuscript accepted for publication in *Cognition* on 9 January, 2024

Abstract

Researchers have long interpreted the presence or absence of semantic interference in picture naming latencies as confirming or refuting theoretical claims regarding competitive lexical selection. But inconsistent empirical results challenge any mechanistic interpretation. A behavioral experiment first verified an apparent boundary condition in a blocked picture naming task: when orthogonally manipulating association type, taxonomic associations consistently elicit interference, while thematic associations do not. A plausible representational difference is that thematic feature activations depend more on supporting contexts. Simulations show that context-sensitivity emerges from the distributional statistics that are often used to measure thematic associations: residual semantic activation facilitates the retrieval of words that share semantic features, counteracting learning-based interference, and training a production model with greater sequential cooccurrence for thematically related words causes it to acquire stronger residual activation for thematic features. Modulating residual activation, either directly or through training, allows the model to capture gradient values of interference and facilitation, and in every simulation competitive and noncompetitive selection algorithms produce qualitatively equivalent results.

Introduction

The phonological form of a word is largely independent of its semantic content, so mapping from shared semantic representations to phonology in language production requires some form of winner-take-all process to select a single word. That much is agreed. However, a major debate spanning the past two decades has focused on the particular functional characteristics of that selection process and its association with the timecourse of word production. Chronometric studies of word production assume that more difficult lexical selection should take more time and therefore use naming latencies from tasks like primed picture naming to identify factors that increase or decrease selection difficulty. The simplest interpretative framework assumes that the naming latencies for a stimulus directly index the activation of the strongest candidate name (i.e., the ‘target’). For instance, each word might accumulate activation at its own rate, and speakers simply select the first word whose activation exceeds an absolute threshold. Therefore, if speakers were quicker to name a picture of a dog in one instance than in another, the difference would be assumed to reflect a modulation in the underlying activation of the name ‘dog’. A more complex, but also more popular, interpretive framework instead assumes that naming latencies index the activation of the strongest candidate name *relative to the activations of one or more selection alternatives* (i.e., ‘competitors’).¹ Under this account, speakers might require the activation of one word to exceed that of other candidates by a certain critical amount: a relative threshold. Therefore, if speakers were slower to name a picture of a dog in one instance than in another, the difference would be assumed to reflect a modulation in the activation of competitors for ‘dog’ (e.g., ‘cat’). Crucially, the empirical support for these contrasting accounts reflects contrasting—and apparently contradictory—effects of semantic context manipulations, with semantic facilitation claimed to support the former (“non-competitive”) account, and semantic interference claimed to support the latter (“competitive”) account (e.g., Abdel Rahman & Melinger, 2009; Mahon et al., 2007). In this paper, we will argue that the association of these contrasting behavioral effects with contrasting selection mechanisms is mistaken, and the contrasting effects are better understood as reflecting differences in the activation of representations, placing little if any constraint on plausible selection mechanisms.

Behavioral effects of semantic similarity on word production

The coactivation of multiple taxonomically related words is, as a general rule, associated with behavioral interference. Perhaps the best-known example of this phenomenon comes from the picture-word interference paradigm, in which researchers present to-be-ignored written or auditory ‘distractor’ word at the same time as a to-be-named picture (e.g., Schriefers et al., 1990). The distractor presentation elicits a general increase in error rates and slowing of correct productions, even when the target and distractor are wholly unrelated (e.g., target: dog, distractor: chair), and these effects are exaggerated when the target and distractor are exemplars of the same taxonomic category (e.g., target: dog, distractor: wolf). This semantic picture-word interference effect has proven empirically robust over several decades of research (see Bürki et al., 2020, for a recent meta-analysis, suggesting an effect size of approximately 21 ms), and the task has allowed

¹ Or that target activations directly encode this difference, for instance via lateral inhibition.

researchers to control and manipulate target-distractor relations with enviable precision. But the task's complexity, combining simultaneous presentations of multiple stimuli in multiple modalities with strong ad hoc control demands, has proven a perennial liability: although its semantic interference effects are usually interpreted as emerging from competition between potential responses during the lexical selection process (e.g., Abdel Rahman & Melinger, 2009; Roelofs, 1992), after four decades and hundreds of published experiments, controversy remains about what, if anything, results from the task reveal about typical communicative language production (Mahon et al., 2007; Oppenheim & Balatsou, 2019; Starreveld & La Heij, 1995).

It is therefore useful to note that other production tasks—such as those that we will primarily focus on in this paper—provide converging evidence for this general empirical trend. Similar interference has been shown in picture naming tasks without concurrently presented distractors, when naming is preceded at some lag by the production of a semantically related prime word. Producing a related prime has been shown to decrease speed and accuracy of picture naming several seconds later, whether the prime is elicited via picture naming (Vitkovitch & Humphreys, 1991), naming-to-definition (Wheeldon & Monsell, 1994), or word naming (Vitkovitch et al., 2019), though it has been suggested that the last of these has a different source. In the so-called continuous naming paradigm, semantic interference incrementally grows as speakers name pictures from the same semantic category, interspersed with pictures from other categories (e.g., **bear**, chair, bottle, **wolf**, apple, trombone, waterfall, **dog**), producing an ordinal position effect (e.g., Howard et al., 2006; Schnur, 2014). In that example, dog would typically be about 10-20 ms slower when it occurs as the third mammal in the sequence than when it occurs as the second. Notably, this effect appears to increase with each instance rather than each unique exemplar (e.g., Navarrete et al., 2010; Oppenheim, 2018), linking it to interference in another paradigm, called blocked cyclic naming. In blocked cyclic naming, speakers repeatedly cycle through naming pictures in small sets (or 'blocks') that are either semantically related (e.g., bear, wolf, dog) or unrelated (e.g., chair, apple, dog). Naming pictures in single-category blocks is typically slower and more error prone, compared to unrelated blocks (e.g., Damian et al., 2001; Schnur et al., 2006), and the effects usually, but not invariably, emerge with repetition (for exceptions, see e.g., Belke, 2008; Belke et al., 2005; Damian & Als, 2005). Finally, while the priming manipulations discussed thus far are thought to externally induce semantic interference, picture name agreement effects are often claimed as evidence that endogenous co-activation produces the same competitive effects: potential names inhibit each other, so speakers are slower to name pictures that more frequently elicit alternative names (e.g., Alario et al., 2004; Bates et al., 2003; but see Oppenheim, 2024, for evidence against that interpretation). Thus, a wide variety of tasks appear to link lexical coactivation to behavioral interference, particularly in the case of taxonomic category coordinates.

But semantic similarity does not always lead to interference. Semantically related distractors have sometimes been linked to faster naming latencies in picture-word interference (e.g., Abel et al., 2009; Alario et al., 2000; Costa et al., 2005; de Zubicaray et al., 2013; Heij et al., 1990; Mahon et al., 2007), and semantic relations in blocked cyclic naming have sometimes failed to elicit the expected strong interference effects, even trending toward facilitation (e.g., de Zubicaray et al., 2014). It would be tempting to attribute these findings to false negatives or procedural aberrations. Methodological discussions over the past decade have raised awareness of sampling error in

inferential statistics (e.g., Cumming, 2014); some variation in magnitude and even direction is to be expected for any true effect. And decades of research have considered many seemingly trivial details that might underlie picture-word interference effects or undermine their common theoretical interpretations (e.g., SOA, ITI, distractor modality, response set membership, set size, distractor frequency, pre-experiment familiarization, mediated priming; see Bürki et al., 2020, for a recent review). But these studies share an important attribute that may explain their divergent results: their semantic manipulations are not based on structural taxonomic relations, but rather thematic associations (see Mahon et al., 2007, and references therein).

Although research participants can readily provide subjective judgements of concepts' semantic relatedness, defining general theories of such relations and optimal objective measures to quantify them remains the subject of active debate (e.g., Golonka & Estes, 2009; Recchia & Jones, 2009; Tenenbaum & Griffiths, 2001; Tversky, 1977). For present purposes, it will suffice to point out that semantic relatedness can be assessed along at least two major dimensions: (1) taxonomic, and (2) thematic. Taxonomic relatedness describes the extent to which two (or more) words could be used to describe the same referent, and can thus be readily represented in terms of shared structural semantic features (e.g., 'dog' and 'bear' would share structural features such as 'animate', 'furry', and 'quadrupedal'). Experimental manipulations of semantic priming have predominantly focused on this dimension, usually combining basic-level exemplars of a single taxonomic category (e.g., rodents: mouse, rabbit; fruits: apple, pear; buildings: house, church). Across paradigms there is strong and consistent evidence that such taxonomy-based manipulations of semantic relatedness elicit behavioral interference (see Nozari & Pinet, 2020, for a recent review): although taxonomic effects appear to be nonmonotonic in the picture-word interference paradigm (with synonyms, near-synonyms, superordinates, and translation equivalents eliciting weaker interference or even facilitation, e.g., Costa et al., 1999; Dylman & Barry, 2018; Mahon et al., 2007; Peterson & Savoy, 1998; Vitkovitch & Tyrrell, 1999; but cf. Rose et al., 2019), there is thus far no converging evidence for such inconsistency in either the continuous naming (Alario & Moscoso Del Prado Martín, 2010) or blocked cyclic naming paradigm (Vigliocco et al., 2002).

The thematic relatedness dimension describes a non-structural form of habitual episodic association between concepts, based on their simultaneous appearance or use in the same contexts or 'themes'. It may form the basis of identifiable theme-based or ad hoc categories (e.g., pets), for instance explaining why 'dog' might be considered more closely related to 'cat' (a distantly related domesticated mammal) than to 'wolf' (a closely related undomesticated mammal with which 'dog' shares more superficial features), and although it can emerge alongside taxonomic relatedness (e.g., doctor/nurse, bread/butter), it does not require it (e.g., cake/candle, cow/milk). Because thematic associations are closely related to cooccurrence, both in the real world and in language², it may not be surprising that, in contrast to the robust interference from taxonomic relatedness manipulations, thematic relatedness manipulations in the picture-word interference task generally elicit facilitation (Abel et al., 2009; Alario et al., 2000; Costa et al., 2005; de Zubicaray et al., 2013; La Heij et al., 1990), attenuating taxonomy-based interference in cases of overlap (e.g., dog/cat;

² In fact, recent analysis techniques in the field of distributional semantics have shown that the similarity structure in the real world can be captured by the structure of language (Mikolov et al., 2013).

La Heij et al., 1990; but cf. Lupker, 1979). We could in fact imagine thematic ‘features’ as representations that encode and predict cooccurrence in comprehension and promote cooccurrence in production, for instance by characterizing larger units of discourse than taxonomic features or tracking transitional probabilities.

One might expect thematic manipulations to elicit similar facilitation in blocked cyclic naming but the evidence for facilitation in that paradigm is suggestive, at best. While one group of researchers has reported interference from thematically related sets in multiple experiments (e.g., Frenchman, beret, eclair: Abdel Rahman & Melinger, 2007, 2011; see also Rose & Abdel Rahman, 2016, for a report of such interference in continuous naming), others have interpreted null effects as replication failure (de Zubicaray et al., 2014; see also Aristei et al., 2011). So compared to taxonomic interference, the reports of thematic effects are fewer and less consistent. One possible explanation for the inconsistency is that thematic manipulations in blocked cyclic naming have not been sufficiently controlled to exclude other sources of interference. This seems plausible because blocked cyclic naming involves larger item sets (> 2 pictures) than picture-word interference (1 picture, 1 word), thus creating more relations that would need to be controlled, and studies reporting interference have not used rigorous measures to control them. There is also a statistical concern: reports of thematic interference in blocked-cyclic naming have focused on the presence or absence of the effect, rather than its equivalence to the taxonomy-based effect, but in all studies cited above the thematic interference effect is numerically weaker than the taxonomy-based effect, which would be consistent with the idea that thematic facilitation attenuates taxonomic interference.

In sum, feature-based taxonomic manipulations consistently elicit semantic interference effects in several different production tasks, but this is not true for all kinds of semantic relations. For theme-based semantic manipulations, the evidence is less clear: while picture-word-interference studies have consistently reported facilitation, the few studies using other paradigms have often reported interference. And though the thematic effects in blocked naming studies appear to be weaker and/or more variable than those of taxonomic similarity, which would be more in line with the picture-word interference results, they have not formally controlled or manipulated semantic relations in a way that would allow distinguishing between similarity types for a meaningful evaluation or comparison. Thus, both the empirical claim of a thematic interference effect, independent of taxonomic interference, and its comparison to the taxonomic effect require verification.

Theoretical interpretations attached to interference and facilitation effects

What do semantic interference and facilitation effects reveal about the process of word production? As noted above, theoretical accounts broadly describe word production as a mapping from a conceptual/semantic space to a phonological/articulatory space via the selection of single word. The conceptual/semantic space happily accommodates ambiguity; it can allow the co-activation of an array of ‘semantic features’ (Dell, 1986) or a host of ‘lexical concepts’ (Levelt et al., 1999) with no inherent need to constrain or modulate that activation. Activation spreads freely between semantic and lexical representations. But mapping that message onto a linguistic form requires

committing to a single option: choosing between ‘dog’ and ‘wolf’ or even between ‘couch’ and ‘sofa’. Semantic interference and facilitation effects in primed picture naming are usually claimed to reveal the functional characteristics of this decision process.

Semantic interference, regardless of the task or specific manipulation, has been strongly characterized as evidence that lexical selection is ‘competitive’, in the sense that the process of choosing a word requires more time in the presence of co-activated alternatives. Perhaps the clearest example of the principle is Howard and colleagues (2006) adaptation of McClelland & Rumelhart’s (1981) interactive activation model: each candidate word had direct (i.e., lateral) inhibitory connections to every other word, so any co-activated competitor would necessarily inhibit the target, slowing its activation toward a selection threshold. Thus competition was integral to not only the timing of the decision process but also its outcome: though it did not explicitly address the possibility of selection errors, the model contained no mechanism to identify a target prior to its reaching threshold. The complexity of Roelofs’ (1992 *et passim*) WEAVER model makes the role of competitive selection less clear, and its underlying assumptions are quite controversial (e.g., Caramazza & Costa, 2000, 2001; Starreveld & La Heij, 1995, 1996), but it is nonetheless the most prominent example because (1) it describes competitive selection as a core theoretical claim, and (2) it forms the computationally explicit basis of the standard model of word production (Levelt et al., 1999). WEAVER was originally developed to explain picture-word interference effects, such as why taxonomy-based semantic manipulations robustly increased naming latencies without strongly affecting naming accuracy, and to do that it broke the lexical selection process into two stages that we can describe as (1) target identification and (2) competitor suppression. In the first stage, speakers would cross-reference an array of picture-relevant words with an array of task-licensed words to identify and boost the activation of the target ‘lexical concept’ (a nominally semantic representation of the word) that they intended to produce. This stage was described as happening quickly and accurately, via a noncompetitive spreading activation process, and could itself constitute a complete lexical selection process. But WEAVER additionally assumed that speakers must then select the target lexical concept’s corresponding ‘lemma’ – a nominally syntactic representation of the word that was closely integrated with the nominally semantic ‘lexical concept’ that they had just selected – effortfully suppressing all other lemmas. Selection times were assumed to predominantly reflect this second stage: though the production system had already identified its target lemma and could not fail to select it, it would not do so until some time after its activation exceeded the total activation of other lemmas by some amount (i.e., a relative threshold). Within that framework, distractors in the picture-word interference task were assumed to have a task-specific back-door to lemma activation, creating strong task-unlicensed lemmas that must be quashed (Roelofs, 1992), while tasks like blocked-cyclic naming create task-specific residual conceptual activation (via an additional ‘bias’ parameter) that then flows into quashworthy lemmas (Roelofs, 2018; see also Belke, 2013; Belke et al., 2005; Damian et al., 2001), leading to claim that WEAVER provides a parsimonious account of both sets of phenomena (Roelofs, 2018). Thus, while WEAVER introduces considerably more degrees of freedom than a simple lateral inhibition model, it retains the basic principle that coactivation should somehow cause selection delays.

Critics argue that that principle prevents competitive selection models from offering an internally consistent account of semantic facilitation effects (e.g., Mahon & Navarrete, 2014; in press). The simpler lateral inhibition model certainly offers no obvious way to explain semantic facilitation, and while the spreading activation in WEAVER's first stage could plausibly allow co-active competitors to increase target activation (e.g., in a manner similar to the Dell, 1986, model that we describe below), its subsequent suppression stage would seem obliged to convert any co-activation into net interference; indeed that conversion is the basis of its account of cumulative semantic interference (Belke, 2013; Roelofs, 2018). But WEAVER has eluded such straightforward tests, accumulating a list of relations, tasks, and conditions that a cognitively penetrable suppression process might exempt or otherwise treat specially in some way, and adding architectural and algorithmic assumptions and parameter changes each time another is identified. So far, this list includes synonyms, near-synonyms, translation equivalents, part-whole relations, and the main focus of the current paper, thematic relations, as well as the broad stipulation that lemma selection in a picture-word interference experiment is "more complicated" than the "very simple process" in spontaneous speech (Roelofs, 1992, p117). It is *a priori* unclear why the internals of the lexical selection process should be reconfigurable to address arbitrary task demands, why certain relations should be treated specially in some tasks and not others, or why post-semantic efforts to select the lemma 'sofa' should be concerned with the activation of 'chair' but not 'couch' or 'cushion', but beyond the plausibility of the individual revisions such complexities weaken not only WEAVER's claim of parsimony but more importantly the logical link between its ability to fit empirical data and the validity of its individual assumptions, such as competitive selection. The 'swinging lexical network' account (Abdel Rahman & Melinger, 2009 et passim) claims to address the issue of facilitation more gracefully, by assuming that the number of coactivated alternatives matters more than their strength in determining competitive interference effects, but it has not been computationally implemented.

Due to the difficulty of capturing semantic facilitation effects in competitive selection models, they have often been used to argue that lexical selection is, instead, 'noncompetitive': speakers may co-activate multiple candidate words, but doing so does not slow the selection process. Though Dell's (1986 et passim) interactive spreading activation model is often framed as noncompetitive – which is true in the sense that it omitted the lateral inhibition, relative threshold, and non-target suppression features that link coactivation to delayed responses in competitive selection models – it would be more accurate to say that response time effects were simply beyond its intended scope. It was instead developed to explain error patterns as a result of automatic spreading activation and 'feedback': noisy activation cascaded forward from distributed semantic features to words and on to their constituent phonemes, and then flowed back up to connected words and features. This feedback primarily accomplished a pattern-completion function (cf. Hopfield, 1982), reinforcing the activation of an intended wordform and supporting accurate production overall. But because the model only contained excitatory connections, feedback also increased the activation of alternatives that shared semantic and/or phonological elements with a target, thus explaining important features of error distributions as a side effect of processes that normally support error-free productions. It is conceivable that the same feedback function could also provide a route for semantic manipulations to elicit response time facilitation or attenuate interference effects, but the timing of selection in this model was largely an afterthought: at some

point, the most activated word was selected and received an extra jolt of activation, and in the original implementations its timing was entirely independent of both target and non-target activations.

How might such a model account for the response time interference effects in any of the above-mentioned tasks? Clearly it would require additional mechanisms. One possibility is that it could incorporate competitor inhibition directly into its activation and selection processes (e.g., Schade & Berg, 1992), in which case it would no longer represent a non-competitive account. Another possibility is to invoke ad hoc control processes in response to ad hoc manipulations (e.g., Dell & O'Seaghdha, 1992; Mahon et al., 2007; Oppenheim et al., under review), shifting the locus of response time interference from lexical selection to a later post-selection monitoring process. In fact, this option is not so far off from WEAVER's distinction between an early target identification process and a subsequent non-target suppression process (see also Dhooge & Hartsuiker, 2012); the main functional difference emerges from WEAVER's separate assumption of serial discrete processing, which postpones sublexical processing until all non-exempt non-targets have been thoroughly subdued.

Incremental learning models (e.g., Howard et al., 2006; Oppenheim et al., 2010) offer a more promising alternative to the standard residual-activation-with-competitive-selection account of semantic interference effects, by shifting the origin of these effects outside the current trial. This approach assumes an implicit learning framework in which production creates persistent priming (cf. Chang et al., 2006; Dell et al., 2000; Dell & Kittredge, 2013): speakers develop production skills through lifelong practice and each word retrieval experience modifies the connections that map from semantic representations to words. Such a mechanism is critical for explaining long-lasting changes in semantic-to-lexical mappings. For example, Balatsou, Fischer-Baum, and Oppenheim (2022) demonstrated long-lasting idiosyncratic increases in the probability of choosing certain words over common alternatives: in a study where university students named each of the 520 line drawings from the International Picture Naming Project (Bates et al., 2003) once a week for three weeks, they became increasingly likely to reuse the specific picture names that they personally had produced a week earlier (e.g., choosing 'turtle' again over the picture's dominant name, 'tortoise'). And in a recent neuropsychological study, Hepner and Nozari (2020) showed an increase in the probability of producing a semantic error in three individuals with post-stroke aphasia, as a function of a semantically related item that appeared 35-40 trials earlier (see also Hsiao, Schwartz, Schnur, & Dell, 2009, and Oppenheim et al.'s 2010 Simulation 4 thereof). Incremental learning models initiate word retrieval via the same semantic-to-lexical mapping process that characterizes the models above, and assume that the timing of lexical selection depends on lexical activation, modelled as accumulation toward a selection threshold. But crucially, after each retrieval, an implicit incremental learning process reinforces that mapping: successfully naming a picture of a dog as 'dog' strengthens the connections from semantic features like [furry] to 'dog', making it faster and more likely to retrieve dog again in the future. Combining such repetition priming with a competitive selection process would be one way to account for long-lasting cumulative semantic interference effects, and thus Howard et al. (2006) added persistent increases in the weights of semantic-to-lexical connections to the lateral inhibition model that we already described above: voila, long-lasting competitor priming produced long-lasting

interference. But models of incremental learning in most domains involve weakening less desirable connections as well as strengthening more desirable ones. This ‘dark side’ of incremental learning was integral to the models of Widrow and Hoff (1960) and Rescorla and Wagner (1972), which formed the basis of the ‘delta rule’ and thus backpropagation and connectionism in general (e.g., Rumelhart & McClelland, 1986), so it is not a stretch to assume that incremental learning in the domain of word production also involves weakening the connections that support coactivated alternatives (e.g., [furry]→’wolf’). Thus, Oppenheim and colleagues (2010) paired incremental learning with incremental unlearning in their 2010 ‘Dark Side’ model. Rather than positing task-specific strategies and behaviors (cf. Roelofs, 1992; 2018), it implemented a minimal working model of part of the core word production system, producing words through its normal processes (activating words via shared semantic features and then continually ‘boosting’ them until the first reached a selection threshold) and merely continuing the same error-based learning process by which it had initially acquired its semantic-to-lexical mappings. Though it offered no account of picture-word interference effects, the simulations demonstrated that unlearning in the Dark Side model was sufficient to account for all then-known manifestations of cumulative semantic interference, including both error and RT effects from speakers both typical and impaired, without a competitive selection or competitor suppression process, by directly rendering former competitors less accessible.

However, it is important to note that incremental learning models have been specifically proposed to account for interference from semantic manipulations that most clearly map onto the concept of ‘taxonomic relatedness’, with little or no consideration of thematic manipulations. The simplest solution would be to assume that speakers represent and access theme-based categories (e.g., French things) in much the same way that they represent and access taxonomic categories (e.g., mammals), merely representing themes as additional sources of semantic activation (as suggested by Roelofs, 2018). However, any such approach would need to address the fact that studies of theme-based semantic manipulations have typically elicited less interference than taxonomy-based manipulations, sometimes even eliciting facilitation.

To re-cap, several decades of language production research have consistently linked behavioral observations of interference and facilitation effects to theoretical accounts of lexical selection mechanisms. Though recent work has questioned the necessity of competitive selection for explaining taxonomic interference effects, thematic effects and their theoretical implications remain underexplored.

Current study

The main goal of the current paper is to test the assumption that behavioral interference and facilitation map on to competitive and non-competitive selection mechanisms, respectively. To this end, we choose taxonomic vs. thematic similarity as a test case, since many studies have reported contrasting effects for these two types of similarity, with taxonomic effects consistently on the interference side and thematic effects often on the facilitatory side. Note that our goal is not to propose a comprehensive theory of taxonomic vs. thematic similarity, nor is it to adjudicate between different theories of semantic processing (e.g., see Lambon-Ralph et al., 2017, for a

review). Rather, we are primarily concerned with the distinction between these two types of semantic relations in so far as they elicit opposing effects in production. To this end, we capitalize on a key difference between the two types of similarity that is inherent to the definition of these two types of relationship and, as such, independent of the assumptions of specific theory of semantic processing: whereas taxonomic features can refer to individual entities, themes necessarily integrate multiple entities and thus their relevance to retrieval processes depends more on the cooccurrence of multiple theme-relevant entities. We then test whether this simple and self-evident difference explains the contrasting pattern of behavioral results, or whether the key determinant of the contrasting pattern is the selection mechanism.

Given the variability in the reports of thematic effects and the problems enumerated in the empirical designs of past studies, we begin by behaviorally comparing the magnitude of thematic vs. taxonomic similarity within a more rigorously controlled design. The behavioral study uses a simple paradigm with only two pictures per block, named randomly and repeatedly, to directly compare the influence of taxonomic and thematic context on word production. The task has a number of properties that make it useful for this purpose: first, it has been shown to capture the well-known effects of semantic interference, onset facilitation, and coda interference in production (Harrison et al., 2020; Nozari et al., 2016). Second, it is simple and does not require unnatural operations such as suppressing a written word-form while speaking. Third, with only two items, it is easy to statistically model the effect of switching and repeating (e.g., Braver, 2012). In our task, switch trials are almost twice as common as repeat trials. Given the reports that participants are sensitive to the probability of switches vs. repeats (Monsell & Mizon, 2006; Schneider & Logan, 2006), it is possible that they strategically prepare for a switch, and such strategic processes could, in theory, modulate the basic facilitation/interference effects across conditions. Modeling the interaction between the switch/repeat variable and condition allows us to rule out the influence of strategic processes on our effects of interest. Finally, and most importantly, the reduced set size allows stronger control of pairwise relationships than typically characterizes such interference studies. We used well-established measures from a pre-existing taxonomy (i.e., WordNet) to formally quantify taxonomic relations, and used lexical cooccurrence as a more easily quantifiable proxy for thematic relations³, creating an objective double-dissociation between taxonomically and thematically related pairs that allowed us to compare each to an unrelated baseline condition. To anticipate our results, this experiment demonstrates that thematic relations produce no trace of interference. Rather, the thematic effect is numerically on the facilitation side and significantly differs from the interference effect produced by taxonomic relations.

Having verified the distinction between thematic and taxonomic effects, we turn to the theoretical question of whether they imply distinct selection mechanisms. The computational models that are normally cited as engaging competitive or noncompetitive lexical selection algorithms actually vary along many dimensions, obfuscating that critical distinction, so we test both algorithms within a single model. Oppenheim and colleagues' (2010) incremental learning model is useful in this regard: it is a very simple model of lexical activation and selection, with clearly described

³ The rationale for this choice was the close correspondence between cooccurrence in the world and in language, which even allows for the extraction of relationships in the world from the structure of language e.g., Mikolov et al., 2013).

processes, and it can incorporate competitive or noncompetitive selection rules simply by replacing a relative threshold function with an absolute threshold function (ibid, Simulation 6). Using this model as our base, we present two groups of simulations. The first group (Simulations 1-3) assesses whether a simple hard-wired distinction between taxonomic and thematic relations can readily capture the complex pattern of empirical findings irrespective of the selection rule. Simulation 1 replicates the findings of Oppenheim et al. (2010) in capturing taxonomic similarity, and shows that implementing thematic similarity in exactly the same way would produce similarly strong interference for either selection rule. Simulation 2 then compares the results of each selection rule after introducing our assumption that thematic activation is more context-dependent than taxonomic activation; we implement that assumption by carrying a small amount of residual thematic activation into the next trial, so a thematic feature contributes more when it is used repeatedly across multiple trials. Simulations 3a and 3b extend this approach by showing how the extent of residual thematic activation affects the model's predictions.

To anticipate our results, the first group of simulations shows that the basic assumption of context-dependency is sufficient to capture the contrasting pattern of results reported in the literature without invoking contrasting selection mechanisms. While such a distinction seems quite plausible, one could ask how the language production system acquires it. The second group of simulations (Simulations 4-6) addresses this question. Our proposal is simple: speakers learn to distinguish between taxonomic and thematic features from the statistics of typical language use, in which thematically related items are more likely to cooccur. To demonstrate the viability of a distributionally acquired distinction, we extend our base model to incorporate an SRN-inspired (simple recurrent network; Elman, 1990) 'context layer' that represents a time-delayed copy of the input semantic activations, thus allowing context dependency to emerge from transitional probabilities. Simulations 4-6 then show that the extended model can closely approximate the results of Simulations 1-3, respectively, simply by modulating the lexical cooccurrence statistics in the model's training regimen while holding all model parameters constant. Thus the distinction that we hard-coded in Simulations 1-3 is learnable from the same distributional regularities that we used to characterize thematic relations in our behavioral experiment.

Experiment 1

Methods

Participants

The effect size for semantic interference in a similar paradigm in Nozari et al. (2016) was 0.4. With $\alpha = 0.05$, the estimated sample size to detect a significant effect with a power of 0.8 is 52. We thus recruited sixty native English-speaking undergraduate students ($M_{age} = 19.33$, $SD = 1.63$; 38 female) from Carnegie Mellon University, who participated for credit. Consent was obtained under a protocol approved by the Carnegie Mellon Institutional Review Board.

Materials

Three conditions (taxonomic, thematic, and unrelated) were created, each containing six pairs of monosyllabic words that were related taxonomically (e.g., *cow/bear*) or thematically (e.g., *cow/milk*), or were unrelated (e.g., *cow/pen*). Target words, i.e., those on which statistical analyses

were performed, were the same across the three conditions, while the competitor words, i.e., the other word in the pair, were different in each set, creating a set of 24 unique words (see Table 1). Competitors were matched for lexical frequency from SubtlexUS corpus (Brysbaert & New, 2009; Taxonomic: $M = 1.459$, $SD = 0.289$; Thematic: $M = 1.559$, $SD = 0.434$, Unrelated: $M = 1.466$, $SD = 0.602$), age of acquisition from Kuperman et al., (2012; Taxonomic: $M = 4.172$, $SD = 0.587$; Thematic: $M = 4.057$, $SD = 0.707$, Unrelated: $M = 4.107$, $SD = 0.625$), and word length (Taxonomic: $M = 3.333$, $SD = 0.516$; Thematic: $M = 3.5$, $SD = 0.547$, Unrelated: $M = 3.333$, $SD = 0.516$). Finally, phonological overlap, calculated as the total number of phonemes shared by a pair divided by the total number of phonemes in the two words (Goldrick et al., 2010), was minimized and matched across conditions (Taxonomic: $M = 0.112$, $SD = 0.172$; Thematic: $M = 0.167$, $SD = 0.183$, Unrelated: $M = 0.112$, $SD = 0.172$) to avoid segmental blocking effects (e.g., Harrison, Hepner, & Nozari, 2020).

To index taxonomic relatedness, WordNet-based Resnik (1995) similarity scores were calculated, using the Natural Language Toolkit for Python (NLTK), as $\text{Sim}_{\text{Resnik}}(x, y) = \text{IC}(\text{LCS}(x, y))$, where $\text{LCS}(x, y)$ refers to the lowest common subsumer (i.e., lowest node in the hierarchy that is a hypernym) of x and y , and $\text{IC}(z)$ refers to information content of the LCS in the SUBTLEX-US corpus (Brysbaert & New, 2009), calculated as $-\log P(z)$. To index thematic relatedness, which is based on conceptual cooccurrence, we used the closely associated phenomenon of lexical cooccurrence, using NLTK to calculate bidirectional pairwise mutual information (PMI) and log-likelihood (LL) scores (Evert, 2008) for every word pair in the SUBTLEX-US corpus (Brysbaert & New, 2009), with a window of 5 words (and replicated with a window of 10 for robustness). For any two words (x and y), PMI was calculated as $\text{PMI}(x,y) = \log_2 P(x, y) / P(x) P(y)$ while LL was calculated as $\text{LL} = 2 * \sum (O_i * \ln(O_i / E_i))$, where O_i is the observed frequency of word i and E_i its expected frequency. PMI and LL scores have opposite biases with regard to baseline frequency of words in a corpus (Evert, 2008). Calculating both ensures that such biases are not affecting our attempt to create pairs with true double dissociations. Taxonomically related pairs had high Resnik scores ($M = 5.689$, $SD = 1.205$), and low association scores (PMI: $M = 1.257$, $SD = 1.068$; LL: $M = 0.27$, $SD = 0.419$). Conversely, thematically related pairs had high association scores (PMI: $M = 5.689$, $SD = 2.919$; LL: $M = 123.234$, $SD = 128.363$), and low Resnik scores ($M = 1.179$, $SD = 0.742$; all p values < 0.001). Unrelated pairs had low scores on both sets of measures (Resnik: $M = 1.618$, $SD = 0.738$; PMI: $M = 0.151$, $SD = 0.576$; LL: $M = 2.287$, $SD = 4.14$).

Table 1. The six sets used in the experiment. Each row shows the target words, along with the three competitors. Tax = taxonomic, Them = thematic, U = unrelated.

Set	Target	Tax	Them	U
1	Cow	Bear	Milk	Pen
2	Car	Bike	Road	Doll
3	Hand	Tongue	Glove	Chair
4	Sock	Hat	Foot	Kite
5	Bird	Sheep	Nest	Plate
6	Pear	Corn	Tree	Boat

Twenty-four 400 x 400-pixel, color images corresponding to the 24 words were selected from Google Images, taking care to minimize visual similarity (e.g., for animal pairs, postures were selected that did not always reveal four legs, etc.). For cases where some similarity in the

taxonomic condition was unavoidable due to category membership (*cow/bear*), we included a pair that was more visually similar in the thematic condition (*hand/glove*). The materials are available in Appendix A. Eighteen blocks (6 of each condition) were created, each containing one pair of pictures. Three pseudo-randomized orders were generated, such that the same item was not repeated more than three times in a row within a block and the same target never appeared in adjacent blocks. The orders were counterbalanced across conditions, creating three different lists.

Procedure

The experiment was run in MATLAB R2019a with the Psychtoolbox-3 package (Brainard, 1997). Pictures were displayed at the center of a 20 x 13inch Dell desktop approximately 25 inches in front of the participants. Participants were assigned to one of three lists (see above). Figure 1a shows the structure of a block. On each block, participants first saw both pictures and read their written labels aloud. Next, they completed four practice trials. On each trial, one of the two pictures was presented for 1500 ms (or until a response was recorded by the voice key) and participants were instructed to name the picture as quickly and accurately as possible. The next trial started after 500 ms of a blank screen. After practice, participants completed 18 blocks of 12 experimental trials (6 of each picture), with the same structure as practice trials. Response times (RTs) were registered using an Audio-Technica microphone, and all spoken responses were recorded.

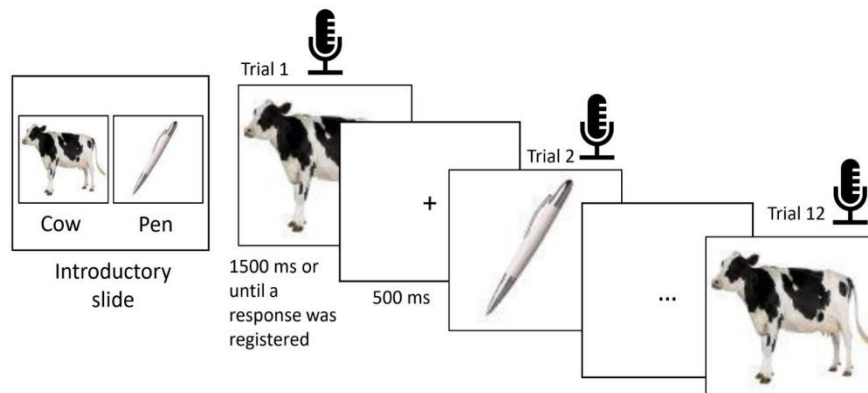


Figure 1. The block structure and a series of example trials in the production experiment. Participants see the picture pair and their labels in an introductory slide. After four practice trials (not shown), they name each of the two pictures six times in pseudorandomized order.

Coding and Analyses

All data were transcribed and coded offline for accuracy. A second coder transcribed a random subset of responses (40% of each participant's response). Inter-rater reliability for accuracy was high (Cohen's Kappa of 0.91). All the RTs registered with the voice key were double checked by viewing the acoustic wave for individual words in PRAAT (Boersma, 2001) and measuring the onset of the word from the onset of the beep (which marked the onset of the picture). Word durations were also calculated by marking the offset of words in PRAAT using text grids, with manual checking, and subtracting the onsets. We included analyses of both RTs and durations because interference may manifest in one or another or as a tradeoff between the two (Fink, Oppenheim, & Goldrick, 2018; Nozari & Hepner, 2019). For example, if taxonomic effects cause

longer RTs but shorter durations than the baseline, it would be difficult to claim that their effect is purely inhibitory.

Analyses were conducted in R (version 4.0.3; R Core Team, 2020). Data were analyzed by linear mixed effects models (LMEMs) using the lme4 package (version 1.1-27.1), in conjunction with lmerTest (version 3.1-3) to calculate p-values using Satterthwaite approximations. For the analyses that compared multiple non-orthogonal contrasts, the multcomp R package version 1.4-16 (Bretz et al., 2011) was used, which returns Tukey-corrected p-values for pairwise comparisons but does not tolerate interaction terms in the model. Values more extreme than 3SD from the mean of RT and duration distributions were excluded from the analyses. Both RT and duration distributions were right-skewed and were log-transformed to approximate normality. The random effect structure was kept maximal according to the recommendation by Barr et al. (2013), unless the model did not converge. In such cases, the random effect structure was reduced by first eliminating the slopes for items and subjects. As most models did not tolerate random slopes, we included the random intercept of subjects and items in all models for consistency.

We first tested the three critical contrasts of interest (taxonomic vs. unrelated, thematic vs. unrelated, and taxonomic vs. thematic) on the whole dataset. Since distractors are carefully matched between conditions, we would expect differences to reflect the relations between items, as opposed to the individual properties of items. Next, to ensure that individual items' properties did not contaminate the results, we ran the same model only on a subset of the data containing the target items. In other words, in the second analysis the same items are compared across the three conditions. It is also important to model the effect of switches that are nearly twice as common as repeats to account for any strategic preparation of a switch and its possible interaction with condition. The multcomp package does not tolerate interaction terms in the model, so we ran a third model that contained a switch variable (i.e., a centered binomial predictor to code whether the target was the same as the previous trial or not) and its possible interaction with the two most critical contrasts (taxonomic vs. unrelated, i.e., the replication of the semantic blocking effect, and taxonomic vs. thematic, i.e., the key difference under investigation).

The data and the analysis code are available online (<https://osf.io/qrg72/>). The full results of all the analyses are reported in Appendix B.

Results

Sixty participants provided 12960 responses. The error rate was low (46) and comparable across the three conditions (< 1% in each condition). In addition to errors, extreme outliers were excluded from the RTs (1%) and durations (< .5%). Figure 2 shows the pattern of results for the RTs. The first-pass analysis tested three pre-planned contrasts, with log(RT) as the DV, semantic context (taxonomic, thematic, and unrelated) as the fixed effect, and intercept of subjects and items, as random effects. The results with Tukey-corrected p-values revealed significantly longer RTs in the taxonomic than the unrelated condition ($\beta = -0.021 \sim -9.3$ ms when back-transformed, $z = -3.84$, $p < .001$), no significant difference between thematic and unrelated conditions ($\beta = 0.004 \sim -1.6$ ms, $z = 0.68$, $p = 0.77$), and significantly longer RTs in the taxonomic compared to the thematic condition ($\beta = -0.024 \sim -10.9$ ms, $z = -4.53$, $p < .001$; Table B1.1). Focusing the analysis on the target words returned similar results; RTs were significantly longer in the

taxonomic condition compared to both unrelated ($\beta = -0.021 \sim -9.5$ ms, $z = -3.91$, $p < .001$) and thematic ($\beta = -0.025 \sim -11.1$ ms, $z = -4.53$, $p < .001$) conditions, whereas the thematic condition did not significantly differ from the unrelated baseline ($\beta = 0.003 \sim -1.6$ ms, $z = 0.62$, $p = 0.81$; Table B1.2).

Next, we checked whether the significant effects uncovered in the above models were altered by repeating or switching. This model's fixed effect structure included repetition (center-coded as repeat vs. switch), semantic context (contrast coded as taxonomic vs. unrelated, and taxonomic vs. thematic), and the interaction between the two. The random effect structure was kept the same as the previous models. As anticipated, there was a significant main effect of switching ($\beta = -0.026 \sim -11.9$ ms, $t = -4.58$, $p < .001$), but crucially both semantic context contrasts remained significant as in the earlier models (both $p < .001$), and neither was moderated by the switching effect (both $p > .5$; see Table B1.3 for full details).

In addition, we analyzed response durations to assess possible trade-offs between the speed of speech initiation and completion of an utterance. Durations were close in the three conditions (all geometric means ~ 320 ms), and in a first-pass analysis that paralleled that reported above for RTs, no significant differences emerged for any of the pairwise comparisons (taxonomic vs. unrelated: $\beta = 0.004 \sim -1.1$ ms, $z = 0.64$, $p = 0.80$; thematic vs. unrelated: $\beta = -0.002 \sim -0.7$ ms, $z = 0.41$, $p = 0.91$; taxonomic vs. thematic: $\beta = 0.006 \sim -1.8$ ms, $z = 1.06$, $p = 0.54$; Table B1.4). The model including only target words returned a similar pattern of results (Table B1.5). Finally, adding repetition and its interaction with the critical contrasts did not change the results (Table B1.6).

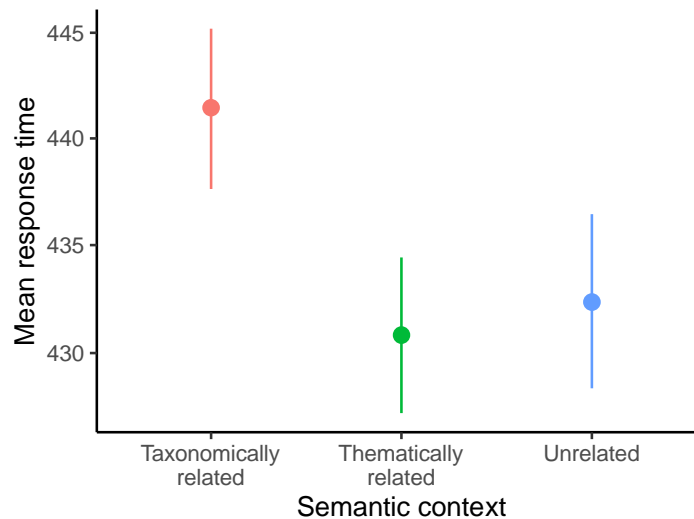


Figure 2. Mean naming latencies for targets words in Experiment 1 (back-transformed log scale) as a function of semantic context. Error bars represent bootstrapped confidence intervals.

To summarize, the results of Experiment 1 replicated the interference effect induced by a taxonomic relationship between items. There was, however, no evidence that thematic relations induced comparable interference. If anything, RTs in the thematic condition were numerically faster than those in the unrelated baseline. Critically, the difference between the effect of

taxonomic and thematic relations was significant, pointing to considerable differences in the influence of these two types of semantic relations on production. A comparable set of analyses for word durations ruled out a possible trade-off between slower speech initiation and shorter durations in the taxonomic condition compared to the other two. These results were robust when the analyses were performed on the entire dataset (with closely matched distractor items) as well as on the target items (which were identical across conditions). In short, we found clear evidence of taxonomic, but not thematic, interference.

Interim discussion

Before considering the neurocomputational bases of the contrast between taxonomic and thematic effects in word production, we needed to empirically verify its existence. Therefore, Experiment 1 first used a highly controlled picture naming procedure to evaluate the three-level distinction between the effects of taxonomically related contexts, thematically related contexts, and unrelated contexts. In picture naming, taxonomic relations create interference, compared to the other two conditions, while the non-significant effect of thematic relations was, if anything, in the direction of facilitating word production. Experiment 2, reported in the Supplementary Information, uses the same materials in a word-picture matching task to strengthen the case that the contrasting effects of taxonomic and thematic relations are indeed specific to word production. Thus, in line with the bulk of prior findings, the empirical contrast between the effects of taxonomic and thematic relations on production survives its most stringent test to date, and one must therefore ask *why* taxonomic and thematic relations elicit contrasting effects in word production?

As noted in the Introduction, these contrasting behavioral effects have been repeatedly claimed as evidence for two contrasting mechanisms for word selection. Interference, whether taxonomic or thematic, has been claimed as evidence of competitive lexical selection (e.g., Roelofs, 2018), whereas facilitation or otherwise diminished interference from thematic manipulations has been claimed as evidence of non-competitive lexical selection. While it is plausible that speakers could engage different executive control processes for distinct production-related tasks (e.g., Nozari & Hepner, 2019; Oppenheim, Balatsou, & Thierry, under review; Roelofs & Piai, 2013), in Experiment 1 both contrasting effects emerged for the same speakers in the same simple picture naming task in the same session, thus casting doubt on such control-based accounts of these semantic effects.

A more compelling explanation is that the taxonomic/thematic contrast reflects differences in the ways that speakers represent and engage such relations in the word retrieval process. While taxonomic features are item-specific, defined by the differences between exemplars of a category (e.g., Cree et al., 2006; Markman & Ross, 2003), thematic features are inherently relational, defined by their commonality to multiple items that typically co-occur in certain contexts (e.g., Barsalou, 1983). Because they are acquired and defined by the binding of a concept to a context, it follows that the activation of thematic features in production ought to similarly depend more their contextual relevance. In the following simulations, we argue that a contrast between taxonomic and thematic effects in blocked naming naturally arises from a relatively minor difference in the timecourse of semantic activations. The item-specificity of taxonomic features requires rapid activation and deactivation, while the multi-item relationality of thematic features requires activation that persists over multiple words (i.e., residual activation) and thus makes their activation more sensitive to a broader discourse context. We will show that implementing this

distinction in a simple incremental learning model of word production is sufficient to produce a taxonomic/thematic contrast in semantic blocking effects that is similar to that observed in Experiment 1, that such a distinction is learnable on that basis of typical language use, and that the same contrasting effects emerge regardless of the particular lexical selection algorithm.

Model description

For these simulations, we adapt the framework of Oppenheim, Dell, & Schwartz's (2010) Dark Side model of incremental learning in semantically driven word production, implemented in R/C++ (code for the model and all simulations is available at <https://osf.io/qrg72/>). This simple model maps shared semantic features onto words via a single layer of connections, adjusting the connections after each retrieval to reduce activation error. It was originally proposed as an explanation of taxonomic cumulative semantic interference in tasks including blocked cyclic picture naming, thus allowing us to straightforwardly compare selection rules under two approaches to representing thematic associations.

Architecture. As illustrated in Figure 3A, the model consists of an array of 18 semantic nodes, described⁴ as semantic features, directly mapped to an array of 36 word nodes via a single layer of learnable connections, initialized at zero. In Figure 3A, the semantic layer is depicted as three subsets of six orthogonal semantic features; the first subset represents nominally thematic features, the remaining subsets represent nominally taxonomic features.

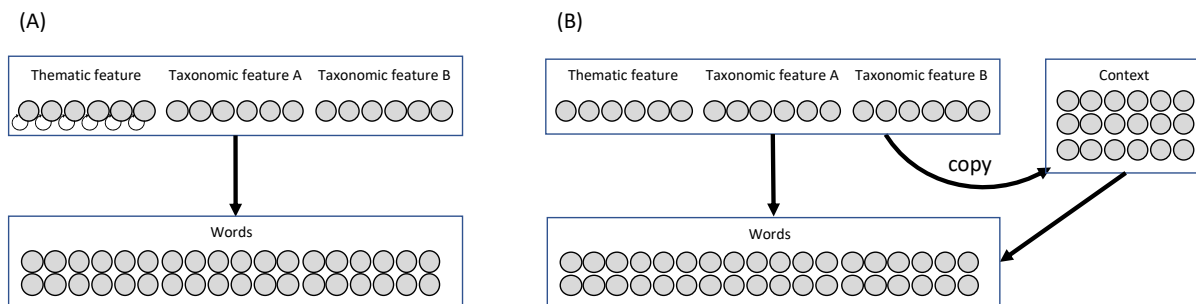


Figure 3. Architecture of the model, as implemented in Simulations 1, 2, and 3 (Panel A) and Simulations 4, 5, and 6 (Panel B). In Simulations 1, 2, and 3, recurrent connections distinguish thematic features from taxonomic features, slowing their activation and deactivation. In Simulations 4, 5, and 6, the recurrent connections for thematic inputs are replaced by an SRN-like ‘context’ layer (i.e., representing the previous value of each input), thus allowing the taxonomic/thematic distinction to emerge from the sequential statistics of the training set.

Representations. Following Oppenheim et al. (2010), this model represents each word as the unique intersection of two taxonomic features. For instance, within its small vocabulary, the output word ‘dog’ is the intersection of [mammal] and [terrestrial] input features, while the output word ‘car’ is the intersection of the [vehicular] and [terrestrial] input features. A simple extension of this approach represents thematic features like [birthday party] in exactly the same way. Note that,

⁴ Though the term ‘feature’ can imply a localist approach to semantic representation, the model is similarly compatible with distributed approaches.

although the model contains three orthogonal subsets semantic features (and thus three positive features for each word), an intersection of two of the three is sufficient to uniquely specify a word so, e.g., thematic features are not required for retrieval but can contribute to retrieval (e.g., [sweet] + [spongy] + [birthday party] \rightarrow “cake”) or compensate for taxonomic underspecification (e.g., [spongy] + [birthday party] \rightarrow “cake”).

Algorithms. Following Oppenheim et al. (2010), each trial consists of three phases: activation, selection, and learning.

Activation. Each trial begins with the activation of semantic features, with nominal values of zero or one, $a_{j,trial\ n}$. To integrate activation across successive trials, these values are adjusted per Eq. 1, below, where $a_{j,trial\ n-1}$ is the activation of input node j in the previous trial and λ is a volatility (or, conversely, inertia) parameter to modulate the carryover of semantic activation from one trial to the next ($0 < \lambda \leq 1$), implementing a form of recursion. Under this equation, $\lambda=1$ eliminates trial-to-trial carryover of activation, as implicitly assumed in Oppenheim et al. (2010), while smaller λ cause a_j to asymptote toward 0 or 1 as node j is consistently activated over successive trials.

$$a_j = a_{j,trial\ n} \lambda + a_{j,trial\ n-1}(1 - \lambda) \quad \text{Eq. 1}$$

From these input activations, we calculate the initial activation of each output word as the sigmoidal transformation of its net input, including a small amount of normally distributed noise. Thus, the net input for each word output node i is calculated in Eq. 2 as the sum of the activation of each input node, a_j , times the weight of its connection to node i , w_{ij} :

$$net_i = \sum_j w_{ij} a_j \quad \text{Eq. 2}$$

And the activation for each word output node i is given in Eq. 3 as a sigmoidal function of the net input, plus a small amount of normally distributed noise (with a mean of 0 and a variance of ν):

$$a_i = \frac{1}{1 + e^{-(net_i + N(0,\nu))}} \quad \text{Eq. 3}$$

Selection. An active selection process plays out over time, linking each word’s activation to its likelihood and timing of selection. We assume that this selection emerges from an evidence accumulation process, which ‘boosts’ task-relevant activations toward a selection threshold, τ , at a rate, β . By ‘task-relevant’, we mean that the selection process does not know which word it ought to select (cf. Roelofs, 1992), but may implement coarse categorical constraints such as specifically boosting English-language nouns but not verbs (Blanco-Elorrieta & Caramazza, 2021; Chang et al., 2006; Dell et al., 2008; Gordon & Dell, 2003; Oppenheim, 2011). To assess the possible role of non-target activations in determining target selection times (i.e., the competitive lexical selection debate), we report selection times according to two formulae:

1. A non-competitive formula (Eq. 4, following Eq. 11 from Oppenheim et al, 2010) assumes that the time of word selection is a simple threshold function of the activation of the most strongly activated word.

$$t_{selection} = \log_{\beta} \left(\frac{\tau}{a_i t_1} \right) \quad \text{Eq. 4}$$

2. A competitive formula (Eq. 5, following Eq. 12 from Oppenheim et al., 2010; cf. Krajbich & Rangel, 2011; Nozari & Hepner, 2019) assumes that the time of word selection is a *relative* threshold function of the activation of the most strongly activated word compared to the most strongly activated alternative. As argued and demonstrated elsewhere (Oppenheim et al., 2010; Oppenheim, 2024), competitive lexical selection algorithms must disproportionately weigh the strongest options if they are to produce identifiable ‘competitive’ characteristics. This is because, as the size of a vocabulary grow, the mean activation of its elements approaches a constant, with the large number of minimally activated elements swamping the much smaller number of highly activated alternatives that are assumed to produce competitive selection effects.

$$t_{selection} = \log_{\beta} \left(\frac{\tau}{a_i t_1 - a_{strongest\ competitor} t_1} \right) \quad \text{Eq. 5}$$

We have selected these algorithms for the sake of expository clarity, but others that we have considered yield similar predictions when considered at the same grain.

Learning. At the end of the trial, the weight of each connection is adjusted according to the Widrow-Hoff (1960) Delta Rule (Eq. 6), implementing a supervised corrective learning algorithm:

$$\Delta w_{ij} = \eta(a_i(1 - a_i)(d_i - a_i))a_j \quad \text{Eq. 6}$$

Although, as we have mentioned, the selection process does not know which word it ought to select, we do assume that speakers can approximate such knowledge as the basis for the learning process, for instance as a result of self-monitoring (e.g., Middleton et al., 2022; Nozari et al., 2011; Postma, 2000) or a contrastive Hebbian learning (e.g., Movellan, 1990; Norman et al., 2007).

Parameters. Model parameters are given in Table 1. These are identical to the parameters used in Oppenheim et al. (2010), with one exception: because we are only interested in response times in these simulations, we reduced the activation noise parameter from 0.5 to 0.1.

Table 2. Model parameters.

Parameter	Value
Learning rate (η)	0.75
Activation noise (ν)	0.10
Boosting rate (β)	1.01
Threshold (τ)	1.00
Deadline (Ω)	100
Activation volatility (λ)	1.00 unless otherwise specified

Simulation 1

The first simulation aims to capture the semantic interference effect that is consistently reported for taxonomic manipulations and has sometimes been reported for thematic manipulations. This

represents a straightforward adaptation of Oppenheim et al.’s (2010) approach, demonstrating robust interference in a task that approximates our Experiment 1 when $\lambda = 1.00$ for all features. Thus, we start by testing competitive and noncompetitive models under the assumption that thematic associations contribute to word production in exactly the same way as taxonomic associations: in each trial, several relevant semantic features are activated—we give these input activations a value of 1, assuming that they quickly asymptote towards this value—while irrelevant features remain inactive—we give them a value of 0, assuming that any such activation must quickly dissipate to allow fluent production, asymptoting toward zero after each selection (e.g., Bock & Griffin, 2000).

Pre-training was implemented exactly as in Oppenheim et al. (2010), via 100 randomly ordered sweeps through the full vocabulary, mapping semantic features to words and updating connection weights via the delta rule (Eq. 6). Then, the model was tested on a task approximating our Experiment 1: six randomly ordered repetitions of two stimuli that shared either a ‘taxonomic’ feature (taxonomically related blocks), a ‘thematic’ feature (thematically related blocks), or no features (the unrelated baseline).⁵ Response times for each trial were calculated according to both the noncompetitive selection rule specified in Eq. 4 and the competitive selection rule specified in Eq. 5. The full simulation consisted of 1000 individually trained models, tested in each condition with each selection rule.

Results and discussion

As illustrated in Figure 4A, Simulation 1 produced equivalently robust interference for target selection times in taxonomically related and thematically related blocks, compared to the unrelated baseline condition, and this same pattern emerged under both competitive and noncompetitive selection rules.

In this simulation, as in previous work (Oppenheim et al., 2010), the interference resulted from error-based incremental learning and unlearning. Learning algorithms typically produce efficient target activation via gradient descent, using small, error-based changes to ensure that inputs activate target outputs more and competing outputs less. Thus the connections that support a former target become persistently strengthened and those that support former competitors become persistently weakened, a process that facilitates retrieving the same target again in the future. Such algorithms implicitly assume the recent past is a good predictor of the near future (e.g., in discourse, one might expect a conversation about cows to remain focused on cows rather than shifting to bison). However, when the structure of a task violates that assumption, requiring former competitors to become the targets in subsequent trials, that same incremental learning produces interference. Thus, as illustrated in Figure 5A, connections from shared taxonomic features were relatively reduced in taxonomically related blocks, and connections from shared thematic features were relatively reduced in thematically related blocks.

The pattern of equivalent interference from overlap in nominally taxonomic and thematic features is a straightforward result of their equivalent representations in this simulation. Nonetheless, it

⁵ *N.b.* because $\lambda = 1$ for all features in Simulation 1, there was effectively no difference between the taxonomic and thematic features.

usefully represents one end of a continuum from context-independence to context-dependence that thematic features could occupy (e.g., Barsalou, 1982). It also illustrates the sometimes-arbitrary distinction between taxonomic and thematic relations: for instance, although nationality may be reasonably considered as a basis for a theme-based category (e.g., Abdel Rahman & Melinger, 2007), associations such as [FRENCH] → “beret” may nonetheless be acquired and engaged as taxonomic-like theory-based features (e.g., Murphy & Medin, 1985) in the absence of direct experience.

Having demonstrated that both types of relations elicit interference when implemented with equally rapid activation and deactivation, in Simulation 2 we consider the consequences of the assumption that the activation and deactivation of thematic features is somewhat more gradual.

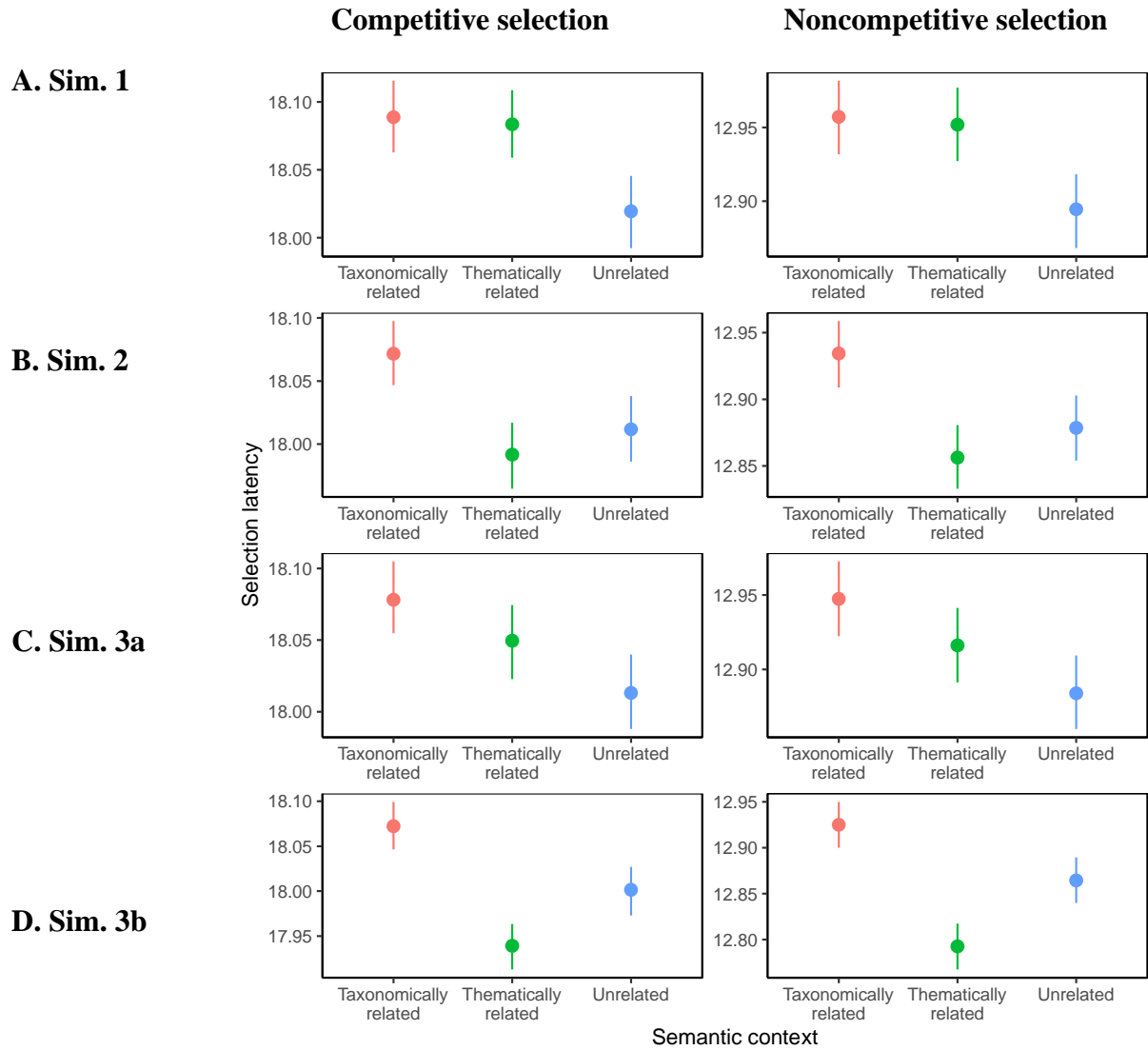


Figure 4. Simulated selection times in Simulations 1 (Panel A), 2 (B), and 3a (C), and 3b (D), under competitive (left column) and noncompetitive (right column) selection rules.⁶

⁶ Note that although holding parameterization constant across all models produces some panel-to-panel variation in y-axis ranges for Figures 4-7, such variation is not theoretically meaningful because the goal is strictly to characterize the ordinal relations between the conditions within each panel. We therefore scale each panel to fit its simulated range.

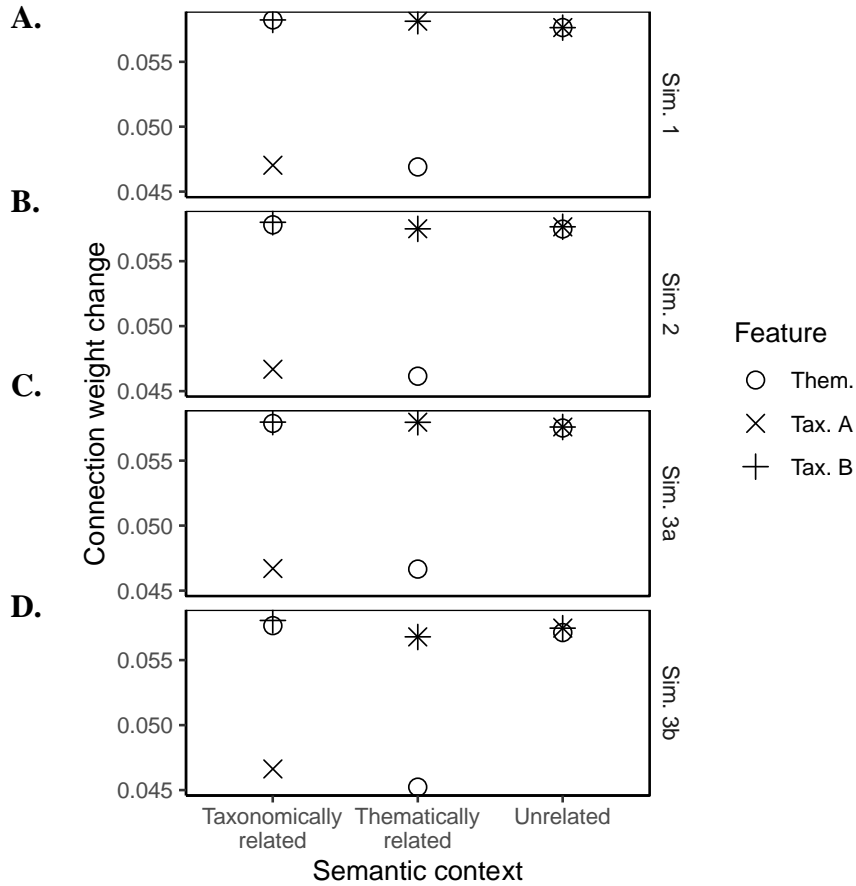


Figure 5. Mean weight changes in each simulated semantic context block for connections to the target word from its critical taxonomic and thematic features, in Simulations 1 (Panel A), 2 (B), 3a (C), and 3b (D). The incremental learning that follows each trial strengthens the current mapping and weakens alternative mappings, per Eq. 6. Thus, accessing a target word strengthens the connections that support it, but subsequently accessing a semantically related word can attenuate or reverse those changes. This dynamic affects both taxonomically and thematically related blocks, and does so to a similar degree in each simulation.

Simulation 2

Simulation 1 showed that the model can capture the semantic interference effect that researchers have consistently reported in blocked cyclic naming experiments that focus on shared taxonomic features. If thematic features contributed to production in the same way—simply as additional shared features—then they would generate the same interference effects. In fact, because semantic representations develop in response to individuals' experiences, which obviously vary, it is possible that some individuals may represent and use some thematic associations in this way (e.g., associations with the Eiffel Tower based on reading about monumental architecture will differ from those based on a visit to Paris).

However, it is difficult to defend the strong claim that taxonomic and thematic associations *necessarily* contribute in exactly the same way. For one thing, their scope of relevance clearly differs: while taxonomic features (e.g., “furry”) typically describe only a single word in a sentence—and must therefore activate and deactivate quickly—thematic associations (e.g., gives (cow, milk)) necessarily integrate multiple terms, so it makes less sense to assume that the same rapid activation/deactivation timecourse applies. In Simulation 2, we implement this insight through the minimal change that the thematic nodes asymptote toward 1 and 0 across multiple trials instead of within each individual trial. Therefore, the activation volatility constant, λ , was reduced from 1.000 to 0.995 for thematic features, but remained at 1.000 for taxonomic features, allowing thematic features to carry a small amount of residual activation over into the next trial.

Results and discussion

As illustrated in Figure 4B, reducing the volatility of thematic feature activation by just half of one percent was sufficient to negate the thematic interference that we had observed in Simulation 1 selection times, producing a null effect for the thematic / baseline comparison. The taxonomic interference effect, however, was left unchanged. Moreover, both competitive and noncompetitive selection rules produced this same pattern of results. Simulation 2 thus captures the empirical contrast between robust taxonomic interference and null thematic effects, such as we observed in Experiment 1, while demonstrating that the contrast is similarly compatible with either selection rule.

It is important to note that this attenuation of the thematic effect is not caused by any attenuation of the connection changes that had created the thematic effect in Simulation 1. As illustrated in Figure 5B, the connection changes in the two simulations are nearly indistinguishable: the same unlearning that produced interference for the thematic blocks in Simulation 1 (and for the taxonomic blocks in both simulations) is also present for the thematic blocks in Simulation 2, and should therefore produce thematic interference. So where does the null effect in Simulation 2 come from? The answer is that reducing thematic feature volatility produces a residual activation-based facilitation effect to offset the weight-based interference effect that would otherwise emerge. That is, Equation 2 specifies that the net input to any word is the sum product of the activated semantic features and the strength of their respective semantic-to-lexical connections, implying that increasing semantic activations can counteract decreasing connection weights. In Simulation 3, we show how reducing or increasing the activation volatility constant could explain empirical variation in thematic blocking effects.

Simulation 3

Simulation 2 showed that a single model can concurrently capture both interference from taxonomic relations and null effects from thematic relations using either selection rule, under the minimal assumption that thematic features activate and deactivate more gradually than taxonomic features. The change that produced these null effects was remarkably subtle: a half percent reduction in a single parameter. Although the resulting pattern matches that of the response times in our Experiment 1, as we discussed in the Introduction, variability seems to be a feature of the thematic context effects in the broader literature. Some studies have reported significant (if apparently attenuated) interference in thematically related blocks (Abdel Rahman & Melinger, 2007, 2011), others have reported nonsignificant effects that trended toward facilitation (de

Zubicaray et al., 2014; our Experiment 1), and if we assume a common source for thematic effects in the picture-word interference literature then we may also need to explain how thematic relations could produce significant response time facilitation. In Simulation 3, we assess the consequences of further reducing or increasing the activation volatility constant. Therefore, the activation volatility constant, λ , for thematic features was set to 0.9975 in Simulation 3a—slightly greater than in Simulation 2—and 0.99 in Simulation 3b—slightly less than in Simulation 2. As in the previous simulations, activation volatility for taxonomic features remained at 1.00.

Results and discussion

As illustrated in Figure 4C and D, a smaller reduction in the volatility of thematic feature activation resulted in a weaker thematic interference effect for selection times in Simulation 3a, while a larger reduction in thematic feature volatility produced thematic facilitation in Simulation 3b. Again, the two selection rules produced equivalent results. And again, the selection time patterns dissociated from the patterns of weight changes, illustrated in Figure 5C and D. Simulation 3 thus demonstrates that reducing semantic feature volatility in this simple model creates residual activation-based facilitation that can reduce, balance-out, or reverse the semantic interference effects that one would expect based on the consistent patterns of weight changes for all simulations Figure 5. Such variation may explain some of the discrepancies in empirical assessments of thematic context effects, a point that we will return to in the General Discussion.

Summary of Simulations 1-3 and motivation for Simulations 4-6

In Simulations 1-3, we demonstrated that differences in the time course of semantic activation can turn semantic interference into facilitation. But how do speakers acquire such a distinction? The most plausible answer is that they *learn it through experience*. In Simulations 4-6, we will show how speakers could acquire such context-sensitive thematic activations through experience, leveraging the insight that thematic associations are often manifest in patterns of lexical (and presumably semantic) cooccurrence. To address this possibility, we must adapt the model to attend to cooccurrences and adapt the training process to provide them.

One classic model that is known for its sensitivity to sequential regularities is the simple recurrent network (SRN; Elman, 1990). The standard architecture for an SRN includes an input layer, an output layer, and a layer of hidden units between them, but what makes it especially useful is a layer of ‘context’ units that provide its memory. When applied to a task of, e.g., predicting the next word in a sentence, activation will initially flow from inputs to the hidden layer to the outputs—a standard procedure for any feedforward network—but then the activation pattern from the hidden layer is ‘copied’ to the context units, which serve as an additional set of time-delayed inputs for the next trial. Error-based learning then allows the model to adjust connection weights from its context nodes in the same way that it would adjust connections from any other input, but the time delay means that these connections come to represent transitions from the preceding units. Tracking unit-to-unit transitional probabilities in this way immediately provides a basis for parsing continuous speech into words and parsing strings of words into clauses, and it indirectly provides a basis for identifying syntactic categories and taxonomic hierarchies (*ibid*; cf. Mikolov et al., 2013). And although the SRN architecture was originally applied to language perception and

comprehension, it also plays an in essential role in contemporary models of sentence production (e.g., Chang et al., 2006).

Without modifying our base model too much⁷, we can adapt from the SRN the idea of a context layer that represents a time-delayed activation pattern without prejudice, so connections from that layer specifically represent the utility of that delayed signal (see Figure 3B). The traditional approach when modeling perception and comprehension tasks is to track the transitions between lower level elements like words, but speakers have the great advantage of knowing what they mean, so we can model semantic transitions directly. The volatility parameter (λ) that we used in Simulations 1-3 can already be described as implementing a form of recursion, but our use of it distinguished between thematic and taxonomic inputs by fiat, so we eliminate it from Simulations 4-6. Instead, at the end of each trial, the current state of the 18 semantic input nodes is copied into the 18 context nodes (thus providing a context analogue of each semantic input), which then provide additional inputs for the next trial. Their connections to the output layer must be learned through experience. Thus, if the model should distinguish between taxonomic and thematic features it must learn to do so on the basis of patterns in its input.

Simulations 4-6 are intended to provide learned analogues of Simulations 1-3, respectively. Thus, Simulation 4 will present our modified base model, including the new context nodes but maintaining the same random-without-replacement training order that we had used in the first three simulations, to demonstrate that the architecture and algorithms do not compel a taxonomic/thematic distinction. Simulations 5, 6a, and 6b will then show that modulating the probabilities of thematic cooccurrences in the model's training orders can approximate the pattern of thematic effects that we had obtained in Simulations 2, 3a, and 3b, thus demonstrating that the full range of outcomes that we reported from those previous simulations is learnable from experience under commonly held assumptions about sequential language processing.

Simulation 4

Aside from the addition of context input nodes as described, Simulation 4 was made identical to Simulation 1 to show that it produced similar results despite the addition of the context units. Thus, during training each target followed each other with equal probability.

Results and discussion

As illustrated in Figure 6A, the SRN-inspired model tested in Simulation 4 produced equivalent semantic interference for both taxonomic and thematic relations, and did so whether using competitive or noncompetitive selection algorithms. This result is reassuring: although the revised architecture should allow the model to detect cooccurrence patterns if they exist, it does not compel the model to treat nominally taxonomic and thematic relations differently. Here it treats them the same because the training set provided no basis for distinguishing between them: the sequencing of items in the training set was random, so thematically related items were no more likely to

⁷ This implementation is restricted to short-range dependencies, but its recursion could be easily be extended (e.g. recursively copying the states of the $t-1$ context units to additional sets of $t-n$ context units) or its architecture augmented to capture longer-range dependencies (e.g. via LSTM). The goal here was to provide a first approximation of the approach to demonstrate the learnability principle while maintaining the transparency of the base model.

cooccur than taxonomically related items, and thus the learned connections from thematic context features were no stronger than those from taxonomic context features (Figure 7). If human thematic facilitation effects reflect processes similar to this model, then limited experiential cooccurrence might explain some of the variability in the human data. In Simulation 5, we will combine the unbiased context layer with a training set in which thematic relations are more likely than taxonomic relations to occur in close proximity, thereby assessing whether that minimal change is sufficient to elicit the expected contrast between taxonomic and thematic blocking effects.

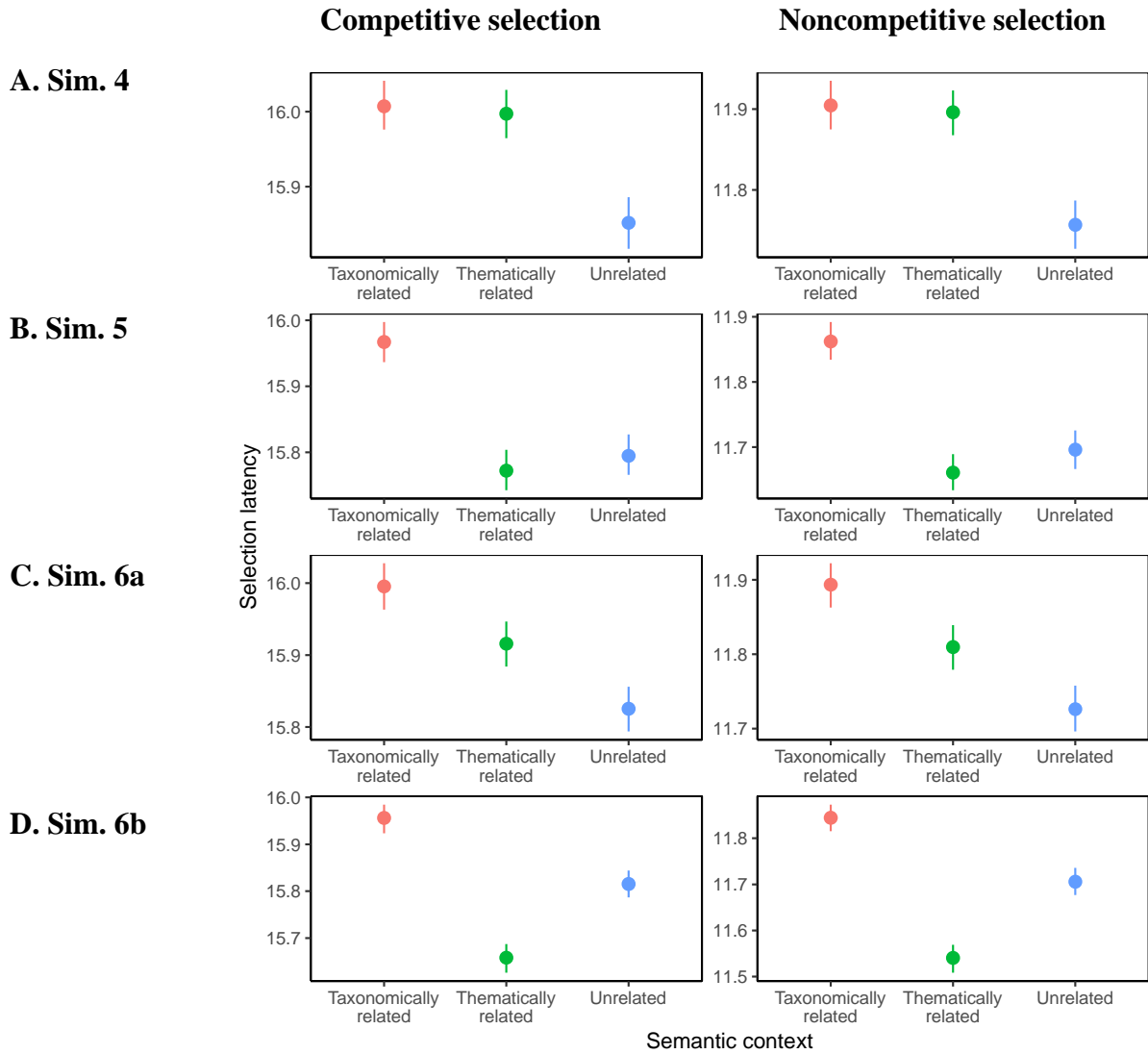


Figure 6. Simulated selection times in Simulations 4 (top), 5 (middle), and 6 (bottom), under competitive (left column) and noncompetitive (right column) selection rules.

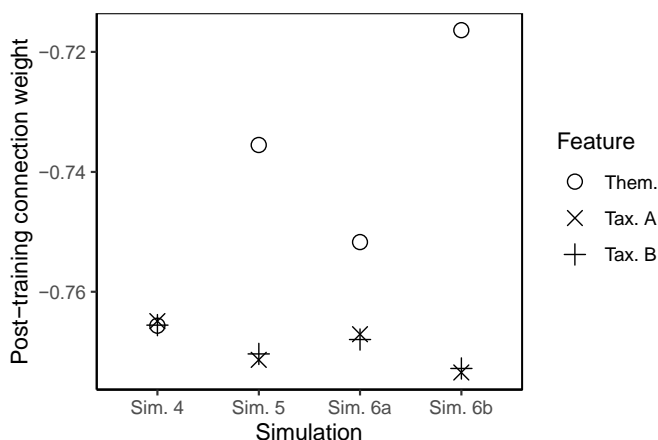


Figure 7. Mean post-training weights for connections to the target word from the context analogues of its critical taxonomic and thematic features in Simulations 4-6. All weights are negative because context node activations only weakly predict lexical activations, but thematic context connections grow stronger (i.e., less negative) as the probability of thematic cooccurrence in the training set increases.

Simulation 5

In Simulation 5, we use exactly the same architecture and algorithms as in Simulation 4. The only difference concerned the ordering of items within the pre-experiment training process: to approximate the sequential cooccurrence that typically characterizes thematic associations in discourse, as measured by PMI and LL metrics in the design of Experiments 1 and 2, we used an algorithm to stochastically ‘clump’ items that shared thematic features. Briefly, for each training trial, $trial_0$, we generated a uniformly distributed random number, $U_{[0,1]}$, for each stimulus that had not yet occurred in the current epoch and cued the stimulus with the largest value. For each subsequent trial, $trial_n$, the random value of any stimulus that was associated with the same thematic feature was increased by a small amount that decayed with each subsequent trial, $0.015 * 0.5^{(n-1)}$. This increased the probability of drawing two successive tokens associated with the same thematic feature, from ~ 0.14 to ~ 0.18 , but left their input activations unchanged.

Results and discussion

As illustrated in Figure 6B, the elevated rates of thematic cooccurrences in the training set that we introduced in Simulation 5 combined with the SRN-like context structure that we introduced in Simulation 4 to produce a null thematic effect that contrasted with taxonomic interference regardless of the selection rule. Because the same architecture and algorithms failed to produce such a distinction in Simulation 4, the only possible explanation is that the model acquired the distinction from the relatively subtle manipulation of thematic cooccurrences in its training set. Figure 7 shows that this distinction is now manifest in stronger connections from the context analogues of thematic features.

The overall patterns in Simulations 4 and 5 are very similar to those that emerged in Simulations 1 and 2 when we directly implemented the assumption that thematic features activate and

deactivate more gradually than taxonomic features. This resemblance is reassuring because it shows that the distinction that we had previously hard-coded could emerge from the statistics of normal communicative language use, and by extension from the normal concept use that language presumably indexes. Simulation 6 will similarly parallel Simulation 3, to demonstrate that thematic cooccurrence similarly provides a continuously valued facilitation effect.

Simulation 6

Simulation 6 replicated Simulation 5 exactly, but decreased the thematic priming value in the training set from 0.015 to 0.005 in Simulation 6a, and increased it to 0.025 in Simulation 6b. As before, these changes modulated the probability of drawing two successive tokens associated with the same thematic feature during training, but left their input activations unchanged.

Results and discussion

As illustrated in Figure 6C and D, a smaller increase in the rate of thematic cooccurrences in the training set resulted in a weaker thematic interference effect for selection times in Simulation 6a, compared to Simulation 4, while a larger increase in the rate of thematic cooccurrences in the training set produced thematic facilitation in Simulation 6b. Again, the two selection rules produced equivalent results. And again, the acquired distinction between taxonomic and thematic features was specifically manifest in the connections from the context features (Figure 7). Simulation 6 thus demonstrates that the reversal of thematic context effects that we identified in Simulation 3 can also emerge from lexical cooccurrence, implying that the dynamics in this model can explain the range of outcomes that been reported in the literature.

General Discussion

We set out to assess whether behavioral interference and facilitation in language production imply competitive or non-competitive selection processes, respectively, using the comparison of taxonomic and thematic similarity effects as a test case. To this end, three subgoals were set: (1) to verify a possible empirical distinction between the influence of taxonomic vs. thematic relations on word production, (2) to test by simulation whether such a distinction would actually imply distinct selection mechanisms in the production system, and (3) to assess whether such a distinction could emerge incidentally from experience. In a tightly controlled design, we confirmed the empirical distinction: thematic relationships elicited significantly faster RTs compared to taxonomic relationships, and unlike taxonomic relationships they showed no evidence of interference when compared to an unrelated baseline. The theoretical importance of this empirical distinction cannot be overstated: it implies a basic distinction between these types of semantic features in terms of their representation and contribution to language production processes. Previous studies have hinted at such a difference (Abdel Rahman & Melinger, 2007, 2011; Alario et al., 2000; Aristei et al., 2011; Costa, Alario, et al., 2005; de Zubicaray et al., 2014; Mahon et al., 2007; Navarrete et al., 2014) but the quantitative orthogonal manipulation allowed this study to robustly evaluate and confirm it.

A series of simulations then showed how these contrasting effects could emerge from a single neural network model, extending Oppenheim et al.'s (2010) incremental learning approach. Two core ideas were that (1) semantic blocking effects reflect the same dynamics of semantic activation,

lexical access, and implicit learning that characterize spontaneous language production, and (2) speakers acquire more context-sensitive thematic activations as a learned response to typical patterns of lexical/semantic cooccurrence. Researchers commonly assume that the activation of thematic associations is, in some way, more context-sensitive than that of taxonomic features, and we posited that such context-sensitivity might relate to the linguistic scope of thematic features: whereas taxonomic features usually refer to smaller chunks of discourse (e.g., single words), when thematic features are relevant to discourse they are relevant to larger chunks, containing multiple words that a speaker would presumably access sequentially, so thematic feature activation ought to be more *legato* (to borrow the musical term) compared to the *staccato* activation of taxonomic features. The first set of simulations (1-3) implemented this idea as a hard-wired difference in the timing of semantic activation: taxonomic features activated and deactivated rapidly, within each trial, while thematic features activated and deactivated slightly more gradually, carrying over a bit of their activation across trials. When a feature's activation was modulated rapidly, feature overlap produced strong net interference (Simulation 1). But when its activation was modulated more gradually, feature overlap produced null effects (Simulation 2), weaker interference (Simulation 3a) or even net facilitation (Simulation 3b). It is worth reiterating that the contrasting patterns of interference versus facilitation in these simulations did not reflect special 'exception' cases where the standard rules no longer applied. The same error-based learning and selection algorithms that produced both taxonomic and thematic interference in Simulation 1 continued to operate in Simulations 2 and 3. The only substantive difference was the residual activation of thematic features, which caused them to contribute more input activation when they were activated more consistently across trials, increasing target activation by overshadowing any learning-based reduction in connection weights.

What evidence is there for our assumption that thematic features are activated and deactivated more gradually than taxonomic features? It is easy to show that thematic features initially contribute less to lexical access in production by considering the kinds of names that participants spontaneously produce when asked to simply name a picture. For instance, when classifying word errors that aphasia patients and age-matched controls made on the Philadelphia Naming Test, Schwartz and colleagues (2011) found that both groups produced roughly five times as many taxonomic errors as they did thematic errors (86 patients: 645:134; 20 controls: 44:8). Similarly, when classifying the most common non-dominant names that emerged from a large norming study, Oppenheim (2024) found that the vast majority were closely taxonomically related to the dominant names (e.g., 'turtle' vs. 'tortoise'; 360 out of 460 items that elicited at least two names); undepicted thematic associates emerged much less frequently (e.g., 'teeth' vs. 'dentist', when no dentist was depicted in the image; 19/460 items). Thus, thematic features do not appear to strongly contribute to lexical activation in theme-irrelevant naming tasks. And although we are not aware of any studies that have directly assessed their timecourse in simple picture naming, thematic effects in other production tasks seem to ramp up and dissipate more slowly than taxonomic effects. For instance, in picture-word interference studies where distractors are both taxonomically and associatively (~thematically) related, associative facilitation at long negative stimulus-onset asynchronies (SOAs) appears to give way to taxonomic interference at short negative or positive SOAs (La Heij et al. 1990). In other PWI studies that have orthogonally manipulated the relationships, it is well established that associative or thematic facilitation effects appear at long negative SOAs (Alario et al., 2000: -234 ms; Sailor et al., 2009: -450 ms; Sailor et al., 2013: -300ms), apparently contiguous with the much longer range effects that have been reported in word-primed picture naming tasks (e.g., Sperber et al., 1979, cited in La Heij et al., 1990, as showing

thematic facilitation when unnamed primes preceded picture targets by up to 2000 ms), and are typically weak or absent for the -100 to 0 ms SOAs in which taxonomic interference emerges. Studies of word naming and other visual word recognition-based tasks also suggest such a timing-based distinction (see e.g., Neely, 1991, for a review, and Plaut & Booth, 2000, for a cooccurrence-based computational model): taxonomic facilitation in such tasks is fully disrupted by inserting an unrelated word between prime and target (Masson, 1991; Ratcliff and McKoon, 1988), but associative facilitation persists at least a bit longer, remaining detectable albeit greatly diminished after one intervening unrelated item (Joordens & Besner, 1992).⁸ Thus there is converging evidence from multiple language tasks that both the onset and offset of thematic activation are relatively gradual.

The proposed contrast between residual activation-based facilitation and learning-based interference offers a number of additional predictions, many of which already have some empirical support. For instance, although it is well established that taxonomic interference in the blocked cyclic naming and continuous naming paradigms is robust to the interleaving of unrelated filler items (Damian & Als, 2005; Howard et al., 2006), similar interleaving should reduce thematic facilitation, thereby increasing net thematic interference. In keeping with this prediction, Rose and Abdel Rahman (2016) found thematic interference, instead of facilitation, when they used a continuous naming task, in which thematically related items were separated by unrelated items. In fact, the thematic interference effect in that study was comparable in magnitude to the interference effects that have typically been reported for taxonomic relations. If thematic facilitation were strictly time-bound, in line with the assumptions of the primed naming studies cited above (but cf. the sentence-based studies cited therein), then simple timing manipulations like introducing longer inter-trial intervals in blocked naming (as in Schnur et al., 2006) should similarly allow it to dissipate. Finally, although we would not consider it a theoretical principle *per se*, under the model's current parameterization thematic facilitation quickly reaches its maximum, providing a relatively constant offset to any interference effect, and in the blocked cyclic naming paradigm it should therefore predominantly affect the main effect of semantic context while leaving its interaction with cycle relatively intact; de Zubicaray and colleagues (2014) reported such a result, and Abdel Rahman & Melinger's (2007) Experiment 1 and (2011) Experiment 2 appear to show similar patterns.

Three additional simulations (4-6) illustrated how such gradual thematic activation could emerge incidentally from the distributional statistics of sequential lexical / semantic experience. Replacing our hard-wired timecourse difference with an SRN-inspired 'context layer' that learns transitional probabilities, we first showed that neither the architecture nor the algorithms nor the parameters compelled a thematic / taxonomic distinction: if thematic associates cooccurred at chance during training, then thematic feature overlap produced net interference that was equivalent to that from taxonomic feature overlap (Simulation 4). But when thematically related words cooccurred at above-chance rates during its training, the model acquired correspondingly stronger connections from thematic context nodes, leading thematic feature overlap during testing to produce null effects (Simulation 5), weaker interference (Simulation 6a) or even net facilitation (Simulation 6b), response time effects that mirrored those in the first set of simulations despite holding all

⁸ There is even some evidence from lexical decision tasks that sentence structure can allow associative priming to persist longer than in simple lists of words, presumably because the representations of the component words become better integrated (e.g., Foss, 1982; Sharkey & Sharkey, 1992; Faust & Chiarello, 1998).

model parameters constant. That is, the same kind of lexical cooccurrences that formed the bases of our semantic manipulations in Experiment 1 led our model to learn in Simulations 5 and 6 that activating thematic features predicted their re-use in the near future, thus demonstrating that the timecourse differences that we had imposed by fiat in Simulations 2 and 3 were not only learnable in principle but even likely to be learned from typical language use.

The simulations also addressed the apparent variability in empirical reports of thematic blocking effects, and the possible role of selection processes in explaining it. Simulations 1 and 4 showed that, when thematic features are implemented exactly like taxonomic features, thematic associations produce similar interference, while Simulations 2, 3, 5, and 6 showed that making the activation and deactivation of thematic features slightly more gradual was sufficient to reduce and even reverse thematic interference effects. One plausible explanation for the discrepancy in empirical observations is that actual humans' representations of thematic relations lie along a continuum between those implemented in these simulations. In some cases, research intended to focus on thematic associations may in fact tap into associations that participants have acquired and represented more like classical taxonomic categories, and if the proposal that we explored in Simulations 4-6 is correct, then corpus-derived estimates of lexical cooccurrence should explain at least some of the observed variance. But even among proper thematic associations, researchers have long characterized context-dependency as a continuum rather than a binary attribute, for example suggesting that theme-based categories can develop into taxonomic-like categories via repeated use (e.g., Barsalou, 1983; Medin et al., 1997).

In any case, it is important to note that the variability discussed above did not depend on the selection process: in every simulation, similar patterns of thematic and taxonomic effects emerged regardless of whether the model assumed a competitive or noncompetitive rule for lexical selection. This insight is important for two reasons. First, as noted in the Introduction, empirical observations of thematic interference (or facilitation) have been repeatedly described as evidence for or against competitive lexical selection (e.g., Abdel Rahman & Melinger, 2009; Mahon et al., 2007; Roelofs, 1992; cf., Nozari & Hepner, 2019). Our simulations instead demonstrate that both classes of selection rules are similarly compatible with both behavioral patterns. While a model requires *some* rule for selection, the choice among plausible selection rules is inconsequential. Second, residual activation at the semantic level has often been claimed as an explanation for cumulative semantic interference effects, under the assumption that competitive lexical selection will convert residual activation to lexical interference (e.g., Belke et al., 2005; Roelofs, 2018). However, our simulations demonstrate that neither competitive selection nor residual activation are necessary or sufficient to explain cumulative semantic interference, and in a reasonable model even their combination can instead produce facilitation⁹. In other computational modeling work, it has often proven surprisingly difficult to get 'competitive selection' models to produce 'competitive selection' effects (e.g., Oppenheim et al., 2010, Simulation 5a). Perhaps that difficulty is why the most cited competitive selection model — Roelofs' (1992) WEAVER model — includes an additional stage with the primary function of suppressing alternatives that it already decided against in the previous stage: by dissociating the outcome of a decision from its timing,

⁹ In fact, residual semantic activation has previously been proposed as an explanation for the first-cycle facilitation that is often seen in the blocked cyclic picture naming paradigm (e.g., Scaltritti, Peressotti, & Navarrete, 2017). Our thematic simulations suggest how that could work. If one assumed smaller amount of residual activation for taxonomic features, facilitation could more quickly give way to interference.

the model can create arbitrarily large ‘competitive’ RT effects without creating correspondingly high error rates. Although Roelofs (2018) showed that WEAVER could generate cumulative taxonomic and thematic¹⁰ interference as a product of competitive lexical selection and residual semantic activation, that result can be seen as a consequence of that model’s complexity and unprincipled flexibility rather than the theoretical principles that should be the primary goal of computational cognitive modeling.

Our simulations’ focus on theoretical principles underlying successful language production is both a strength and a limitation. The strength is that each aspect of the model is driven by the intrinsic demands of real-world communication rather than the arbitrary demands of a laboratory task. To communicate an intended message, a speaker needs to map meaning onto a linguistic form, and so the model learns through experience to do just that. The learning that creates semantic interference effects in certain artificially structured laboratory tasks is the same learning that should usually promote quick and accurate lexical access; the matter is simply that these tasks violate the normally useful heuristic that having retrieved a word in the recent past is a good predictor of wanting to retrieve it again in the future (see Oppenheim et al., 2010, for more on this point). Similarly, thematic facilitation (or attenuated interference) effects emerge from, and reproduce, a tendency to mention multiple thematically related entities in the same utterance. Thus, beyond addressing the competitive selection debate, we would characterize the model’s primary utility as identifying experiment-relevant predictions that readily emerge from small number of task-general processing assumptions, and should also emerge from a broader class of models that implement similar assumptions. But the limitation of our approach is that the simulations’ scope is quite limited; they model only a few processes and make no attempt to model other processes that certainly contribute to humans’ performance of the tasks at hand. For instance, picture naming obviously includes both visual and phonological processing, but the model does not include either. Nor does it implement any theories of within-trial performance monitoring or task-driven controlled retrieval or production (e.g., response set bias, selective inhibition, response exclusion). As such, while our simulations can qualitatively fit a range of observed data patterns, from interference to facilitation, it would be hubris to claim an exact quantitative fit for any specific pattern (cf. Roelofs, 1992 et passim). Moreover, while we can confidently conclude that taxonomic interference and thematic facilitation effects are similarly compatible with competitive and noncompetitive lexical selection processes *in principle* – and we believe that is the much more important question in terms of theory development – they may nonetheless present challenges for specific models, which we will consider in the next section.

Alternative accounts of semantic non-interference

Both sides of the competition debate have often assumed semantic facilitation as the default outcome for non-competitive lexical selection accounts (e.g., Belke, 2013; Mahon et al., 2007; Wheeldon & Monsell, 1994; Roelofs, 2018), so the challenge for them has instead been to explain why semantic manipulations sometimes elicit interference. The simulations in this paper show how

¹⁰ Though Roelofs (2018) simulated cumulative thematic interference under the assumption that thematic relations are represented similarly to taxonomic relations, as in our Simulation 1, WEAVER could conceivably amend that claim, as we did in Simulations 2-3.

the Dark Side model (Oppenheim et al., 2010) can simultaneously capture taxonomic interference and thematic non-interference under either selection rule, and that property should extend to closely related noncompetitive selection proposals (e.g., Navarrete et al., 2012 et passim).

So what about competitive selection models? Although observations of semantic non-interference or facilitation do not challenge competitive selection rules per se, they do challenge the role that competitive selection has played in many explanations of semantic interference effects. For instance, suppressing competitors is the main goal of selection in the WEAVER model (Roelofs, 1992), and Roelofs' (2018) central premise when extending it to account for cumulative semantic interference was that residual semantic activation creates strong competitors that require suppression. In fact, the simulations in that paper produced a thematic interference effect that appeared comparable in magnitude to its taxonomic interference effect, so it seems likely that increasing residual semantic activation per our proposal would only make WEAVER's thematic interference even larger. Better developing WEAVER's semantic representations might allow it to accommodate thematic non-interference, though it is not immediately clear how such changes would interact with its established selection processes or accounts of other phenomena. In the past it has often explained away observations of non-interference by assuming that certain relations are somehow excluded from consideration, but as we noted earlier doing so weakens the logical link between a model's principles and its predictions.

Because observations of non-interference have challenged competitive selection accounts for so long, at least two recent revisions have attempted to explain diverging results more gracefully. The first of these, Abdel Rahman & Melinger's (2009) *swinging lexical network*, maintains that lexical selection is fundamentally competitive, but such competition is sometimes masked by facilitatory effects. The account combines two key claims: (1) "semantic" manipulations often elicit facilitation at a semantic level that may be difficult to distinguish from interference at the lexical level, and (2) whether a semantic manipulation produces net facilitation or interference depends on how many lexical competitors it activates. It thus follows that the differences observed in the effects of taxonomic vs. thematic relations may stem not from their differential representations in the production system, but rather from the different number of lexical items they co-activate with the target. We agree with the first claim—the combination of semantic facilitation and lexical interference is key to our own model's account—and acknowledge that the second claim may indeed explain some of the variability observed in the empirical data. However, the second claim also makes a testable prediction that allows us to assess its importance: if one does equate the number of competitors in taxonomic and thematic conditions, then one should expect them to elicit similar effects. Our paradigm did just this, strongly highlighting a single competitor in each condition, yet we did not observe comparable effects for taxonomic and thematic relations. One could argue that, despite their irrelevance to the task, other competitors may have been incidentally activated and the number of such competitors may have been different across different conditions. The modeling results are useful to address this concern; the number of competitors were set to be equal in both simulated conditions. Nevertheless, we were able to model the differential effects with a simple change to the timeline of the activation of the semantic features relevant to each condition. In short, while the current results do not reject the notion that the number of competitors may indeed modulate the effects of semantic similarity, they conclusively demonstrate that a key assumption of the swinging lexical network is not necessary to explain the observed interference effects or the differential effects of taxonomic vs. thematic relations.

The second account, Nozari & Hepner’s (2019) *criterion account*, suggests that speakers use an adjustable threshold to compare candidate responses, allowing to be increased or decreased according to task goals, analogous to a classic speed-accuracy tradeoff. Tasks demanding high accuracy (e.g., choosing between the closely related words “probability” and “likelihood” in the context of statistical analysis when delivering a lecture), increase the value of the comparison parameter, emphasizing the competitive selection process, whereas tasks that do not (e.g., casually talking to a friend about a piece of furniture that would be best described as a “settee” but “sofa” will do as well), reduce the comparison parameter to the point that selection may appear entirely noncompetitive. Thus, selection is claimed to be cognitively penetrable in a particular, limited way that ties competition to accuracy demands. But the mapping of semantic representations to lexical items is outside of the scope of the account. Rather, it provides an explanation for how the *same mapping* can produce interference vs. facilitation with *different task goals*, and the design of our Experiment 1 provides no reason to assume that participants’ task goals differed across block types. Our model can be thought of as addressing the reverse problem: how subtly *different mappings* between the same input and output can produce interference vs. facilitation with *the same task goal*.

To summarize, the current results are novel in the sense that they do not fit within any of the previously proposed accounts that interpreted behavioral facilitation and interference from semantic relations as evidence for or against competitive lexical selection. They also show that, far from causing interference, residual semantic activation tends to produce semantic facilitation, and does so even under a strongly competitive selection rule. Moreover, they provide a parsimonious and theoretically motivated account of the empirical differences between the effects of taxonomic and thematic relations on word production, which naturally stem from how these two types of relations are experienced and represented in the semantic space that feeds into the lexicon, and which connects both interference and facilitation in laboratory experiments to functions that support fluent production in more typical communicative contexts.

Thematic effects in picture-word interference

In this paper we have primarily focused on thematic effects in semantic blocking paradigms, but we have also discussed related effects in picture-word interference and production models based on that task. There is an open question about how the paradigms’ semantic effects relate and whether it would be appropriate to assign both to a common source. Recent evidence from our own labs suggests that the neural states can differ substantially based on the underlying task, even when the behavioral outcome is similar. For instance, Oppenheim, Balatsou, and Thierry (under review) compared the first-pass EEG correlates of picture name agreement to those of an intervention that orthogonally either confirmed participants’ initial responses (e.g., ‘sofa’) or ‘corrected’ them to use dominant or secondary names from previous norms (e.g., ‘couch’). After the intervention, the early (230-320ms) frontocentral negativity that was initially associated with low name agreement disappeared, and a late (>500ms) frontocentral positivity associated with directed name change appeared, suggesting fundamentally different mechanisms for resolving endogenous versus exogenous name conflict. More to the current point, Pinet and Nozari (2023)

compared interference resulting from contextual similarity to interference resulting from Stroop-like manipulations when a prepotent lexical label is to be suppressed in favor of another. They found that the two tasks produced distinct EEG patterns. Representational Similarity Analysis applied to EEG data further suggested that the Stroop-like manipulation induced active suppression of the prepotent competing label, whereas no such overt suppression was observed in either semantically or phonologically similar context, suggesting different underlying control mechanisms for dealing with these two types of difficulty. In the same vein, the unnamed distractors in the latter paradigm have no equivalent in tasks that simply manipulate contextual similarity among all-to-be-named items, and most likely recruits a different set of control mechanisms, a matter that is especially important because picture-word effects are often ascribed to task-specific distractor-suppression processes (e.g., Mahon et al., 2007; Roelofs, 1992), rather than processes that are more integral to target word retrieval. Despite these differences, it has been argued that parsimony demands an account to explain findings from both paradigms (e.g., Roelofs, 2018). Although we believe that the empirical evidence is sufficiently strong to obviate the need for an umbrella explanation, as we discuss below, the current account is capable of providing such an explanation and thus meets the principle of parsimony.

A key difference between the manipulation of contextual similarity and PWI is the timecourse. Because the weight changes that we described as the basis of semantic interference for serially presented targets in semantic blocking tasks (following Oppenheim et al., 2010) would presumably require some time, they could not play any obvious role in explaining the interference from simultaneously presented distractors in picture-word interference. In a sense, that limitation is good news: if we simply eliminated the learning-based interference from our model, then even the weaker thematic manipulations in Simulations 3a and 6a would produce thematic facilitation, in line with the consistent reports of thematic facilitation in the picture-word interference literature. On the other hand, it is also less clear how the specific transitional probability-based dynamics that we explored in Simulations 5-6 would apply to simultaneous or near-simultaneous target-distractor presentations. Therefore, although the strong distractor suppression in picture-word interference remains beyond the scope of our model, and we assume that taxonomic interference in that paradigm is a result of that suppression, we can offer a possible account of the paradigm's thematic facilitation: Simulations 1-3 showed that residual semantic activation tends to produce semantic facilitation under any decision rule, and Simulations 4-6 suggest that residual semantic activation would tend to be more beneficial for thematic features than for taxonomic features, and might therefore be greater. Near-simultaneous target/distractor presentation in picture-word interference should maximize residual activation-based facilitation, as in Simulation 3b, and obviate learning-based target/distractor interference as discussed above, leaving only a net effect of thematic facilitation.

Broader implications for models of language production

Some of the most influential models of language production, e.g., DIVA and GODIVA (Guenther, 1995; see Tourville & Guenther, 2011, for a review) focus entirely on mapping a phonological

form onto motor movements, thus ignoring any earlier (i.e., message-level) operations. A major contribution of psycholinguistic models (e.g., Dell, 1986; Levelt et al., 1999, and their variants) has been the inclusion of semantic concepts and their mapping on to lexical units. Adding concepts and a lexicon has allowed researchers to provide mechanistic explanations for a range of critical observations, such as the commission of semantic errors and their similarity across in neurotypical adults, children, and individuals with brain damage (e.g., Budd et al., 2011; Dell et al., 1997; Oppenheim et al., 2010, 2020), detection of semantic errors in mono- and bilingual adults, children, and individuals with brain damage (Hanley et al., 2016; Martin & Nozari, 2021; McMillen et al., 2020; Nozari et al., 2011, 2019), and finally, the influence of semantic relations on the speed of production, the evidence for which has been extensively reviewed in the earlier sections. However, despite the clear utility of the semantic-lexical module, the nature of representations in this part of the system has received considerably less attention than those of the lower layers, such as phonemes and phonetic features.

Generally speaking, the theoretical commitment to modeling the semantic layer in production models has been weak at best. Most models (e.g., Caramazza, 1997; Dell, 1986 et passim) have maintained agnosticism regarding the nature of such representations beyond assuming that concepts are represented as collections of features in a distributed network. Others have adopted a non-decompositional approach (Levelt et al., 1999; Roelofs, 1992), in which concepts are instead represented as holistic nodes. Drawing on an Aristotelian tradition in which features are considered propositions that require verification, they have argued against featural representations on the basis that (a) research participants tend not to substitute superordinates for specific targets (e.g., rarely saying “animal” for “cat”; Levelt's, 1989, ‘hyperonym problem’) and (b) words with more features are no harder to retrieve than words with fewer features (Levelt et al.'s, 1978, ‘semantic complexity’ problem). But categorization research has long since moved beyond the Aristotelian approach (e.g., Rosch & Mervis, 1975) and the specific arguments do not hold up to scrutiny. The premise of the hyperonym argument can be rejected on empirical grounds by noting hyperonym errors in individuals with aphasia (e.g., Butterworth et al., 1984) and the prevalence of hyperonym relations in picture naming norms (Oppenheim, 2024). And the semantic complexity problem has been computationally irrelevant since the advent of parallel processing (e.g., Rumelhart & McClelland, 1986): implemented connectionist models such as ours typically sum activation without regard to ‘complexity’, so a subset of features can be sufficient to activate a word. In fact, to be flexible and context-dependent, conceptual representations must be able to be activated via different subsets of their features (Yee & Thompson-Schill, 2016). In short, of the two stands taken in production models for instantiating semantic representations, the distributed view is better aligned with the evidence from the field of semantic processing, including embodied views on concept processing (e.g., Barsalou, 1999). Still, the representation of semantic knowledge in production models remains quite impoverished when compared to the more nuanced models in the fields of semantic processing (e.g., Lambon Ralph et al., 2016; Patterson et al., 2007), event representation (e.g., Elman & McRae, 2017), and category learning (e.g., Kruschke, 1992; Love et al., 2004).

One of the key contributions of the current study is to point out that maintaining an impoverished representation of the semantic space has nontrivial consequences for theories of language production. When nuances regarding the nature of representations and dynamics of activation in the semantic space are overlooked, there is a risk of misattribution of differences arising from that space to other parts of the system, such as a process of lexical selection. The current case is

particularly informative because taxonomic and thematic relations are clearly different in the semantic space, but there is no logical reason why they should differentially affect a process such as lexical selection. Our simulations show that even simple improvements in representing the dynamics of semantic-to-lexical mapping to make the process more consistent with the reality of taxonomic vs. thematic relations can easily explain a discrepancy that has caused one of the biggest theoretical debates in the field of language production for over two decades. More generally, this demonstration highlights the importance of better connecting models of semantic knowledge with the current models of language production.

Conclusion

Although our behavioral and computational results vindicate both competitive and noncompetitive accounts of lexical selection, in the sense of showing that both can be compatible with evidence that has been claimed to challenge them, it is a pyrrhic victory for both sides. The idea of competitive lexical selection has captured researchers' imaginations, motivating decades of research and serving as the cornerstone of one of the great theories of language production. But the computational reality is far more mundane. The choice of whether to implement a competitive or noncompetitive selection algorithm often has little if any detectable consequence, and behavioral observations of interference versus facilitation are not sufficient to settle the matter. Competition is far from the only possible source of response time interference effects, so one cannot simply draw a line from the presence or absence of response time effects to the presence or absence of a competition-based selection process.

CRedit author statement

Conceptualization, GMO & NN; Experiments 1 & 2 (Methodology, NN; Software, NN; Validation, NN; Formal analysis, NN; Investigation, NN; Resources, NN; Data Curation, NN; Supervision, NN; Project administration, NN; Funding acquisition, NN); Model and simulations (Methodology, GMO; Software, GMO; Validation, GMO; Formal analysis, GMO; Investigation, GMO; Resources, GMO; Data Curation, GMO); Writing - Original Draft, GMO & NN; Writing - Review & Editing, GMO & NN; Visualization, GMO

Acknowledgments

We would like to thank Delaney McDonagh for her help with data collection and Anna Fisher for her comments on an earlier version of this manuscript. This work was partially supported by the NSF awards BCS-1949631 and BCS-206378 to N.N.

References

- Abdel Rahman, R., & Melinger, A. (2007). When bees hamper the production of honey: lexical interference from associates in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3), 604–614.
- Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes*, 24(5), 713–734.
- Abdel Rahman, R., & Melinger, A. (2011). The dynamic microstructure of speech production: semantic interference built on the fly. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 149–161.
- Abel, S., Dressel, K., Kümmerer, D., Saur, D., Mader, I., Weiller, C., & Huber, W. (2009). Correct and erroneous picture naming responses in healthy subjects. *Neuroscience Letters*, 463(3), 167–171.
- Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, 36(1), 140–155.
- Alario, F.-X., & Moscoso Del Prado Martín, F. (2010). On the origin of the “cumulative semantic inhibition” effect. *Memory & Cognition*, 38(1), 57–66.
- Alario, F.-X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology*, 53A(3), 741–764.
- Aristei, S., Melinger, A., & Abdel Rahman, R. (2011). Electrophysiological chronometry of semantic context effects in language production. *Journal of Cognitive Neuroscience*, 23(7), 1567–1586.
- 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.
- Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, 10(1), 82–93.
- Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, 11(3), 211–227.
<http://www.ncbi.nlm.nih.gov/pubmed/8817461>
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577–609; discussion 610–60.
- Bates, E., D’Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Ching Lu, C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K. D., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., ... Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, 10(2), 344–380.
- Belke, E. (2008). Effects of working memory load on lexical-semantic encoding in language production. *Psychonomic Bulletin & Review*, 15(2), 357–363.
- Belke, E. (2013). Long-lasting inhibitory semantic context effects on object naming are necessarily conceptually mediated: Implications for models of lexical-semantic encoding. *Journal of Memory and Language*, 69(3), 228–256.
- Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 58(4), 667–692.

- Blanco-Elorrieta, E., & Caramazza, A. (2021). A common selection mechanism at each linguistic level in bilingual and monolingual language production. *Cognition*, 213(February), 104625.
- Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, 129(2), 177–192.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436.
- Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, 16(2), 106–113.
- Bretz, F., Hothorn, T., & Westfall, P. (2011). Multiple comparisons using R. In *Multiple Comparisons Using R*. CRC Press.
- Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990.
- Budd, M. J., Hanley, J. R., & Griffiths, Y. (2011). Simulating children’s retrieval errors in picture-naming: A test of Foygel and Dell’s (2000) semantic/phonological model of speech production. *Journal of Memory and Language*, 64(1), 74–87.
- Bürki, A., Elbuy, S., Madec, S., & Vasishth, S. (2020). What did we learn from forty years of research on semantic interference? A Bayesian meta-analysis. *Journal of Memory and Language*, 114(April), 104125.
- Butterworth, B., Howard, D., & McLoughlin, P. (1984). The semantic deficit in aphasia: The relationship between semantic errors in auditory comprehension and picture naming. *Neuropsychologia*, 22(4), 409–426.
- Caramazza, A. (1997). How Many Levels of Processing Are There in Lexical Access? *Cognitive Neuropsychology*, 14(1), 177–208.
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113(2), 234–272.
- Costa, A., Alario, F.-X., & Caramazza, A. (2005). On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin & Review*, 12(1), 125–131.
- Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical Selection in Bilinguals: Do Words in the Bilingual’s Two Lexicons Compete for Selection? *Journal of Memory and Language*, 41(3), 365–397.
- Costa, A., Santesteban, M., & Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language*, 94(1), 94–103.
- Cree, G. S., McNorgan, C., & McRae, K. (2006). Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4), 643–658.
- Cumming, G. (2014). The New Statistics: Why and How. *Psychological Science*, 25(1), 7–29.
- Damian, M. F., & Als, L. C. (2005). Long-lasting semantic context effects in the spoken production of object names. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6), 1372–1384.
- Damian, M. F., Vigliocco, G., & Levelt, W. J. M. (2001). Effects of semantic context in the naming of pictures and words. *Cognition*, 81(3), 77–86.
<http://linkinghub.elsevier.com/retrieve/pii/S0010027701001354>
- De Zubicaray, G. I., Hansen, S., & McMahan, K. L. (2013). Differential processing of thematic and categorical conceptual relations in spoken word production. *Journal of Experimental Psychology: General*, 142(1), 131–142.

- de Zubicaray, G. I., Johnson, K., Howard, D., & McMahon, K. (2014). A perfusion fMRI investigation of thematic and categorical context effects in the spoken production of object names. *Cortex*, *54*(1), 135–149.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.
- Dell, G. S., & Kittredge, A. K. (2013). Prediction, Production, Priming, and implicit learning: A framework for psycholinguistics. In M. Sanz, I. Laka, & M. Tanenhaus (Eds.), *Language Down the Garden Path* (pp. 174–183). Oxford University Press.
- Dell, G. S., & O’Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*(1–3), 287–314.
<http://www.sciencedirect.com/science/article/pii/001002779290046K>
- Dell, G. S., Oppenheim, G. M., & Kittredge, A. K. (2008). Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. *Language and Cognitive Processes*, *23*(4), 583–608.
- Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1355–1367.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*(4), 801–838.
<http://www.ncbi.nlm.nih.gov/pubmed/9337631>
- Dylman, A. S., & Barry, C. (2018). When having two names facilitates lexical selection: Similar results in the picture-word task from translation distractors in bilinguals and synonym distractors in monolinguals. *Cognition*, *171*, 151–171.
- Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science*, *14*(2), 179–211.
- Elman, J. L., & McRae, K. (2017). A Model of Event Knowledge. *CogSci 2017 - Proceedings of the 39th Annual Meeting of the Cognitive Science Society: Computational Foundations of Cognition*, 337–342.
- Evert, S. (2008). Corpora and collocations. *Corpus Linguistics: An International Handbook*, *2*, 1212–1248.
- Goldrick, M., Folk, J. R., & Rapp, B. (2010). Mrs. Malaprop’s neighborhood: Using word errors to reveal neighborhood structure. *Journal of Memory and Language*, *62*(2), 113–134.
- Golonka, S., & Estes, Z. (2009). Thematic Relations Affect Similarity Via Commonalities. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(6), 1454–1464.
- Gordon, J. K., & Dell, G. S. (2003). Learning to divide the labor: An account of deficits in light and heavy verb production. *Cognitive Science*, *27*(1), 1–40. <http://www.uiowa.edu/delta-center/pubs/pdfs/gordon1.pdf>
- Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, *102*(3), 594–621.
<http://www.ncbi.nlm.nih.gov/pubmed/7624456>
- Hanley, J. R., Cortis, C., Budd, M. J., & Nozari, N. (2016). Did I say dog or cat? A study of semantic error detection and correction in children. *Journal of Experimental Child Psychology*, *142*, 36–47.
- Harrison, W., Hepner, C. R., & Nozari, N. (2020). Is Segmental Interference Position-dependent? *CogSci 2020*, 681–687.
- Hepner, C. R., & Nozari, N. (2020). The dual origin of lexical perseverations in aphasia: Residual activation and incremental learning. *Neuropsychologia*, *147*, 107603.

- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(April), 2554–2558.
- Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: experimental and computational studies. *Cognition*, 100(3), 464–482.
- Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of the United States of America*, 108(33), 13852–13857.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99(1), 22–44.
<http://psycnet.apa.org/index.cfm?fa=fulltext.journal&jcode=rev&vol=99&issue=1&format=html&page=22&expand=1>
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44(4), 978–990.
- La Heij, W., Dirkx, J., & Kramer, P. (1990). Categorical interference and associative priming in picture naming. *British Journal of Psychology*, 81(4), 511–525.
- Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2016). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, 18(1), 42–55.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. The MIT Press.
<http://books.google.com/books?hl=en&lr=&id=LbVCdCE-NQAC&oi=fnd&pg=PP15&dq=Speaking+:+from+intention+to+articulation&ots=JXgGUa9bzM&sig=2QfsTFI70iE0vQIK2TZh5yxrYZQ>
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(01), 1–38.
<http://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=30916>
- Levelt, W. J. M., Schreuder, R., & Hoenkamp, E. (1978). Structure and use of verbs of motion. In *Recent advances in the psychology of language* (pp. 137–162). Springer.
- Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: a network model of category learning. *Psychological Review*, 111(2), 309–332.
- Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, 7(6), 485–495. <http://doi.apa.org/?uid=1981-04769-001>
- Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3), 503–535.
- Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, 129(4), 592–613.
<http://psycnet.apa.org/index.cfm?fa=fulltext.journal&jcode=bul&vol=129&issue=4&format=html&page=592&expand=1>
- Martin, C. D., & Nozari, N. (2021). Language control in bilingual production: Insights from error rate and error type in sentence production. *Bilingualism*, 24(2), 374–388.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88(5), 375–407.
- McMillen, S., Griffin, Z. M., Peña, E. D., Bedore, L. M., & Oppenheim, G. M. (2020). “Did I Say Cherry?” Error Patterns on a Blocked Cyclic Naming Task for Bilingual Children With

























- and Without Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, 63(4), 1148–1164.
- Middleton, E. L., Schwartz, M. F., Dell, G. S., & Brecher, A. (2022). Learning from errors: Exploration of the monitoring learning effect. *Cognition*, 224, 105057.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Nips*, 1–9.
- Movellan, J. (1990). Contrastive Hebbian learning in the continuous Hopfield model. *Proceedings of the 1990 Connectionist Models Summer School*, 10–17. <http://mplab.ucsd.edu/~movellan/mypapers/Movellan91.pdf>
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92(3), 289–316. <http://psycnet.apa.org/journals/rev/92/3/289.html>
- Navarrete, E., Del Prato, P., Peressotti, F., & Mahon, B. Z. (2014). Lexical selection is not by competition: Evidence from the blocked naming paradigm. *Journal of Memory and Language*, 76, 253–272.
- Navarrete, E., Mahon, B. Z., & Caramazza, A. (2010). The cumulative semantic cost does not reflect lexical selection by competition. *Acta Psychologica*, 134(3), 279–289.
- Norman, K. a, Newman, E. L., & Detre, G. J. (2007). A neural network model of retrieval-induced forgetting. *Psychological Review*, 114(4), 887–953.
- Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, 63(1), 1–33.
- Nozari, N., Freund, M., Breining, B. L., Rapp, B., & Gordon, B. (2016). Cognitive control during selection and repair in word production. *Language, Cognition and Neuroscience*, 31(7), 886–903.
- Nozari, N., & Hepner, C. R. (2019). To select or to wait? The importance of criterion setting in debates of competitive lexical selection. *Cognitive Neuropsychology*, 36(5–6), 193–207.
- Nozari, N., Martin, C. D., & McCloskey, N. (2019). Is repairing speech errors an automatic or a controlled process? Insights from the relationship between error and repair probabilities in English and Spanish. *Language, Cognition and Neuroscience*, 34(9), 1230–1245.
- Nozari, N., & Pinet, S. (2020). A critical review of the behavioral, neuroimaging, and electrophysiological studies of co-activation of representations during word production. *Journal of Neurolinguistics*, 53(April 2019), 100875.
- Oppenheim, G. M. (2011). Incremental lexical learning in speech production: a computational model and empirical evaluation. In *Doctoral thesis*. University of Illinois at Urbana-Champaign.
- Oppenheim, G. M. (2018). The paca that roared : Immediate cumulative semantic interference among newly acquired words. *Cognition*, 177(August 2018), 21–29.
- Oppenheim, G. M. (2024). Lexical selection in language production [Manuscript submitted for publication]. PsyArXiv. <https://doi.org/10.31234/osf.io/k2s3p>
- Oppenheim, G. M., & Balatsou, E. (2019). Lexical competition on demand. *Cognitive Neuropsychology*, 36(5–6), 216–219.
- Oppenheim, G. M., Balatsou, E., & Thierry, G. *Endogenous conflict and exogenous competition in word production: an EEG study of name agreement in overt picture naming* [Manuscript submitted for publication].
- Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: a model of cumulative semantic interference during lexical access in speech production.

- Cognition*, 114(2), 227–252.
- Oppenheim, G. M., Griffin, Z. M., Peña, E. D., & Bedore, L. M. (2020). Longitudinal Evidence for Simultaneous Bilingual Language Development With Shifting Language Dominance, and How to Explain It. *Language Learning*, 70(June 2020), 20–44.
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12), 976–987.
- Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(3), 539–557.
- Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, 77(2), 97–132.
<http://linkinghub.elsevier.com/retrieve/pii/S0010027700000901>
- Recchia, G., & Jones, M. N. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis. *Behavior Research Methods*, 41(3), 647–656.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variation in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning: II. Theory and research* (pp. 64–99). Appleton-Century-Crofts.
- Roelofs, A. (1992). A spreading activation theory of lemma retrieval in speaking. *Cognition*, 42(1–3), 107–142. <http://www.sciencedirect.com/science/article/pii/001002779290041F>
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573–605.
<http://linkinghub.elsevier.com/retrieve/pii/0010028575900249>
- Rose, S. B., & Abdel Rahman, R. (2016). Cumulative semantic interference for associative relations in language production. *Cognition*, 152, 20–31.
- Rose, S. B., Aristei, S., Melinger, A., & Abdel Rahman, R. (2019). The closer they are, the more they interfere: Semantic similarity of word distractors increases competition in language production. *Journal of Experimental Psychology: Learning Memory and Cognition*, 45(4), 753–763.
- Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Vol. 1. Foundations*. MIT Press.
- Schade, U., & Berg, T. (1992). The role of inhibition in a spreading-activation model of language production. II. The simulational perspective. *Journal of Psycholinguistic Research*, 21(6), 435–462.
- Schnur, T. T. (2014). The persistence of cumulative semantic interference during naming. *Journal of Memory and Language*, 75, 27–44.
- Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, 54(2), 199–227.
- Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, 29(1), 86–102.
- Schwartz, M. F., & Kimberg, D. Y. (2011). Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proceedings of the National Academy of Sciences*

- of the United States of America, 108(20), 8520–8524.
- Starreveld, P. A., & La Heij, W. (1995). Semantic Interference, Orthographic Facilitation, and Their Interaction in Naming Tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(3), 686–698.
- Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, 24(4), 629–640.
- Tourville, J. a., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26(October 2011), 952–981.
- Tversky, A. (1977). Features of Similarity. *Psychological Review*, 84(4), 327–352.
- Vigliocco, G., Vinson, D. P., Damian, M. F., & Levelt, W. J. M. (2002). Semantic distance effects on object and action naming. *Cognition*, 85(3), B61–B69.
- Vitkovitch, M., Batoool Ghadiri, T., & Hersi, F. (2019). Word naming slows picture naming but does not affect cumulative semantic interference. *Acta Psychologica*, 195, 30–38.
- Vitkovitch, M., & Humphreys, G. W. (1991). Perseverant responding in speeded naming of pictures: It's in the links. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(4), 664–680. <http://psycnet.apa.org/journals/xlm/17/4/664.html>
- Wheeldon, L. R., & Monsell, S. (1994). Inhibition of Spoken Word Production by Priming a Semantic Competitor. *Journal of Memory and Language*, 33(3), 332–356.
- Widrow, B., & Hoff, M. E. (1960). *Adaptive switching circuits*.
- Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin and Review*, 23(4), 1015–1027.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly INC: Beijing. <http://www.nltk.org/book/>
- Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glott International* 5:9/10, 341-345. <https://www.fon.hum.uva.nl/praat/>
- Miller, G. A. (1998). *WordNet: An electronic lexical database*. MIT press. <https://wordnet.princeton.edu/>
- Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: do all roads lead to Rome? *Cognitive Psychology*, 32(1), 49–96.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92(3), 289–316. <http://psycnet.apa.org/journals/rev/92/3/289.html>
- Scaltritti, M., Peressotti, F., & Navarrete, E. (2017). A joint investigation of semantic facilitation and semantic interference in continuous naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(5), 818.

Appendix A

Stimulus Images

Target	Taxonomically Related	Thematically Related	Unrelated
			
			
			
			
			
			

Appendix B

Full results of the analyses

Experiment 1 - RT analyses

Table B1.1. Three comparisons of interest on all words (Tukey-corrected p-values)

Effect	Estimate	SE	z	p-value
Thematic - Taxonomic	-0.024	0.005	-4.525	< .001
Unrelated - Taxonomic	-0.021	0.005	-3.844	< .001
Unrelated - Thematic	0.004	0.005	0.682	0.774

Table B1.2. Three comparisons of interest on target words only (Tukey-corrected p-values)

Effect	Estimate	SE	z	p-value
Thematic - Taxonomic	-0.025	0.005	-4.525	< .001
Unrelated - Taxonomic	-0.021	0.005	-3.912	< .001
Unrelated - Thematic	0.003	0.005	0.615	0.812

Table B1.3. Testing the effect of repeating/switching on the critical contrasts

Effect	Estimate	SE	t	p-value
Intercept	6.101	0.018	344.54	< .001
Unrelated - Taxonomic	-0.019	0.007	-2.607	0.009
Thematic - Taxonomic	-0.03	0.007	-4.076	< .001
Repetition	-0.026	0.006	-4.584	< .001
Unrelated - Taxonomic x Repetition	-0.003	0.008	-0.361	0.718
Thematic - Taxonomic x Repetition	0.004	0.008	0.514	0.607

Experiment 1 - Duration analyses

Table B1.4. Three comparisons of interest on all words (Tukey-corrected p-values)

Effect	Estimate	SE	z	p-value
Thematic - Taxonomic	0.006	0.005	1.056	0.541
Unrelated - Taxonomic	0.004	0.005	0.643	0.796
Unrelated - Thematic	-0.002	0.005	-0.414	0.91

Table B1.5. Three comparisons of interest on target words only (Tukey-corrected p-values)

Effect	Estimate	SE	z	p-value
Thematic - Taxonomic	0.006	0.005	1.109	0.509
Unrelated - Taxonomic	0.004	0.005	0.71	0.758
Unrelated - Thematic	-0.002	0.005	-0.399	0.916

Table B1.6. Testing the effect of repeating/switching on the critical contrasts

Effect	Estimate	SE	t	p-value
Intercept	5.752	0.033	174.295	< .001
Unrelated - Taxonomic	0.001	0.008	0.182	0.856
Thematic - Taxonomic	0.002	0.008	0.209	0.835
Repetition	0.001	0.006	0.246	0.806
Unrelated - Taxonomic x Repetition	0.005	0.008	0.574	0.566
Thematic - Taxonomic x Repetition	0.007	0.008	0.866	0.386

Supplement for Similarity-induced interference or facilitation in language production reflects representation, not selection

Gary M. Oppenheim^{1,2}, Nazbanou Nozari^{3,4}

¹ Department of Psychology, Bangor University, Bangor, Wales, UK

² Department of Psychology, The University of Texas at Austin, USA

³ Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

⁴ Cognitive Science Program, Indiana University, Bloomington, IN, USA

Abstract

Researchers have long interpreted the presence or absence of semantic interference in picture naming latencies as confirming or refuting theoretical claims regarding competitive lexical selection. But inconsistent empirical results challenge any mechanistic interpretation. Experiment 1, reported in the main text, verified a boundary condition in a blocked picture naming task: when orthogonally manipulating association type, taxonomic associations consistently elicit interference, while thematic associations do not. Experiment 2, reported in this supplement, uses a word-picture matching version of the same task to strengthen the case that this pattern of results is specific to word production.

Experiment 2

Participants

The sample size was similar to Experiment 1 for consistency. Sixty native English-speaking adults ($M_{age}= 20.42$, $SD=1.80$; 32 female) participated in the study for course credit or payment. Recruitment took place online, from Carnegie Mellon University's undergraduate subject pool and Amazon Mechanical Turk. Consent was obtained under a protocol approved by the Carnegie Mellon Institutional Review Board.

Materials

The same materials as Experiment 1 were used in Experiment 2. To convert the task into a word-to-picture matching task, 24 audio stimuli, corresponding to the labels of the 24 pictures, were created via text-to-speech voice cloning using an Overdub stock voice in the Descript audio editing software (<https://www.descript.com>). Each audio file began with 100ms of silence and ended immediately after the exact length of a particular word. The average length of audio files was matched across words in all conditions (Target: $M = 464$ ms, $SD = 35$ ms; Taxonomic competitor:

$M = 464$ ms, $SD = 73$ ms; Thematic competitor: $M = 482$ ms, $SD = 80$ ms, Unrelated competitor: $M = 472$ ms, $SD = 64$ ms).

Similar to Experiment 1, 18 blocks were created (6 in each condition) with a pair of target and competitor pictures and their corresponding audio files in each block. Half of the trials in each block were “match” trials (picture and sound file corresponding to the same item), and the other half, the “mismatch” trials (picture and sound file corresponding to different items). Half of the match trials were on the target item, and the other half on the competitor. Three pseudo-randomized orders were created, where the same item was not repeated more than three times in a row within a block (as audio or as image) and the same target never appeared in adjacent blocks. The orders were counterbalanced across conditions, creating three different lists.

Procedures

The experiment was programmed in jsPsych (de Leeuw, 2015) with JavaScript and hosted on JATOS (Lange et al., 2015) and psiTurk (Gureckis et al., 2016). Participants first completed an orientation and practice phase, using a picture pair that was not part of the experimental materials (a fork and a shoe), to learn how to respond in the word-to-picture task. On each trial, they heard a word followed quickly by a picture, and were instructed to press one of the two buttons (G or H, counterbalanced across participants) to indicate whether the trial was a match or a mismatch. The central G and H keys were selected so that participants could use the index and middle fingers of their dominant hand for responding, regardless of whether they were left-handed or right-handed. After the instructions and four self-paced trials, participants completed an additional 10 practice trials with a deadline of 1500 ms and received feedback after each trial. If their accuracy was lower than 60% after 10 trials, practice was repeated in the same manner to ensure that the mapping between the keys and the match/mismatch trials was well learned before moving on to the experimental phase.

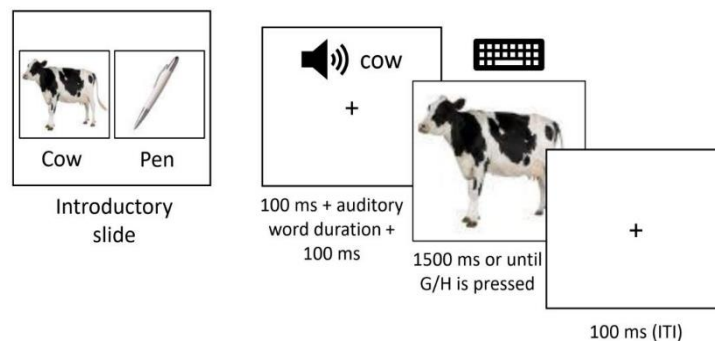


Figure S8. Example trials from Exp 1 (production; panel a) and Exp 2 (comprehension; panel b). Panel (a) shows the block structure in the production experiment. Participants see the picture pair and their labels in an introductory slide. After four practice trials (not shown), they name each of the two pictures six times in pseudorandomized order. The block structure is generally similar in the comprehension experiment (Panel b), except that participants hear a word, followed by a picture, and indicate the match/mismatch between the two by pressing a button the keyboard.

In the experimental phase, participants were assigned to one of three block orders. In each of the 18 blocks, they were first presented with the picture pair in that block along with their labels. In the next two practice trials, participants first heard the audio file, followed 100 ms later by a picture which remained on the screen for 1500 ms or until G or H keys were pressed. If an incorrect or no response was made on the practice trials, corrective feedback was provided, and the next trial began after an ITI of 100ms. Participants then completed 12 experimental—6 presentations of each picture—with the same trial structure but without feedback. The only exception was three consecutive trials with no response, which elicited a warning to “Please respond as quickly as possible”.

Coding and Analyses

Analyses were conducted in R (version 4.0.3; R Core Team, 2020). Data were analyzed by linear mixed effects models (LMEs) using the lme4 package (version 1.1-27.1), in conjunction with lmerTest (version 3.1-3) to calculate p-values using Satterthwaite approximations. For the analyses that compared multiple non-orthogonal contrasts, the multcomp R package version 1.4-16 (Bretz et al., 2011) was used, which returns Tukey-corrected p-values for pairwise comparisons. Values more extreme than 3SD from the mean of RT and duration distributions were excluded from the analyses. RT distributions were right-skewed and were therefore log-transformed to approximate normality. The random effect structure was kept maximal according to the recommendation by Barr et al. (2013), unless the model did not converge; in such cases, the random effect structure was reduced by first eliminating the slopes for items and subjects. As most models did not tolerate random slopes, we included the random intercept of subjects and items in all models for consistency.

As in Experiment 1, we first tested the three critical contrasts of interest (taxonomic vs. unrelated, thematic vs. unrelated, and taxonomic vs. thematic) on the whole dataset. Since distractors are carefully matched between conditions, we would expect differences to reflect the relations between items, as opposed to the individual properties of items. Next, to ensure that individual items’ properties did not contaminate the results, we ran the same model only on a subset of the data containing the target items. In other words, in the second analysis the same items are compared across the three conditions. Both of these models were run using the multcomp package, which returns Tukey-corrected p-values for multiple non-orthogonal contrasts, but does not tolerate interaction terms in the model. To double check the results of these models on the two most critical contrasts (taxonomic vs. unrelated, i.e., the replication of the semantic blocking effect, and taxonomic vs. thematic, i.e., the key difference under investigation), we ran a third model containing possible nuisance variables and their interactions with the critical contrasts. Because the paradigm contains frequent repetitions and switches between two items, we expect a canonical switch cost (e.g., Braver, 2012), with switch trials having longer RTs than repeat trials. To ensure that the switch cost did not alter the effect of semantic relations, a repetition variable (repeat vs. switch, coded based on whether the target was the same as the previous trial or not) was added to the check the models. Additionally, responses in Experiment 2 were in the form of “yes” (match)/”no”(mismatch). Since “match” responses are known to be faster than “mismatch” responses in such tasks (i.e., the ‘fast same effect’, e.g., Bamber, 1969; Farell, 1985; Stadthagen-Gonzalez et al., 2009), this canonical effect was modeled by adding the match variable (match vs. mismatch) to the check model in Experiment 2.

The data and the analysis code are available online (<https://osf.io/qrg72/>). The full results of these analyses are reported in Appendix SA.

Results

The error rate was much higher in Experiment 2 (884) than it had been in Experiment 1 but was comparable between the three conditions (7% of responses in each condition). To verify that there was no significant differences between the error rates, a logistic version of LMEM was run with accuracy as the DV, condition with three pairwise contrasts as IV, and random effects of subjects and items. The Tukey-corrected p-values suggested that there were no significant differences between conditions ($p > 0.80$ for all comparisons; Table SA1.1). The error responses were excluded from further analysis. In addition, extreme values were also excluded (2% of the durations for correct trials).

Figure S2 shows the pattern of RTs in this experiment. As can be seen, the pattern is quite different from Experiment 1. The first-pass analysis revealed no significant differences in any of the comparisons (taxonomic vs. unrelated: $\beta < .001 \sim -0.1$ ms when back-transformed, $t = -0.02$, $p \approx 1$; thematic vs. unrelated: $\beta = -0.011 \sim -5.7$ ms, $t = -1.32$, $p = 0.38$; taxonomic vs. thematic: $\beta = 0.011 \sim -5.7$ ms, $t = 1.30$, $p = 0.40$; Table SA1.2). The model including only target words returned a similar pattern of results with no significant effect for any of the pairwise comparisons (Table SA1.3).

Finally, the last model tested the effect of repetition, as well as response match on the critical contrasts from Experiment 1. Table S1 shows the RTs relevant to this analysis. The model included condition (contrast coded as taxonomic vs. unrelated and taxonomic vs. thematic), repetition (repeat or switch), and match (match or mismatch) and all the two- and three-way interactions between these factors, as its fixed-effect structure. The random effect structure was the same as the earlier models. As expected, RTs were significantly shorter on match compared to non-match trials ($\beta = -0.17 \sim -88.7$ ms, $t = -32.92$, $p < .001$). Additionally, there was a main effect of repetition, as well as a significant interaction between match and repetition ($\beta = 0.062 \sim 33.0$ ms, $t = 6.074$, $p < .001$), such that RTs were shorter for repeats vs. switches on match trials, but the opposite was true for mismatch trials. There was also a marginal three-way interaction between repetition, match, and the contrast representing the difference between taxonomic and thematic conditions ($\beta = 0.043 \sim 23.2$ ms, $t = 1.73$, $p = 0.08$). This interaction reflects the fact that the RT differences between match and mismatch trials was 25ms larger for thematic compared to taxonomic conditions on repeat trials, but 4ms smaller on switch trials (see Table S1). No other effect, and importantly, neither of the critical contrasts, approached significance in the model (Table SA1.4). To summarize, there was no significant evidence of taxonomic interference in comprehension, nor of an attenuated or reversed interference effect for thematic relations.

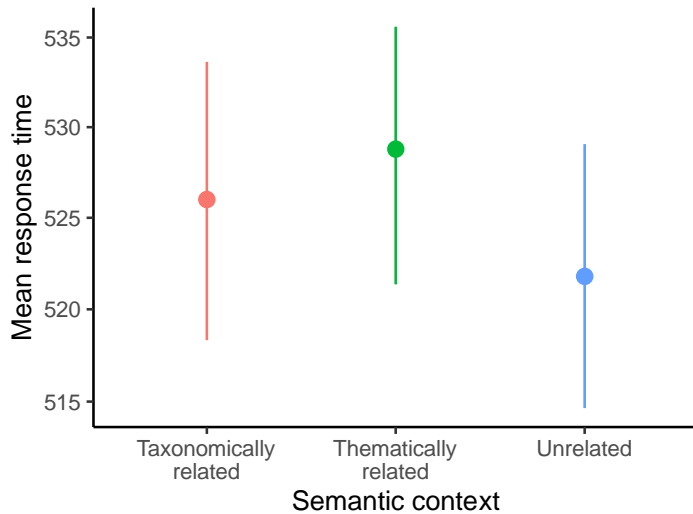


Figure S9. Mean word-picture matching times for target words in Experiment 2 (Panel B: back-transformed inverse scale) as a function of competitor type. Error bars represent bootstrapped confidence intervals.

Table S1. Mean RTs (SE) in the three conditions in Experiment 2, broken down by repetition/switching and match/mismatch responses.

Repetition	Match	Condition		
Repeats		Taxonomic	Thematic	Unrelated
	match	511 (22)	504 (22)	505 (23)
	mismatch	606 (23)	624 (26)	615 (26)
Switches	match	510 (23)	522 (22)	510 (23)
	mismatch	575 (21)	583 (23)	569 (21)

Discussion

Experiment 2 adapted the methods and stimuli from Experiment 1 to assess corresponding effects in word comprehension, via a word-picture matching task. There was no evidence of attenuated interference effects for thematic relations in the matched comprehension task, suggesting that the contrasting effects of taxonomic and thematic relations that we reported in Experiment 1 are indeed specific to word production.

References

- Bamber, D. (1969). Reaction times and error rates for "same"- "different" judgments of multidimensional stimuli. *Perception & Psychophysics*, 6(3), 169–174.
- 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.
- Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, 16(2), 106–113.
- Bretz, F., Hothorn, T., & Westfall, P. (2011). Multiple comparisons using R. In *Multiple Comparisons Using R*. CRC Press.
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12.
- Farell, B. (1985). “Same”-“Different” Judgments. A Review of Current Controversies in Perceptual Comparisons. *Psychological Bulletin*, 98(3), 419–456.
- Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., Halpern, D., Hamrick, J. B., & Chan, P. (2016). psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behavior Research Methods*, 48(3), 829–842.
- Lange, K., Kühn, S., & Filevich, E. (2015). “Just another tool for online studies” (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLoS ONE*, 10(6), 1–14.
- Stadthagen-Gonzalez, H., Damian, M. F., Pérez, M. A., Bowers, J. S., & Marín, J. (2009). Name-picture verification as a control measure for object naming: A task analysis and norms for a large set of pictures. *Quarterly Journal of Experimental Psychology*, 62(8), 1581–1597.

Appendix SA

Full results of the analyses for Experiment 2

Experiment 2 – Accuracy analysis

Table SA1.1. Three comparisons of interest on all words (Tukey-corrected p-values)

Effect	Estimate	SE	z	p-value
Thematic - Taxonomic	-0.057	0.11	-0.524	0.86
Unrelated - Taxonomic	-0.063	0.109	-0.579	0.831
Unrelated - Thematic	-0.006	0.11	-0.052	0.998

Experiment 2 - RT analyses

Table SA1.2. Three comparisons of interest on all words (Tukey-corrected p-values)

Effect	Estimate	SE	z	p-value
Thematic - Taxonomic	0.011	0.008	1.296	0.397
Unrelated - Taxonomic	< .001	0.008	-0.024	1
Unrelated - Thematic	-0.011	0.008	-1.32	0.384

Table SA1.3. Three comparisons of interest on target words only (Tukey-corrected p-values)

Effect	Estimate	SE	z	p-value
Thematic - Taxonomic	0.006	0.011	0.537	0.853
Unrelated - Taxonomic	-0.007	0.011	-0.654	0.79

Unrelated - Thematic	-0.013	0.011	-1.19	0.459
----------------------	--------	-------	-------	-------

Table SA1.4. Testing the effect of repeating/switching and match/mismatch on the critical contrasts

Effect	Estimate	SE	t	p-value
Intercept	6.283	0.023	278.32	< .001
Match	-0.166	0.005	-32.916	< .001
Repetition	-0.032	0.005	-5.891	< .001
Unrelated - Taxonomic	-0.003	0.008	-0.328	0.743
Thematic - Taxonomic	0.011	0.008	1.397	0.163
Match x Repetition	0.062	0.01	6.074	< .001
Unrelated - Taxonomic x Match	-0.001	0.012	-0.12	0.905
Thematic - Taxonomic x Match	-0.004	0.012	-0.285	0.776
Unrelated - Taxonomic x Repetition	0.007	0.013	0.52	0.603
Thematic - Taxonomic x Repetition	0.01	0.013	0.777	0.437
Unrelated - Taxonomic x Match x Repetition	0.031	0.025	1.249	0.212
Thematic - Taxonomic x Match x Repetition	0.043	0.025	1.727	0.084