



The role of word frequency and morpho-orthography in agreement processing

Laurel Brehm, Erika Hussey & Kiel Christianson

To cite this article: Laurel Brehm, Erika Hussey & Kiel Christianson (2019): The role of word frequency and morpho-orthography in agreement processing, Language, Cognition and Neuroscience, DOI: [10.1080/23273798.2019.1631456](https://doi.org/10.1080/23273798.2019.1631456)

To link to this article: <https://doi.org/10.1080/23273798.2019.1631456>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 27 Jun 2019.



[Submit your article to this journal](#)



Article views: 105



[View Crossmark data](#)

The role of word frequency and morpho-orthography in agreement processing

Laurel Brehm^{a,c}, Erika Hussey^{b,c} and Kiel Christianson^{c,d}

^aMax Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; ^bUS Army Natick Soldier Research, Development, and Engineering Center, Natick, MA, USA; ^cBeckman Institute for Advanced Science and Technology, Urbana, IL, USA; ^dDepartment of Educational Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, USA

ABSTRACT

Agreement attraction in comprehension (when an ungrammatical verb is read quickly if preceded by a feature-matching local noun) is well described by a cue-based retrieval framework. This suggests a role for lexical retrieval in attraction. To examine this, we manipulated two probabilistic factors known to affect lexical retrieval: local noun word frequency and morpho-orthography (agreement morphology realised with or without *-s* endings) in a self-paced reading study. Noun number and word frequency affected noun and verb region reading times, with higher-frequency words not eliciting attraction. Morpho-orthography impacted verb processing but not attraction: atypical plurals led to slower verb reading times regardless of verb number. Exploratory individual difference analyses further underscore the importance of lexical retrieval dynamics in sentence processing. This provides evidence that agreement operates via a cue-based retrieval mechanism over lexical representations that vary in their strength and association to number features.

ARTICLE HISTORY

Received 22 February 2018
Accepted 28 May 2019

KEYWORDS

Sentence processing; number agreement; lexical retrieval; vocabulary; individual differences

Introduction

Number is expressed in English on nouns, verbs, pronouns, and determiners. Expressing number on these words makes it clear which nouns control which verbs, allowing a comprehender to easily establish syntactic dependencies in sentences. The typical pattern in English is that plural nouns receive an *-s* ending, such that the plural of *skirt* is *skirts*; however, some nouns mismatch this pattern. For example, some words, like *cactus* or *dress*, match the typical plural pattern (*-s*) when singular and some words, like *men* and *cacti*, match the typical singular pattern (no *-s*) when plural. These morphological patterns further co-vary with word frequency, such that atypical plurals like *men* are frequent, while atypical plurals like *cacti* are infrequent. Both factors could lead to trouble with agreement, such that probabilistic co-occurrences between nouns and verbs and between noun spelling and verb inflection might cause infrequent, atypically spelled singular nouns like *cactus* to be mis-read as plural. In the current study, we investigate how morpho-orthography and word frequency impact number agreement processing. This showcases the lexical properties influencing the processing of subject-verb dependencies and in turn sheds light on the

underlying mechanisms of sentence processing and the representations these mechanisms operate over.

Cue-based retrieval mechanisms in agreement processing

In order to process agreement, a reader must retrieve and maintain lexical items, making a guess as to what noun will be the subject of a downstream verb from partial, continuously changing information. Though this processing is typically successful, mishaps occur. One common mishap is *agreement attraction*, where properties of an intervening but grammatically irrelevant noun influence readers' processing speed and acceptability judgments, as well as speakers' production difficulty. An example of this appears in sentences 1 and 2. When sentences contain an ungrammatical verb (*were*) that matches in number with a *local* (non-subject) noun (*cabinets*, 1a), there is reduced processing difficulty compared to sentences where the local noun does not match the ungrammatical verb (1b; Pearlmuter, Garnsey, & Bock, 1999; Wagers, Lau, & Phillips, 2009). This means that ungrammatical verbs paired with a feature-matching local noun elicit a processing

CONTACT Laurel Brehm  laurel.brehm@mpi.nl, laurel.brehm@gmail.com

 Supplemental data for this article can be accessed <https://doi.org/10.1080/23273798.2019.1631456>.

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

pattern more similar to a comparable grammatical sentence (2a vs. 2b).

(1a) *The key to the cabinets were rusty.

(1b) *The key to the cabinet were rusty.

(2a) The key to the cabinets was rusty.

(2b) The key to the cabinet was rusty.

The cue-based retrieval framework provides a set of mechanisms to account for agreement attraction in ungrammatical sentences (Lewis & Vasishth, 2005; McElree, Foraker, & Dyer, 2003; Van Dyke & McElree, 2011; for reviews, see Lewis, Vasishth, & Van Dyke, 2006; and Van Dyke & Johns, 2012; for meta-analysis, see Jäger, Engelmann, & Vasishth, 2017). The premise is that retrieval cues allow the reader to establish syntactic dependencies between words in sentences; if retrieval cues are weak, processing is more difficult and interference is more likely to occur during retrieval (e.g. Van Dyke, Johns, & Kukona, 2014).

For agreement, this means that encountering an unpredicted ungrammatical verb triggers a memory search for a noun that is a plausible controller for the verb. When a feature-matching *non*-subject noun is retrieved instead of the subject, a sentence can appear to be grammatical when it is not, eliciting a reduced penalty for ungrammatical verbs when an attractor is present (e.g. Dillon, Mishler, Sloggett, & Phillips, 2013; Lago, Shalom, Sigman, Lau, & Phillips, 2015; Martin, Nieuwland, & Carreiras, 2012; Tanner, Nicol, & Brehm, 2014; Wagers et al., 2009). This means that retrieval dynamics elegantly capture much of the data on how grammatical number affects number agreement.

Probabilistic factors in agreement

However, there is also clear evidence for probabilistic factors distinct from grammatical number impacting number agreement production. This has led to the development of probabilistic agreement models, providing an alternate theoretical framework accounting for attraction (e.g. Haskell, Thornton, & MacDonald, 2010; Mirković & MacDonald, 2013; Smith, Franck, & Tabor, 2018). Probabilistic factors known to impact agreement production include syntactic properties of the language, such as the degree of inflectional morphology (Foote & Bock, 2012; Franck, Vigliocco, Antón-Méndez, Collina, & Frauenfelder, 2008) and the reliability of inflectional morphology (Mirković & MacDonald, 2013). They also include conceptual properties, such as the likelihood of a number cue matching the number concept (e.g. Brehm & Bock, 2013, 2017; Eberhard, 1999; Haskell & MacDonald, 2003; Haskell et al., 2010; Humphreys & Bock,

2005; Mirković & MacDonald, 2013; Smith et al., 2018; Vigliocco, Butterworth, & Garrett, 1996), and morphophonological properties, such as the transparency of number marking in spoken forms (Haskell & MacDonald, 2003; Lorimor, Jackson, Spalek, & van Hell, 2016). These probabilistic cues tend to have small effects in isolation, often only reaching significance in combination with each other: in particular, effects of morphophonology and number marking are shown in combination with conceptual number.

It is also the case that other types of interference occur in processing: interference also occurs during encoding (Villata, Tabor, & Franck, 2018), such that more recent elements can overwrite older ones when they have similar semantic or morphological features. Encoding interference is not captured in classic cue-based retrieval model. However, as Villata et al. (2018) demonstrate, encoding interference and other similarity effects arise naturally as a consequence of lexically-rooted, probabilistic structure assembly (SOSP) models, providing strong evidence that probabilistic factors are likely to be an essential piece of agreement comprehension.

While probabilistic models have to date mainly gained traction in agreement production research, we also note that what is produced has consequences for what is observed. This means that when a factor impacts agreement production, it also increases the likelihood of observing an error in comprehension. As such, a goal of the current study is to assess whether there is evidence for probabilistic factors in agreement *comprehension*, establishing a link at the lexical level between agreement in production and comprehension. This situates the current work within larger theoretical frameworks integrating language production and comprehension (e.g. Dell & Chang, 2014; Pickering & Garrod, 2013).

Linking cue-based retrieval and probabilistic agreement: the role of lexical retrieval

The larger goal of the current study is to assess the underlying mechanisms and representations behind agreement processing. There is a potential mechanistic link between cue-based retrieval and the mechanisms by which probabilistic factors affect language processing, as noted by Lorimor and colleagues (Lorimor et al., 2016) for agreement production and by Van Dyke and colleagues (Van Dyke et al., 2014; Van Dyke & Johns, 2012) for comprehension. The critical dimension is lexical retrieval itself.

As reviewed in the first section, lexical retrieval is the primary driver of agreement attraction in a cue-based

retrieval framework. Ungrammatical verbs trigger the need to retrieve nouns from earlier in the sentence that have a matching plural feature. It is also the case that lexical retrieval is affected by the same sorts of probabilistic factors that impact agreement production. Work on single-word production has carefully examined the factors that impact lexical retrieval, which include number morphology (e.g. Baayen, Dijkstra, & Schreuder, 1997; Schreuder & Baayen, 1997) and semantics (e.g. Abdel Rahman & Melinger, 2009; Damian, Vigliocco, & Levelt, 2001; see Strijkers & Costa, 2011, for a recent review of findings from multiple paradigms). Since these factors also influence agreement, the implication is that probabilistic variations in retrieval might provide coverage for a variety of findings within agreement, linking cue-based retrieval with probabilistic models of agreement. Investigating the role of lexical properties in agreement therefore extends and links these theoretical models of psycholinguistics.

We operationalised lexical retrieval difficulty in the current study with two probabilistic factors. The first factor was word frequency, manipulated by the frequency of inflected nouns: Infrequent words tend to be more difficult to retrieve (e.g. Bates et al., 2003; Jescheniak & Levelt, 1994). The second factor was morpho-orthography, as manipulated based upon the presence of *-s* endings crossed with plural morphology: Infrequently observed spelling patterns make lexical retrieval more difficult (e.g. Andrews, 1997; Coltheart, Davelaar, Jonasson, & Besner, 1977; Grainger, O'Regan, Jacobs, & Segui, 1989; Snodgrass & Mintzer, 1993).

For agreement, there is also a probabilistic relationship between spelling and number inflection, such that words that end in *-s* are typically plural, with the *-s* affix serving as the regular plural marker (*skirts, boys, yuccas*; but c.f. *dress, cactus*), and such that words with no *-s* ending are typically singular (*skirt, boy, yucca*; but c.f. *men, cacti*). Therefore, the presence of a misleading *-s* on local nouns could have an impact on attraction. Because it appears frequently as an independent morpheme, readers may aim to decompose the *-s* affix even when this is not licensed (e.g. as outlined in Rastle & Davis, 2008), leading to attraction effects. Similarly, the *-s* ending itself may have a direct but probabilistic impact on agreement processing (see, e.g. reduced attraction production for irregular *-s* absent plurals but only when paired with other cues; Haskell & MacDonald, 2003).

Morphology and frequency also tend to co-vary in English, such that irregular plurals are often extremely frequent (*men*) or extremely infrequent (*cacti*). This may have a strong impact on agreement, with interactive effects such that frequency effects are stronger for

irregular words (e.g. Allen, Badecker, & Osterhout, 2003) and lower-frequency words are more likely to be stored in a decomposed fashion (e.g. Alegre & Gordon, 1999; Baayen et al., 1997). This means that morpho-orthography and word frequency are ideal to manipulate in tandem.

We assess the dynamics of lexical retrieval in agreement by examining processing effects separately on regions that relate to noun and verb processing. This is a way in which the present study differs from many previous studies of agreement attraction, which tend to focus on processing costs on the verb region only. This analysis approach allows us to observe whether there are differences in how lexical retrieval related factors affect processing of nouns and copular verbs. Processing both nouns and verbs is likely to require lexical retrieval, but the representations that retrieval operates upon could differ. Evidence for this claim comes from the fact that different factors affect reading times in the noun and verb regions of sentences (see Wagers et al., 2009). Consistent with increased lexical retrieval difficulty, plural nouns tend to elicit slower reading times as a main effect at the noun and following word. However, the effect disappears at later regions, such that reading times are slower for only ungrammatical verbs that are preceded by a local singular noun. This shows a dissociation between what affects noun processing (number marking) and verb processing (the presence of a plural in memory once retrieval has been triggered); the question raised in the current work is whether the same set of probabilistic factors affect both sentence regions. Answering this question shows whether lexical retrieval affects noun and verb processing equivalently, showcasing the way memory retrieval supports sentence processing.

Current study

In the present study, we examine whether word frequency and morpho-orthography, two factors that can hinder lexical retrieval, also impact subject-verb number agreement and the preceding processing of a local noun. We manipulated the frequency, morpho-orthography, and number of local nouns and the grammaticality of the following verb embedded in attraction-inducing sentences like “*The landscaper who planted the cactus already was/*were anticipating the dry summer*” (critical local noun and verb underlined, see Methods for more details) to test whether probabilistic retrieval cues modulate attraction. Local nouns were varied in their morpho-orthography (e.g. *man/men; dress/dresses; cactus/cacti*) and compared with

frequency-matched typical morphology count nouns to isolate the role of frequency from the role of morpho-orthography in comprehension.

Predicted differences in the sentence's verb region can dissociate the relative roles of cue-based retrieval and probabilistic factors in agreement. A hypothesis derived from a strict, non-probabilistic cue-based retrieval framework is that only *number* will impact agreement. The cue-based retrieval framework suggests that agreement attraction occurs due to the retrieval of a plural noun (or plural feature), irrespective of how number was instantiated. As such, we predict that ungrammatical plural verbs preceded by a local singular noun will increase reading times versus local plural nouns. This is the standard attraction pattern.

In contrast, if a reader uses any and all cues for number agreement, as suggested by a strong probabilistic model of agreement, an alternate hypothesis is that *noun morpho-orthography* will also lead to differences in verb reading time such that *-s* endings will increase attraction, with less attraction observed for *men* and *cacti* type items than matched controls. Such a pattern would suggest that probabilistic cues such as *-s* endings can lead to the perception of grammaticality. In the case that the quality of lexical representations leads to probabilistic agreement difficulty (as suggested by Van Dyke et al., 2014; Van Dyke & Johns, 2012), a third prediction is that *infrequent nouns* should lead to increased attraction, such that the most attraction is observed for the least frequent items. Such a pattern would suggest that the strength of noun representations impacts agreement.

In contrast to the varied hypotheses for the verb region, the noun region predictions are relatively simple: processing time in the noun region is likely to be driven by retrieval difficulty.¹ It is predicted that frequency, morpho-orthography, and number will all impact reading times on the local noun, given their established link to lexical retrieval. Previous work has shown that plural nouns require more processing time than singular nouns (see Wagers et al., 2009). We predict that this will also be the case for infrequent nouns, which may have a more weakly specified lexical representation, and nouns with misleading morpho-orthography (*-s* ending in singular form, like *dress* and *cactus*; absence of *-s* ending in plural, like *men* or *cacti*); both properties affect lexical retrieval, and as such, are predicted to affect noun reading times.

Method

Participants

One hundred twenty-five participants were recruited from the University of Illinois at Urbana-Champaign community for \$7 compensation or course credit. Data from three participants were excluded (one due to technical issues, one who refused to put away her phone during the experiment, and one who was an older adult).

Design

All participants completed five computerised tasks in the same order. Participants first performed a self-paced sentence reading task. This was followed by four individual-difference measures that were used in exploratory analyses that focus on individual differences in lexical representations and retrieval ability (a vocabulary test, a reading span test, a Stroop test, and verb generation, see Appendix A for details). The entire experimental session took approximately 1 hour to complete; individual difference measures were omitted for time as necessary to keep the session under an hour.

The main experiment had 48 critical stimuli like, "*The landscaper who planted the cactus already was/*were anticipating the dry summer*" that varied in the frequency, morpho-orthography (presence of a word-final *-s*), and grammatical number of local nouns embedded in subject-extracted relative clauses, paired with verbs varying in grammaticality (see Examples 1–3 in Table 1). All items contained an adverb before the verb, serving as a spillover region for the noun. This allows better dissociation of reading time effects based upon noun and verb processing (e.g. Wagers et al., 2009).

To manipulate local noun morpho-orthography, we varied whether items ended in *-s* in their singular and plural forms. These patterns also co-varied with word frequency. As can be seen in Table 1, items came in three atypical orthographic patterns: high frequency with no *-s* in either form (*man/men*), medium frequency with *-s* for both forms (*dress/dresses*), and low-frequency with an *-s* in the singular but not plural (*cactus/cacti*).

To separate the role of word frequency from morpho-orthography, each orthographically atypical item was paired with a fully regular, orthographically typical noun (*control*, see Table 1), matched across relevant properties as detailed in the Material Creation section below. Local noun orthographic type (*atypical* vs. *control*) was fully crossed with verb grammaticality (*was* (correct) vs. *were* (error)) and local noun number (*plural*

¹Encoding interference also affects processing speed (see Villata et al., 2018). However, due to the consistent semantic relationships between head and local nouns in this study, comparisons between conditions should overwhelmingly reflect retrieval interference.

Table 1. Example stimuli for each condition. Local noun regions are underlined with bold lines and local noun spillover regions are underlined with dotted lines. Main verb regions are bold and italicised and main verb spillover regions are italicised.

	Sentence	Orthography	Frequency class
1a	The physician who cured the <u>man/men</u> occasionally <i>was/were</i> <i>incorrect</i> about the diagnosis.	Critical	High
1b	The physician who cured the <u>boy/boys</u> occasionally <i>was/were</i> <i>incorrect</i> about the diagnosis.	Control	High
2a	The celebrity who promoted the <u>dress/ dresses</u> seldom <i>was/were</i> <i>seen</i> without a big entourage.	Critical	Medium
2b	The celebrity who promoted the <u>skirt/ skirts</u> seldom <i>was/were</i> <i>seen</i> without a big entourage.	Control	Medium
3a	The landscaper who planted the <u>cactus/ cacti</u> already <i>was/were</i> <i>anticipating</i> the dry summer.	Critical	Low
3b	The landscaper who planted the <u>yucca/ yuccas</u> already <i>was/were</i> <i>anticipating</i> the dry summer.	Control	Low

vs. *singular*) for a total of eight versions of each item. These were distributed into eight lists such that one version of each item was presented to each participant and such that each list contained an equal number of items of each form.

The 48 critical sentences were mixed with 60 fillers for a total of 108 sentences. Fillers included a mixture of plural and singular nouns with regular spelling contained in structures varying in difficulty. These included items containing noun phrases with prepositional phrase modifiers, items with subject- and object-extracted relative clauses containing copulas and lexical verbs, and locally ambiguous sentences containing DO- and SC-biased verbs. Most fillers were grammatical (52), with eight ungrammatical fillers. Items were pseudorandomized within each list using the program Mix (van Casteren & Davis, 2006) to prevent any within-condition repeats on back-to-back trials. A list of critical items can be found in Appendix B.

Material creation

Critical items were created with a search in the MRC Psycholinguistic database (Coltheart, 1981) for items that ended in -s in the singular form and/or did not end in -s in the plural form. We then extracted log frequencies for these items from the Corpus of Contemporary American English (COCA, Davies, 2008). Stimulus development was done in the spring of 2014; frequencies reflect numbers taken from the corpus at that time.

As noted above, the three orthographic classes of items varied inherently in word frequency, leading to creation of a control condition. Control items were

matched for suitability in the same sentence context based upon the N-gram log frequency in COCA for the association of each head noun with the control and critical local noun (occurrence of local noun within nine words after head noun) and the N-gram log frequency for the association of the relative clause verb and local noun (occurrence of local noun within five words after the RC verb). In addition to matching suitability in context, we aimed to match the frequency of control versus atypical items, as done by minimising the difference in log frequency for each version of an item (see Table 2). We also statistically controlled for the word frequency of both singular and plural local nouns by including log local noun token frequency as a covariate in analyses. See Appendix B for frequency measurements by item.

Procedure

Participants read sentences in a non-cumulative, self-paced moving-window display (Just, Carpenter, & Woolley, 1982) presented using Paradigm software (Perception Research Systems, 2007). Sentences appeared on a single line with dashes replacing letters and punctuation but with spaces preserved between words. With each mouse click, a single word was revealed at a time, with all other words masked; participants were not allowed to revisit words of the sentence. Following the final word, a yes/no comprehension question appeared on a new screen. This gauged comprehension of information unrelated to the subject-verb agreement or local noun manipulation (e.g. “*Did the summer tend to be dry?*” for Sentence 3 in Table 1). Across the items in a list, odds of correct yes/no responses and the side of the screen on which yes or no appeared (left or right) were counterbalanced. Word-by-word reading times and comprehension question accuracy were recorded.

Data analysis

Self-paced reading performance was indexed by the amount of time elapsed before clicking the mouse to progress to the next word in the sentence. We excluded critical trials with incorrect responses to comprehension questions, which resulted in a loss of 8.97% of trials (range by participant: 4.76–25.68% incorrect trials). Of the correct response trials, we then removed word-by-word reading times that were less than 200 ms and greater than 3000 ms (2.53% of words, resulting in a loss of 0.49% of data) and word-by-word reading times more than three standard deviations above participants’ mean reading times, resulting in a loss of an additional

Table 2. Average log word frequency in COCA for local nouns by noun class, atypical orthography, and number. Range is reported in parentheses.

Frequency class	Singular nouns		Plural nouns		<i>M</i>
	Orthography		Orthography		
	Control	Atypical	Control	Atypical	
High (man/men; boy/boys)	3.97 (2.57, 5.20)	3.91 (2.19, 5.48)	3.81 (2.70, 5.05)	4.23 (2.69, 5.90)	3.98
<i>Mean item-wise difference score</i>	0.05 (−0.62, 0.69)		−0.42 (−2.27, 0.56)		
Medium (dress/dresses; skirt/skirts)	3.35 (1.00, 4.62)	3.39 (1.83, 5.18)	2.87 (0.90, 4.40)	2.61 (0.60, 4.37)	3.05
<i>Mean item-wise difference score</i>	−0.05 (−0.83, 1.02)		0.26 (−0.69, 1.47)		
Low (cactus/cacti; yucca/yuccas)	2.72 (1.74, 4.01)	2.69 (1.15, 3.81)	2.29 (1.34, 3.88)	1.88 (0.30, 3.41)	2.40
<i>Mean item-wise difference score</i>	0.03 (−0.97, 1.40)		0.41 (−0.85, 1.81)		

2.53% of the data.² It is likely that these slow reading times reflect trials where a word was unknown to the participant or where the response time was slowed for reasons outside the process of interest.

Data analysis used Bayesian linear mixed-effects models in R (version 3.3.3, R Core team, 2017) using the package brms (version 1.10, Bürkner, 2017). Bayesian analyses are concerned with the likely magnitude of effects rather than statistical significance, making them well-suited to quantifying evidence for various frameworks (see e.g. Nicenboim & Vasishth, 2018; Nicenboim, Vasishth, Engelmann, & Suckow, 2018; for an application of Bayesian modelling to sentence processing; see Sorensen, Hohenstein, & Vasishth, 2016, for a tutorial). Here, our interest is in quantifying the size of parameters and the uncertainty around them. The size of reported betas reflects estimated effect sizes; betas with larger absolute values reflect larger effects. In the main text, we report the parameters for which the 95% Credible Intervals do not contain zero, which is analogous to the frequentist null hypothesis significance test: the parameter has a non-zero effect with high certainty. We also report any parameters for which the point estimate for the beta is about twice the size of its error, as this also provides evidence for an effect: the estimated effect is large compared to the uncertainty around it. We also report the posterior probability of these weak effects, indicating the proportion of samples with a value equal to or more extreme than the beta estimate.

All models had four chains and each chain had 12,000 or 14,000 iterations (listed in model output tables), with the first half representing a burn-in period. All models had weak normal priors with an SD of two, with the distribution centred at six for the intercept and centred at zero for all other parameters. This represents a weak expectation for an intercept that reflects the average reading time across the experiment, and a widely-spread, weak expectation of null effects for all other

parameters such that a wide range of effects, including minimal and extreme ones, would be consistent with the prior. All models were run until the \hat{R} value for each estimated parameter was 1.00, indicating full convergence.

Log-transformed reading times were predicted in four regions. Reading times at the local noun and the following word (local noun spillover region) were used to assess the effects of predictors on retrieving and processing the local noun; reading times at the main verb and following word (main verb spillover region) were used to assess the effects of predictors on retrieving and processing the verb's controller.

We included two continuous predictors in our models. The first accounted for the frequency of local nouns: the log-transformed Local Noun Frequency (for plural and singular tokens separately) was centred and entered as a covariate in all models. In addition, the length of the word in the region of interest (Word Length, centred) was entered as a covariate in each model. Both factors were entered as main effects only.³

Categorical predictors included Local Noun Number (Singular, Plural), Local Noun Class (Low frequency = *cactus/cacti/yucca/yuccas*; Medium frequency = *dress/dresses/skirt/skirts*; High frequency = *man/men/boy/boys*), Local Noun Morpho-orthography (Control, Atypical), and Grammaticality of the verb (Grammatical, Ungrammatical). Contrasts for the two-level variables (Local Number, Local Noun Morpho-orthography, and Grammaticality) were always $-.5$ and $.5$. Local Noun Class was coded with a simple effects contrast-coding scheme, which test pairwise comparisons between levels while preserving the interpretation of main effects, such that the intercept corresponds to the mean of cell means. The first contrast compared the two Noun Classes containing irregular plurals, with the following coefficients assigned to each level (Low: *cactus/cacti/yucca/yuccas* = $-1/3$ versus High: *man/men/*

²After trimming at 200 ms, there were no remaining words read faster than 3 standard deviations below participant means.

³We also ran a series of models that allowed Local Noun Frequency to interact with all other parameters, in addition to the binned Local Noun Class frequency predictor. This did not improve model fit by LOOIC; as such, we report the simpler models.

boy/boys = 2/3, Medium: *dress/dresses/skirt/skirts* = -1/3) and the second compared the two Noun Classes ending in -s in the singular, with the following coefficients assigned to each level (Low: *cactus/cacti/yucca/yuccas* = -1/3 versus Medium: *dress/dresses/skirt/skirts* = 2/3; High: *man/men/boy/boys* = -1/3). Effects of Local Noun Class without taking into account Local Noun Morpho-orthography reflect properties of binned word frequency, with comparisons between low and high frequency nouns in the first contrast and comparisons between low and medium frequency nouns in the second. In interaction with Local Noun Morpho-orthography, the Local Noun Class contrasts reflect the effect of type frequency for each of the morpho-orthographic patterns. The first Local Noun Class contrast represents the effect of irregularly-marked nouns versus typical controls, and the second represents the effect of the local -s for atypical nouns versus typical controls. In addition to these contrasts, we also created 95% credible intervals for relevant pairwise comparisons using the functions *emmeans*, version 1.3.3 (Lenth, 2019) and *tidybayes*, version 1.0.4 (Kay, 2019).

The random effect structure for the noun-region models included random intercepts for Participants and Items, random slopes of Local Noun Number, Local Noun Class, Local Noun Morpho-orthography, Local Noun Frequency, and Word Length by Participants, and random slopes of Local Noun Number and Local Noun Morpho-orthography by Items. The random effect structure for the verb-region models contained all of the same predictors, as well as random slopes of Grammaticality by Participants and Items.

Results

Local noun

See Figure 1 for log-transformed reading times by Word Length and Local Noun Frequency; see Figure 2(a) for mean log-transformed reading times by Local Noun Number, Local Noun Class, and Local Noun Morpho-orthography. Outputs of Bayesian mixed effect models appear in Table 3. Word Length ($\beta = 0.024$, $SE = 0.003$, 95% credible interval = [0.018, 0.030]) and log Local Noun Frequency ($\beta = -0.010$, $SE = 0.003$, 95% credible interval = [-0.016, -0.004]) both reliably impacted reading time, such that longer and infrequent words were read more slowly.

Local noun spillover

See Figure 1 for log-transformed reading times by Word Length and Local Noun Frequency; see Figure 2(b) for

mean log-transformed reading times by Local Noun Number, Local Noun Class, and Local Noun Morpho-orthography. Outputs of Bayesian mixed effect models appear in Table 3. Again, Word Length impacted reading time ($\beta = 0.017$, $SE = 0.004$, 95% credible interval = [0.008, 0.025]), such that longer words were read more slowly, as did log Local Noun Frequency ($\beta = -0.021$, $SE = 0.004$, 95% credible interval = [-0.028, -0.014]) such that the word following an infrequent noun was also read more slowly. There was also evidence for an interaction between Local Noun Number and Local Noun Morpho-orthography ($\beta = 0.039$, $SE = 0.018$, 95% credible interval = [0.004, 0.073]), such that the word following atypical plural nouns (*cacti/dresses/men*) was read more slowly than would be expected from either marginal effect (95% credible interval for pairwise difference between control and atypical plurals = [-0.061, -0.007])

Verb

See Figure 1 for log-transformed reading times by Word Length and Local Noun Frequency; see Figure 3(a) for mean log-transformed reading times by Grammaticality, Local Noun Number, Local Noun Class, and Local Noun Morpho-orthography. Outputs of Bayesian mixed effect models appear in Table 4. In this region, there was evidence for an interaction between Local Noun Number and Local Noun Morpho-orthography, ($\beta = 0.047$, $SE = 0.016$, 95% credible interval = [0.016, 0.078]), such that verbs following atypical plural nouns (*cacti/dresses/men*) were read more slowly than their typical counterparts compared to what would be predicted from the marginal effects alone (95% credible interval for pairwise difference between control and atypical plural nouns = [-0.059, -0.011]). There was also moderate evidence for an interaction between Verb Grammaticality, Noun Number, and the first Local Noun Class contrast, ($\beta = 0.070$, $SE = 0.040$, 95% credible interval = [-0.008, 0.147]); note that while the 95% credible interval contains zero, the point estimate of the beta is high relative to the error around it, and 96% of the posterior distribution around the estimated effect is greater than zero. The pattern was that grammatical verbs following singular high-frequency nouns (*man/boy*) were read more slowly than grammatical verbs following plural or singular low-frequency nouns (*cactus/cacti/yucca/yuccas*, see Figure 3).

Verb spillover

See Figure 1 for log-transformed reading times by Word Length and Local Noun Frequency; see Figure 3(a) for mean log-transformed reading times by Grammaticality, Local Noun Number, Local Noun Class, and Local Noun

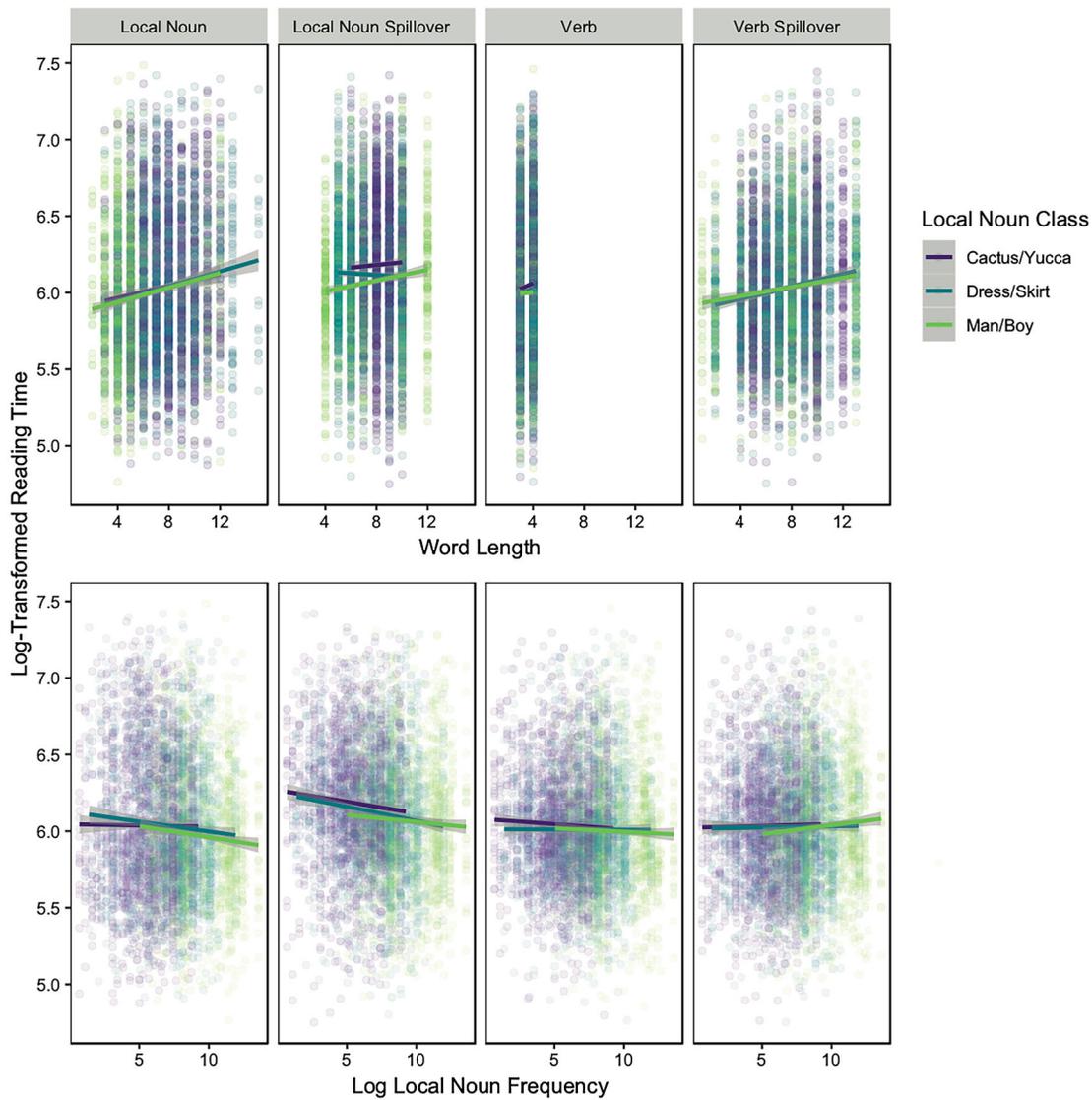


Figure 1. Log-transformed reading times and linear estimates by word region (panels, left to right) by Word Length (top row) and log Local Noun Frequency (bottom row). Colours reflect Local Noun Class, which co-varies with Local Noun Frequency.

Morpho-orthography. Outputs of Bayesian mixed effect models appear in Table 4. Word Length impacted reading time ($\beta = 0.016$, $SE = 0.003$, 95% credible interval = [0.012, 0.021]), such that longer words were read more slowly.

There was also evidence for a main effect of Verb Grammaticality ($\beta = 0.076$, $SE = 0.009$, 95% credible interval = [0.058, 0.094]), such that the spillover region after ungrammatical verbs was read more slowly. This was supported by moderate evidence for an interaction between Verb Grammaticality and Noun Number ($\beta = -0.031$, $SE = 0.016$, 95% credible interval = [-0.062, 0.001]); while the 95% credible interval contains zero, the point estimate of the beta is high relative to the error around it, and 97% of the posterior distribution around the estimated effect is below zero. The pattern

was the typical “illusion of grammaticality”, such that the spillover region after ungrammatical plural verbs following local singular nouns was read more slowly than those following plural nouns (95% credible interval for pairwise difference = [0.050, 0.004]).

Verb Grammaticality and Local Noun Number also interacted with the first Local Noun Class contrast, showing modulation of attraction by word frequency ($\beta = 0.092$, $SE = 0.040$, 95% credible interval = [0.013, 0.170]). The pattern was such that the highest-frequency words elicited the least attraction, with no reliable verb spillover reading time difference between the singular (*man/boy*) and plural (*men/boys*) ungrammatical conditions (Point estimate of singular-plural difference = 0.026; 95% credible interval for pairwise difference = [-0.020, 0.072]).

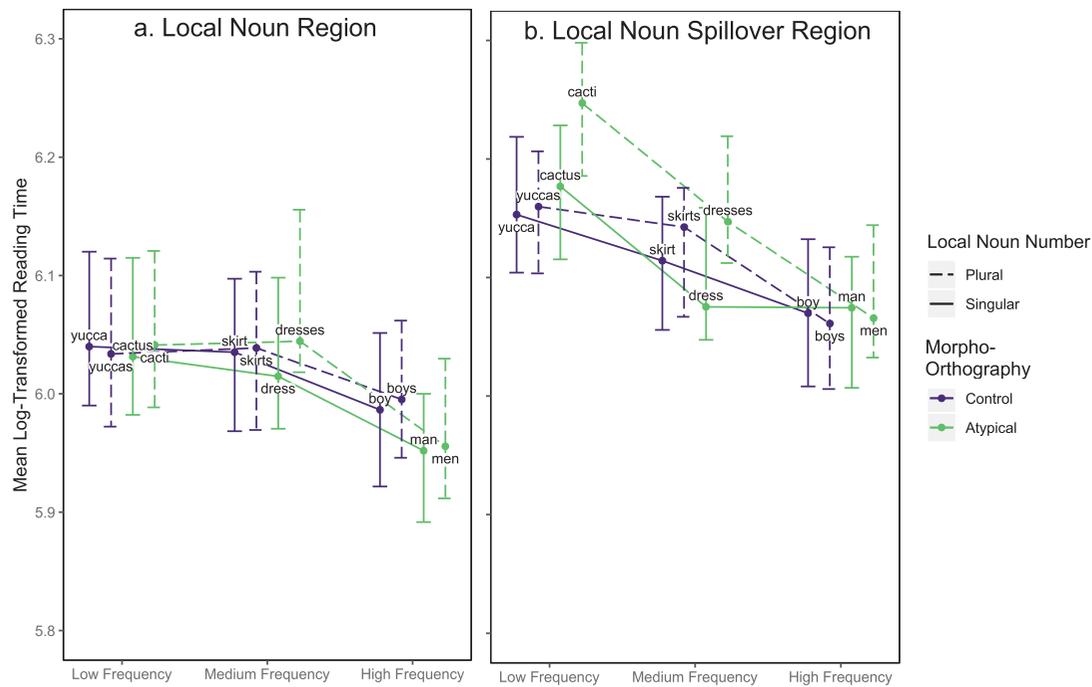


Figure 2. Mean reading times on the local noun region (left) and local noun spillover region (right), split by Local Noun Frequency class, local noun morpho-orthography and local noun number. Error bars reflect non-parametric 95% bootstrapped confidence intervals of participant means by condition.

While the second Local Noun Class contrast did not interact reliably with Verb Grammaticality and Local Noun Number, pairwise comparisons also indicate that the attraction effect observed in the aggregate was carried by the least frequent items (ungrammatical verbs following *cactus/yucca*, versus *cacti/yuccas*; point estimate of singular-plural difference = -0.063 , 95% credible interval = $[-0.109, -0.017]$). In contrast, although the point estimate suggests evidence for attraction in the middle frequency items (point estimate of singular-plural difference = -0.32), the 95% credible interval for the pairwise difference contains zero (95% CrI = $[-0.077, 0.013]$).

Finally, we also observed interactions between Morpho-orthography and Frequency class, showing the specific effect of atypical local noun plural marking on reading time in the verb spillover region. Importantly, these interactions did not involve Verb Grammaticality, indicating that they do not reflect attraction. We observed a two-way interaction between Morpho-orthography and the second Local Noun Class contrast ($\beta = 0.049$, $SE = 0.022$, 95% credible interval = $[0.005, 0.092]$) such that atypical nouns of the “*dress/dresses*” type led to faster verb spillover reading times than typically-spelled, frequency-matched items in the same sentence frames than the marginal effects would predict, with no reliable pairwise difference between control and atypical nouns in the “*dress/dresses*” condition (95% credible

interval = $[-0.052, 0.009]$). We also observed a three-way interaction between Morpho-orthography, Noun Number, and the first Local Noun Class contrast ($\beta = 0.090$, $SE = 0.041$, 95% credible interval = $[0.009, 0.168]$) such that singular atypical nouns of the “*cactus*” type led to equally-fast verb spillover reading times as regular items such as “*yucca*” (95% credible interval for difference = $[-0.046, 0.035]$), whereas the other atypical nouns in the contrast (*man/men/cacti*) led to faster verb spillover reading times than their frequency-matched regular counterparts (see Figure 3).

Exploration of individual differences in agreement processing

To further examine the mechanisms and representations supporting agreement, we now turn to exploratory analyses that further examine the cognitive mechanisms supporting agreement. Cue-based retrieval suggests that processing agreement requires activating and deploying information about lexical items. Other research suggests that the underlying cognitive mechanisms of lexical processing, memory, and executive function may differ between individuals. Previous work on similar processing phenomena (Van Dyke et al., 2014) showed vocabulary size to be an important predictor of retrieval ability. Working memory and executive control have also been linked to agreement processing

Table 3. Results of Bayesian linear mixed-effects models for log-transformed reading times at the local noun and following spillover region. All \hat{R} values were 1.00. Noun region model was run for 1400 iterations, and Noun Spill region model was run for 1200 iterations. Bolded values indicate effects where the 95% Credible Interval does not contain zero.

Population-level effects:	Estimate	Est. error	Noun region		Effective samples	Estimate	Est. error	Noun spill region		Effective samples
			95% CI					95% CI		
			Lower	Upper				Lower	Upper	
Intercept	6.023	0.029	5.965	6.082	989	6.128	0.025	6.078	6.177	1320
Word Length	0.024	0.003	0.018	0.030	17,437	0.017	0.004	0.008	0.025	9805
Log Local Noun Frequency	-0.010	0.003	-0.016	-0.004	11,325	-0.021	0.004	-0.028	-0.014	11,993
Local Noun Number	-0.016	0.011	-0.037	0.005	28,000	0.009	0.010	-0.011	0.030	24,000
Local Noun Class 1 (Low vs High Freq)	0.017	0.018	-0.019	0.052	16,947	-0.028	0.023	-0.073	0.019	9811
Local Noun Class 2 (Low vs Med Freq)	0.015	0.015	-0.015	0.045	18,557	-0.010	0.020	-0.049	0.030	10,425
Local Noun Morpho-orthography	0.004	0.011	-0.018	0.025	28,000	0.015	0.010	-0.006	0.035	24,000
Local N Number x Local Noun Class 1	0.029	0.025	-0.020	0.079	28,000	0.022	0.024	-0.026	0.069	24,000
Local N Number x Local Noun Class 2	-0.021	0.025	-0.070	0.028	28,000	0.011	0.023	-0.034	0.056	24,000
Local N Number x Local Noun Morpho-ortho.	0.033	0.019	-0.003	0.069	28,000	0.039	0.018	0.004	0.073	24,000
Local N Class 1 x Local Noun Morpho-ortho.	-0.029	0.025	-0.078	0.018	28,000	-0.018	0.025	-0.067	0.031	24,000
Local N Class 2 x Local Noun Morpho-ortho.	-0.001	0.025	-0.049	0.048	28,000	-0.039	0.024	-0.087	0.008	24,000
Local N Num x Local N Cl 1 x Local N M-ortho.	0.010	0.047	-0.082	0.102	28,000	0.020	0.045	-0.068	0.108	24,000
Local N Num x Local N Cl 2 x Local N M-ortho.	-0.051	0.046	-0.141	0.041	28,000	-0.014	0.044	-0.099	0.072	24,000
Group-level effects:	Estimate	Est. error	95% CI		Effective samples	Estimate	Est. error	95% CI		Effective samples
Subject			Lower	Upper				Lower	Upper	
sd(Intercept)	0.317	0.021	0.279	0.361	2455	0.261	0.017	0.229	0.298	2754
sd(Loc N Num)	0.016	0.012	0.001	0.044	11,170	0.029	0.016	0.002	0.061	5399
sd(Loc N Class 1)	0.022	0.016	0.001	0.058	8587	0.019	0.014	0.001	0.052	10,219
sd(Loc N Class 2)	0.022	0.016	0.001	0.058	7932	0.017	0.013	0.001	0.049	8657
sd(Loc N Morpho-orth)	0.044	0.019	0.005	0.079	4509	0.029	0.017	0.002	0.065	3761
sd(Word Length)	0.017	0.003	0.010	0.023	9152	0.013	0.003	0.006	0.019	10,212
sd(Log Loc N Freq)	0.013	0.003	0.007	0.019	10,908	0.016	0.003	0.010	0.022	7280
cor(Intercept, Loc N Num)	-0.056	0.319	-0.653	0.579	28,000	-0.305	0.260	-0.747	0.290	24,000
cor(Intercept, Loc N Class 1)	-0.063	0.315	-0.650	0.575	28,000	0.088	0.321	-0.559	0.680	24,000
cor(Loc N Num, Loc N Class 1)	-0.026	0.355	-0.687	0.654	17,493	-0.021	0.347	-0.666	0.651	24,000
cor(Intercept, Loc N Class 2)	0.105	0.307	-0.540	0.670	28,000	0.025	0.320	-0.604	0.638	24,000
cor(Loc N Num, Loc N Class 2)	0.004	0.351	-0.659	0.662	21,067	-0.046	0.349	-0.685	0.628	24,000
cor(Loc N Class 1, Loc N Class 2)	0.055	0.353	-0.639	0.698	19,344	0.025	0.355	-0.653	0.686	24,000
cor(Intercept, Loc N Morpho-orth)	-0.141	0.221	-0.563	0.322	28,000	-0.112	0.266	-0.613	0.455	24,000
cor(Loc N Num, Loc N Morpho-orth)	0.010	0.349	-0.655	0.667	6546	0.015	0.340	-0.633	0.658	11,336
cor(Loc N Class 1, Loc N Morpho-orth)	-0.100	0.347	-0.718	0.591	6975	-0.031	0.349	-0.677	0.639	9213
cor(Loc N Class 2, Loc N Morpho-orth)	-0.070	0.348	-0.696	0.605	8433	0.003	0.351	-0.661	0.661	10,584
cor(Intercept, Word Length)	0.346	0.141	0.059	0.616	20,658	0.610	0.152	0.282	0.869	24,000
cor(Loc N Num, Word Length)	-0.069	0.346	-0.702	0.604	3405	-0.207	0.306	-0.750	0.429	10,760
cor(Loc N Class 1, Word Length)	-0.119	0.335	-0.719	0.563	3934	0.063	0.337	-0.595	0.687	11,359
cor(Loc N Class 2, Word Length)	0.054	0.332	-0.592	0.675	5298	0.055	0.338	-0.606	0.678	12,927
cor(Loc N Morpho-orth, Word Length)	0.132	0.277	-0.447	0.635	7278	-0.095	0.313	-0.67	0.532	11,829
cor(Intercept, Log Loc N Freq)	-0.551	0.145	-0.813	-0.247	19,475	-0.026	0.145	-0.304	0.261	21,592
cor(Loc N Num, Log Loc N Freq)	0.106	0.341	-0.575	0.714	6838	0.130	0.308	-0.513	0.671	2666
cor(Loc N Class 1, Log Loc N Freq)	0.046	0.337	-0.600	0.681	8260	-0.055	0.342	-0.690	0.613	2604
cor(Loc N Class 2, Log Loc N Freq)	-0.076	0.333	-0.687	0.585	9747	0.001	0.341	-0.646	0.650	2609
cor(Loc N Morpho-orth, Log Loc N Freq)	-0.108	0.285	-0.623	0.488	11,435	0.210	0.312	-0.467	0.740	3331
cor(Word Length, Log Loc N Freq)	0.003	0.248	-0.517	0.453	12,814	-0.117	0.244	-0.565	0.382	6669
Item										
sd(Intercept)	0.022	0.008	0.004	0.038	5686	0.041	0.007	0.027	0.057	9812
sd(Loc N Num)	0.021	0.014	0.001	0.051	9582	0.018	0.012	0.001	0.045	9818

<i>sd(Loc N Morpho-orth)</i>	0.021	0.014	0.001	0.052	9386	0.026	0.016	0.001	0.059	5746
<i>cor(Intercept, Loc N Num)</i>	0.273	0.458	-0.744	0.932	15,288	-0.351	0.425	-0.939	0.653	24,000
<i>cor(Intercept, Loc N Morpho-orth)</i>	0.208	0.455	-0.751	0.913	18,040	0.053	0.398	-0.737	0.804	17,560
<i>cor(Loc N Num, Loc N Morpho-orth)</i>	0.040	0.489	-0.854	0.878	17,043	-0.08	0.486	-0.891	0.844	8813
<i>Family specific parameters:</i>	<i>Estimate</i>	<i>Est. error</i>	<i>95% CI</i>		<i>Effective samples</i>	<i>Estimate</i>	<i>Est. error</i>	<i>95% CI</i>		<i>Effective samples</i>
			<i>Lower</i>	<i>Upper</i>				<i>Lower</i>	<i>Upper</i>	
<i>sigma</i>	0.326	0.003	0.319	0.333	28,000	0.312	0.003	0.305	0.318	24,000
<i>Model comparisons</i>	<i>Interaction model</i>		<i>Main effects only model</i>		<i>Interaction model</i>		<i>Main effects only model</i>			
	LOOIC	SE	LOOIC	SE	LOOIC	SE	LOOIC	SE		
<i>Vocabulary Score</i>	-17.33	6.36	0.23	0.65	-11.53	7.88		0.06		0.94
<i>Reading Span</i>	-16.65	6.23	-0.56	0.33	-12.71	6.7		-0.16		0.64
<i>Stroop Cost</i>	-9.51	7.06	2.62	1.65	-7.81	8.72		-0.17		0.42
<i>Verb Selection Cost</i>	-4.45	8.3	-0.34	1.99	-8.32	8.32		-1.39		1.62
<i>Trial Exclusion Count</i>	-14.54	7.06	-0.30	0.81	-17.36	6.59		1.17		0.91

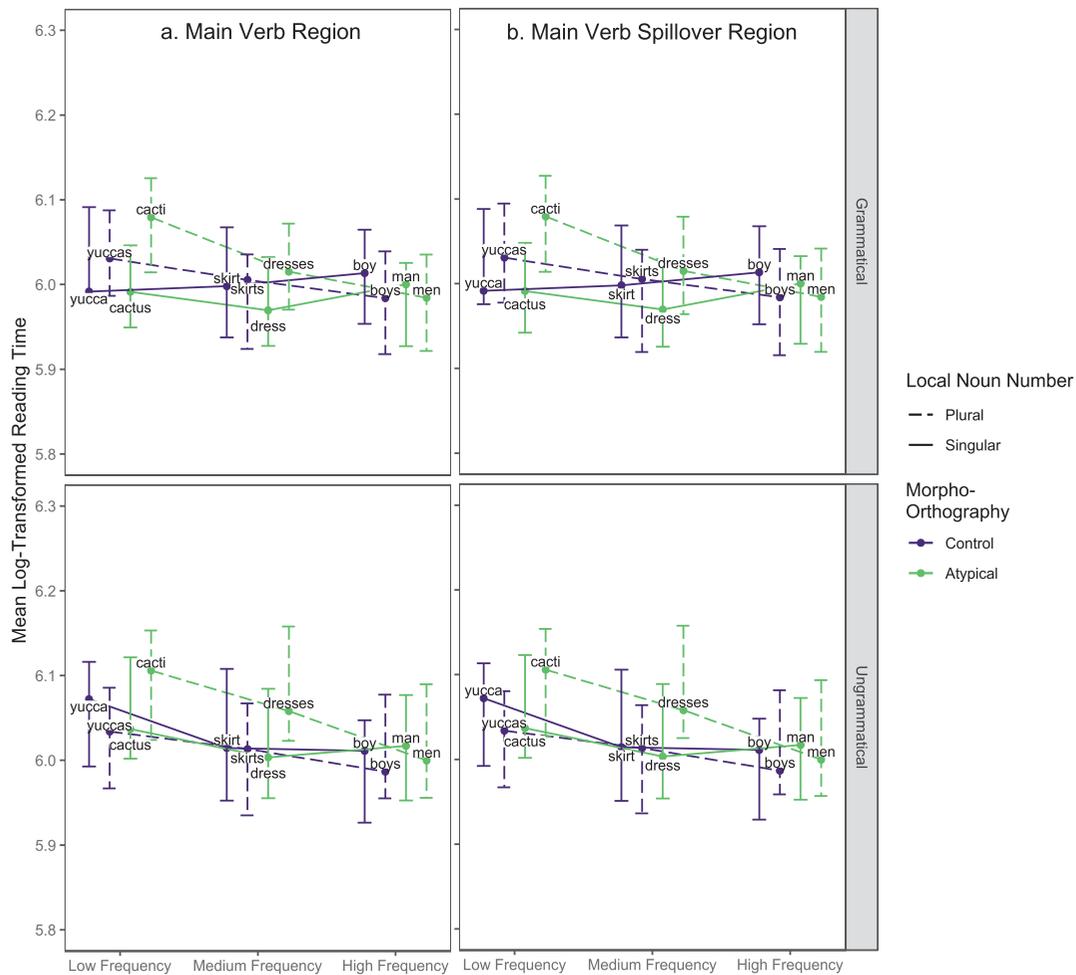


Figure 3. Mean reading times on the main verb region (left) and main verb spillover region (right), split by verb grammaticality (rows), split by Local Noun Frequency class, local noun morpho-orthography and local noun number.

and production, often in combination with each other (e.g. Fyndanis, Arcara, Christidou, & Caplan, 2018; Hartsuiker & Barkhuysen, 2006; Reifegerste, Hauer, & Felser, 2017; Slevc & Martin, 2016; Vandierendonck, Loncke, Hartsuiker, & Desmet, 2017; Veenstra, Antoniou, Katsos, & Kissine, 2018; though see Bock & Cutting, 1992 for null results), and these have also been critically linked to other domains of sentence processing (Badre & Wagner, 2007; Christianson, Williams, Zacks, & Ferreira, 2006; Daneman & Carpenter, 1980; Hussey et al., 2017; Hussey & Novick, 2012; Hussey, Ward, Christianson, & Kramer, 2015; January, Trueswell, & Thompson-Schill, 2009; King & Just, 1991; MacDonald, Just, & Carpenter, 1992; Novick, Trueswell, & Thompson-Schill, 2005; Snyder & Munakata, 2008; Swets, Desmet, Hambrick, & Ferreira, 2007; Vuong & Martin, 2014; Ye & Zhou, 2009).

Given the plausible role for these cognitive mechanisms in agreement processing, and given the large sample of participants used in this experiment, we

decided to use any time remaining in the 1 hour experimental session after the self-paced reading study to collect four continuous individual difference measures (Vocabulary Size, Reading Span, Verb Selection Cost, and Stroop Cost, see Appendix A for methods). We hypothesised that these factors might contribute to agreement processing or reading times on irregular nouns. We predicted that high vocabulary, high working memory, low verb selection cost, and low Stroop cost might lead to faster reading times overall and might especially do so when number cues conflict. We also included the number of trials each participant got wrong as a predictor for reading time in the successfully-answered trials, providing an index of individual reading ability and/or attention to the reading task.

To each of the four baseline models described in the previous sections, we added each of the five centred continuous predictors in turn, in one model as a main effect only and in another, as a main effect plus all of

Table 4. Results of Bayesian linear mixed-effects models for log-transformed reading times at the verb and following spillover region. All \hat{R} values were 1.00. Verb region model was run for 1200 iterations, and Verb Spill region model was run for 1400 iterations. Bolded values indicate effects where the 95% Credible Interval does not contain zero; italicised values indicate effects where the beta estimate is twice the estimate of the standard error.

Population-level effects:	Estimate	Est. error	Verb region		Effective Samples	Estimate	Est. error	Verb spill region		Effective Samples
			95% CI					95% CI		
			Lower	Upper				Lower	Upper	
Intercept	6.019	0.024	5.972	6.066	1634	6.032	0.024	5.985	6.078	1356
Word Length	0.003	1.415	-2.778	2.796	19,590	0.016	0.002	0.012	0.021	13,561
Log Local Noun Frequency	-0.005	0.004	-0.012	0.002	9844	-0.001	0.003	-0.006	0.004	17,689
Verb Grammaticality	0.025	1.415	-2.770	2.809	19,591	0.076	0.009	0.058	0.094	28,000
Local Noun Number	0.011	0.011	-0.011	0.034	19,842	-0.008	0.011	-0.029	0.014	24,569
Local Noun Class 1 (Low vs High Freq)	-0.025	0.025	-0.073	0.024	7711	0.004	0.017	-0.029	0.038	14,551
Local Noun Class 2 (Low vs Med Freq)	-0.024	0.022	-0.066	0.021	7586	0.004	0.014	-0.024	0.032	15,279
Local Noun Morpho-orthography	0.012	0.009	-0.006	0.030	24,000	-0.010	0.009	-0.028	0.007	28,000
Verb Grammaticality x Local Noun Number	-0.017	0.016	-0.048	0.015	24,000	-0.031	0.016	-0.062	0.001	28,000
Verb Grammaticality x Local Noun Class 1	-0.013	0.02	-0.053	0.027	24,000	0.003	0.021	-0.039	0.046	28,000
Verb Grammaticality x Local Noun Class 2	0.002	0.02	-0.037	0.042	24,000	0.011	0.021	-0.031	0.054	28,000
Local N Number x Local Noun Class 1	-0.021	0.025	-0.071	0.027	17,887	0.012	0.026	-0.039	0.063	20,643
Local N Number x Local Noun Class 2	-0.012	0.024	-0.059	0.034	18,280	-0.034	0.025	-0.084	0.016	21,440
Verb Gramm. x Local Noun Morpho-ortho	0.014	0.016	-0.018	0.046	24,000	-0.013	0.016	-0.045	0.018	28,000
Local N Number x Local Noun Morpho-ortho.	0.047	0.016	0.016	0.078	24,000	-0.013	0.016	-0.045	0.019	28,000
Local N Class 1 x Local Noun Morpho-ortho.	-0.006	0.023	-0.051	0.038	24,000	0.000	0.022	-0.044	0.044	28,000
Local N Class 2 x Local Noun Morpho-ortho.	0.013	0.022	-0.031	0.057	24,000	0.049	0.022	0.005	0.092	28,000
Verb Gramm x Local N Number x Local N Class 1	<i>0.070</i>	<i>0.040</i>	-0.008	<i>0.147</i>	<i>24,000</i>	0.092	0.040	0.013	0.170	28,000
Verb Gramm x Local N Number x Local N Class 2	0.006	0.040	-0.071	0.084	24,000	0.005	0.039	-0.071	0.082	28,000
Verb Gramm x Local N Num x Local N M-ortho	0.024	0.032	-0.038	0.086	24,000	0.034	0.032	-0.028	0.096	28,000
Verb Gramm x Local N CI 1 x Local N M-ortho	0.001	0.040	-0.076	0.079	24,000	0.032	0.040	-0.046	0.111	28,000
Verb Gramm x Local N CI 2 x Local N M-ortho	-0.007	0.039	-0.084	0.071	24,000	-0.028	0.040	-0.106	0.050	28,000
Local N Num x Local N CI 1 x Local N M-ortho.	0.001	0.040	-0.078	0.080	24,000	0.090	0.041	0.009	0.168	28,000
Local N Num x Local N CI 2 x Local N M-ortho.	-0.004	0.039	-0.081	0.073	24,000	0.064	0.040	-0.014	0.143	28,000
Verb Gram x LocN Num x LocN CI 1 x LocN M-ortho	-0.009	0.079	-0.165	0.145	24,000	0.013	0.080	-0.144	0.169	28,000
Verb Gram x LocN Num x LocN CI 2 x LocN M-ortho	0.042	0.079	-0.113	0.195	24,000	-0.016	0.080	-0.173	0.140	28,000
Group-level effects:	Estimate	Est. Error	95% CI		Effective samples	Estimate	Est. Error	95% CI		Effective samples
Subject			Lower	Upper				Lower	Upper	
sd(Intercept)	0.235	0.016	0.207	0.268	2706	0.259	0.017	0.228	0.294	2470
sd(Verb Grammaticality)	0.034	0.002	0.002	0.073	7017	0.033	0.016	0.003	0.063	4847
sd(Local Noun Number)	0.047	0.017	0.008	0.076	3675	0.018	0.012	0.001	0.045	8805
sd(Local Noun Class 1)	0.016	0.012	0.001	0.044	11,326	0.023	0.015	0.001	0.056	8062
sd(Local Noun Class 2)	0.030	0.016	0.002	0.061	5512	0.019	0.014	0.001	0.051	8142
sd(Word Length)	0.034	0.020	0.002	0.073	4404	0.008	0.002	0.003	0.012	11,327
sd(Log Local Noun Frequency)	0.008	0.003	0.001	0.014	4745	0.004	0.003	0.000	0.009	7482
cor(Intercept, Verb Grammaticality)	0.374	0.342	-0.449	0.847	9169	0.289	0.232	-0.228	0.698	28,000
cor(Intercept, Local Noun Number)	-0.026	0.196	-0.413	0.370	24,000	-0.175	0.300	-0.702	0.467	28,000
cor(Verb Grammaticality, Local Noun Number)	0.056	0.322	-0.577	0.662	5421	-0.114	0.340	-0.717	0.564	28,000
cor(Intercept, Local Noun Class 1)	0.128	0.330	-0.550	0.713	24,000	0.098	0.297	-0.516	0.654	28,000
cor(Verb Grammaticality, Local Noun Class 1)	0.065	0.353	-0.629	0.707	24,000	0.080	0.346	-0.601	0.708	19,649
cor(Local Noun Number, Local Noun Class 1)	0.024	0.346	-0.638	0.674	24,000	-0.077	0.353	-0.712	0.616	14,906
cor(Intercept, Local Noun Class 2)	0.298	0.253	-0.278	0.730	24,000	-0.025	0.303	-0.608	0.579	28,000
cor(Verb Grammaticality, Local Noun Class 2)	0.077	0.333	-0.576	0.692	10,863	-0.097	0.349	-0.714	0.598	21,842
cor(Local Noun Number, Local Noun Class 2)	-0.034	0.315	-0.623	0.583	15,189	0.053	0.347	-0.628	0.692	20,796
cor(Local Noun Class 1, Local Noun Class 2)	0.111	0.355	-0.589	0.738	10,158	0.042	0.351	-0.637	0.690	20,611
cor(Intercept, Word Length)	0.383	0.339	-0.421	0.848	9729	0.582	0.164	0.228	0.861	28,000
cor(Verb Grammaticality, Word Length)	-0.101	0.366	-0.751	0.616	24,000	0.015	0.303	-0.545	0.610	11,699

<i>cor(Local Noun Number, Word Length)</i>	0.051	0.320	-0.578	0.643	12,725	-0.034	0.329	-0.662	0.592	16,124
<i>cor(Local Noun Class 1, Word Length)</i>	0.069	0.348	-0.618	0.697	16,217	0.117	0.331	-0.564	0.703	14,479
<i>cor(Local Noun Class 2, Word Length)</i>	0.082	0.332	-0.584	0.690	16,203	0.006	0.334	-0.629	0.643	17,513
<i>cor(Intercept, Log Local Noun Frequency)</i>	0.266	0.208	-0.169	0.652	24,000	-0.081	0.299	-0.640	0.526	28,000
<i>cor(Verb Gram, Log Local Noun Frequency)</i>	0.135	0.328	-0.548	0.707	5779	0.017	0.342	-0.638	0.657	22,863
<i>cor(Local Noun Number, Log Local Noun Frequency)</i>	0.316	0.292	-0.365	0.780	8020	0.029	0.351	-0.645	0.679	20,269
<i>cor(Local Noun Class 1, Log Local Noun Frequency)</i>	-0.031	0.353	-0.677	0.651	9166	-0.020	0.352	-0.676	0.649	20,512
<i>cor(Local Noun Class 2, Log Local Noun Frequency)</i>	0.310	0.314	-0.402	0.807	7971	0.008	0.351	-0.657	0.671	18,977
<i>cor(Word Length, Log Local Noun Frequency)</i>	0.139	0.329	-0.544	0.718	10,220	-0.022	0.330	-0.649	0.616	28,000
<i>Item</i>										
<i>sd(Intercept)</i>	0.052	0.008	0.038	0.068	8801	0.024	0.007	0.007	0.038	5600
<i>sd(Verb Grammaticality)</i>	0.011	0.008	0.000	0.031	14,814	0.019	0.013	0.001	0.047	8244
<i>sd(Local Noun Number)</i>	0.037	0.014	0.008	0.064	6728	0.043	0.015	0.010	0.070	6062
<i>sd(Local Noun Morpho-orthography)</i>	0.026	0.014	0.002	0.054	5925	0.024	0.014	0.001	0.053	7114
<i>cor(Intercept, Verb Grammaticality)</i>	-0.073	0.421	-0.815	0.748	24,000	-0.026	0.419	-0.791	0.759	17,812
<i>cor(Intercept, Local Noun Number)</i>	-0.403	0.260	-0.843	0.157	16,512	-0.032	0.327	-0.659	0.615	9932
<i>cor(Verb Grammaticality, Local Noun Number)</i>	0.031	0.438	-0.784	0.815	6654	0.124	0.414	-0.713	0.828	5277
<i>cor(Intercept, Local Noun Morpho-orthography)</i>	0.266	0.326	-0.459	0.815	24,000	0.052	0.397	-0.724	0.773	16,796
<i>cor(Verb Gram, Local Noun Morpho-orthography)</i>	0.002	0.442	-0.806	0.805	8822	-0.128	0.436	-0.852	0.736	11,451
<i>cor(Local N Number, Local N Morpho-orthography)</i>	0.090	0.397	-0.716	0.783	13,358	-0.134	0.401	-0.826	0.681	18,195
<i>Family specific parameters:</i>										
	<i>Estimate</i>	<i>Est. error</i>	<i>95% CI</i>		<i>Effective samples</i>	<i>Estimate</i>	<i>Est. error</i>	<i>95% CI</i>		<i>Effective samples</i>
			<i>Lower</i>	<i>Upper</i>				<i>Lower</i>	<i>Upper</i>	
<i>sigma</i>	0.286	0.003	0.28	0.292	24,000	0.289	0.003	0.283	0.295	28,000
<i>Model comparisons</i>										
	<i>Interaction model</i>		<i>Main effects only model</i>			<i>Interaction model</i>		<i>Main effects only model</i>		
	LOOIC	SE	LOOIC	SE		LOOIC	SE	LOOIC	SE	
<i>Vocabulary Score</i>	-29.74	9.24	2.28	1.31		-23.73	11.47	0.04	0.74	
<i>Reading Span</i>	-23.86	9.61	1.14	1.39		-15.81	11.29	0.00	0.75	
<i>Stroop Cost</i>	-23.44	9.89	-0.78	0.72		-30.25	9.12	-1.56	0.66	
<i>Verb Selection Cost</i>	-22.97	8.88	0.17	1.61		-25.52	9.57	-1.24	1.23	
<i>Trial Exclusion Count</i>	-29.11	9.16	-0.23	0.8		-24.26	10.65	0.12	0.48	

its interactions with other predictors. We used the leave-one-out information criterion (LOOIC), calculated using the *brms* function `loo(model1, model2)` to assess whether adding individual difference variables improved model fit. This function estimates the predictive accuracy of a model by re-running it while leaving out one observation at a time; as with other information criteria, the smaller value indicates the preferred model. We report the difference between the LOOIC for the original model and the LOOIC for the model with added predictors; a negative value indicates that the original model is preferred, while a positive value indicates that the model containing the individual difference factor is preferred.

In the noun region (see Table 4), there was weak evidence for Stroop Cost impacting reading time. Adding Stroop Cost as a main effect only to this model tended to improve model fit ($\beta = 0.547$, $SE = 0.283$, 95% credible interval = $[-0.017, 1.094]$), such that participants who suffered more interference in the Stroop task also had slower reading times on local nouns, providing evidence that cognitive control affects retrieval ability. In the noun spillover region, (see Table 4), the number of excluded trials also had an impact on reading time ($\beta = -0.012$, $SE = 0.010$, 95% credible interval = $[-0.031, 0.007]$), where individuals with more excluded trials tended to have faster reading times on local nouns; this is likely because the trials excluded tended to be the ones with the slowest reading times and excluding them made the condition averages smaller. There was no evidence for any of the other factors affecting reading time in the noun or noun spillover regions.

In the verb region, Vocabulary size improved model fit as a main effect only ($\beta = -0.222$, $SE = 0.146$, 95% credible interval = $[-0.508, 0.066]$) such that individuals with larger vocabulary scores tended to have faster verb reading times. This is consistent with the findings of Van Dyke et al. (2014), supporting their argument that a large vocabulary contains lexical items with strong representations and this influences retrieval in sentence processing – but this has no impact on attraction *per se*. There was no evidence for any of the other factors affecting reading time in the verb or verb spillover regions.

Discussion

The present study examined the role of local noun frequency and morpho-orthography in agreement processing. Broadly speaking, we found that frequency impacts both noun retrieval and agreement attraction, with the most attraction observed for low-frequency local nouns. In contrast, morpho-orthography impacts

noun retrieval and creates local verb processing difficulty that is distinct from attraction. This supports cue-based retrieval models as well as the role of probabilistic factors in retrieval, demonstrating that both sets of mechanisms contribute to agreement comprehension. It also clearly shows that while many factors impact lexical retrieval, it is number that impacts number agreement. This in turn implies that it is number features that are retrieved to resolve agreement, not the full representation of the controlling noun. We unpack each finding and its contribution to cue-based retrieval and probabilistic understandings of agreement below.

In the local noun region and following spillover word, we showed that individuals tended to slow down for low-frequency local nouns, consistent with the slower retrieval of low-frequency nouns shown in single-word tasks (e.g. Baayen et al., 1997; Bates et al., 2003; Jescheniak & Levelt, 1994; Schreuder & Baayen, 1997). Noun number also impacted noun reading times, with individuals slowing down on words following atypical plural nouns. This is broadly consistent with effects of number morphology on lexical retrieval (e.g. Andrews, 1997; Coltheart et al., 1977; Grainger et al., 1989; Snodgrass & Mintzer, 1993). Together, these findings indicate the importance of interference in processing nouns embedded in sentences, which is fully consistent with the difficulty of reading nouns of these types in isolation. Nouns that are harder to retrieve due to their low frequency (“*yuccas*”) or atypical number specification (“*cacti*”) are generally harder to process, which affects reading times in sentences.

At the verb spillover region, we showed evidence for the typical attraction effect, such that ungrammatical verbs following local plural nouns elicited faster processing than those following local singular nouns. This “illusion of grammaticality” is the core agreement attraction pattern, and it replicates previous work on the comprehension of subject-verb agreement (e.g. Dillon et al., 2013; Lago et al., 2015; Martin et al., 2012; Tanner et al., 2014; Wagers et al., 2009). In this region, noun morpho-orthography did not impact attraction, consistent with work in production (Bock & Eberhard, 1993; Haskell & MacDonald, 2003). The lack of an effect of morpho-orthography at the verb processing regions combined with the effect observed at the noun regions is of critical importance for several reasons.

First, the literature on single-word processing suggests that readers may aim to decompose the *-s* affix even when this is not licensed (e.g. Rastle & Davis, 2008), especially for low-frequency irregular nouns (e.g. Alegre & Gordon, 1999; Baayen et al., 1997). Our results show that while this might occur at the local noun, it

has no consequences for attraction. The implication is that unlicensed (“pseudo”) morphological decomposition effects are transient during sentence processing, and the correct decomposition of words dominates number agreement.

Second, given that only number features matter for verb processing, the implication is that a full representation of the local noun is not retrieved at the verb region – only the relevant number feature. Number features on nouns conveyed with any type of morphology are the main driver of agreement attraction. If, as suggested by Wagers et al. (2009), attraction occurs when an ungrammatical verb triggers re-analysis, this suggests that it is plural features that are drawn from memory and create the observed illusion of grammaticality. This provides an important boundary on cue-based retrieval in comprehension: features of words, not words themselves, are retrieved.

Importantly, attraction was also qualified by local noun frequency: the most attraction occurred for low frequency nouns (*cacti/yuccas*), and there was no evidence for attraction with high frequency nouns regardless of morpho-orthography (*men/boys*). This provides further support that difficulty in lexical retrieval is the main driver of attraction, which is consistent with the current view of attraction as operating under a cue-based retrieval framework. It also suggests that attraction does not always occur as a last-resort strategy: instead, fast processing might induce a shallow or underspecified parse of the verb (e.g. as in *good enough processing*, see Christianson et al., 2006, among others). This suggests that not all local nouns trigger re-analysis: easy-to-retrieve local nouns may induce the reader to proceed without a full representation of the sentence structure. This finding is highly novel, and serves to integrate the “illusion of grammaticality” with the pattern that not all readers process structure veridically: both mis-retrieval and processing without full structure-building allow readers to extract meaning from sentences containing errors.

Though it did not impact attraction, morpho-orthography did lead to processing difficulty in the verb and verb spillover region. First, at the verb region, atypical plural nouns led to processing difficulty regardless of verb grammaticality. Taken in combination with an observed interaction between verb grammaticality, noun number, and noun class, this shows that atypical plural nouns of the “*cacti*” and “*dresses*” type tended to elicit slower verb reading times, and atypical plural nouns of the “*cacti*” type tended to elicit slower verb spillover reading times. This suggests that the atypical morpho-orthography on these nouns has downstream consequences for processing such that later verbs

require more processing time regardless of number. Such a pattern fits with a probabilistic agreement model (e.g. Haskell et al., 2010; Mirković & MacDonald, 2013; Villata et al., 2018). Nouns with atypical morpho-orthography may have an association to both plural and singular number features due to their actual number marking and their pseudo-morphological pattern (e.g. Alegre & Gordon, 1999; Baayen et al., 1997; Rastle & Davis, 2008), leading to later agreement trouble. The critical implication is that while plural number features drive agreement attraction, the activation of these plural number features may in part be probabilistically determined. This means that any successful model of agreement needs to allow for graded representations of number features.

Individual differences in sentence processing are small but consistent

There were few effects of individual differences on sentence reading times, which suggests that as a whole, while reading long, complex sentences requires some effort and may be perceived as difficult, it does not tend to max out individuals’ cognitive abilities. This is consistent with the wider literature on individual differences in sentence processing, which shows that many underlying abilities co-vary, and many underlying abilities relate to variations in processing. The implication is that there are multiple mechanisms that allow us to overcome cognitive difficulty when comprehending language.

We did find evidence for two effects on reading times within the sentence that are consistent with earlier work. Mirroring the finding that executive control impacts the rate of inhibition of local nouns, (e.g. Vandierendonck et al., 2017; Veenstra et al., 2018), we showed that Stroop cost impacted reading times on the local noun region. This suggests the importance of resolving interference when processing local nouns, and is broadly consistent with new findings of encoding interference during noun processing (e.g. Villata et al., 2018).

In addition, consistent with Van Dyke and colleagues (Van Dyke et al., 2014; Van Dyke & Johns, 2012), we showed that vocabulary size impacted verb reading times. The implication is that individuals with larger vocabularies tend to have higher-quality lexical representations, and these lead to improved processing. Individuals with large vocabularies tend to be faster readers overall (e.g. Mainz, Shao, Brysbaert, & Meyer, 2017). The current work suggests that large vocabularies may lead to relative speed in processing particularly difficult regions of sentences, such as encountering a verb with an unclear controller.

Reconciling probabilistic representations with cue-based retrieval

In the introduction, we noted some probabilistic factors that affect agreement production, making some errors more likely to be observed by a comprehender. The present data show the important role of word frequency, a probabilistic factor, on agreement attraction, consistent with the experience-based account of agreement production by MacDonald and colleagues (Haskell et al., 2010; Mirković & MacDonald, 2013). Meanwhile, morpho-orthography, another probabilistic factor, had minimal impact, mirroring the weak effects of morpho-orthography and phonology on agreement in the domain of language production (e.g. Bock & Eberhard, 1993; Haskell & MacDonald, 2003). Combined with the work reviewed in the introduction, the implication is that lexical properties affect production and comprehension similarly: lexical retrieval is an important part of agreement. This provides evidence that the representations and mechanisms operating at the word (or lemma) level are shared across sentence comprehension and production, as in current integrated frameworks (e.g. Dell & Chang, 2014; Pickering & Garrod, 2013).

To reconcile the role of probabilistic factors with the cue-based retrieval framework, we might view our results in light of a feature-based view of memory, such as MINERVA (e.g. Hintzman, 1984), where representational strength impacts the ease of memory retrieval. Previous literature (e.g. Wagers et al., 2009) has suggested that an ungrammatical plural verb cues the retrieval of a plural noun controller; these data suggest that when processing grammatical and ungrammatical verbs, nouns with an unclear feature specification may be hard to retrieve as agreement controllers, and when processing ungrammatical verbs, high-frequency local nouns may be resistant to mis-retrieval.

As a whole, this means that the data support a cue-based retrieval framework where words have lexical representations that vary in quality based upon how many tokens have been encountered. This might be modelled by making the structural or lexical representations that cue-based retrieval operates upon a gradient (see e.g. Brehm & Goldrick, 2017, for a gradient model of structure), or by appealing to an SOSP framework (e.g. Smith et al., 2018; Villata et al., 2018). Importantly, the core finding is that lexical properties such as word frequency and morpho-orthography impact the representation of words, while plural features exist as either a separately represented or more strongly represented aspect of each word. This means that probabilistic factors impact agreement and can lead to a general

slowing as words are read. However, what is most critical for processing number agreement is simply number.

Conclusion

The present work shows how agreement relies upon cue-based mechanisms and probabilistic, graded representations. While number features drive the fortunes and misfortunes of number agreement, lexical properties that affect the robustness of the noun's representation can protect from attraction – making lower-frequency words like *cacti* or *yuccas* better attractors than high-frequency words like *men* or *boys*. This means that future accounts of number agreement processing need to account for the quality of noun lexical representations as well as the mechanisms operating on them.

Acknowledgements

This research was supported by the Departments of Psychology and Educational Psychology at the University of Illinois at Urbana-Champaign via funding from NIH training grant T32-HD055272. The authors thank members of the Educational Psychology Psycholinguistics Lab, especially Kyra Michon and Yasmin Gonzalez, for their assistance with collecting and scoring data. We also are grateful for valuable feedback on the design and implementation of the experiment provided by Darren Tanner. Finally, portions of this work were presented at the 55th Annual Meeting of the Psychonomics Society (Long Beach, CA), the 28th Annual CUNY Conference on Human Sentence Processing (Los Angeles, California), and the 21st Annual Conference on the Architectures and Mechanisms for Language Processing (Valletta, Malta). We thank the attendees for their comments and suggestions. Data and scripts are archived at the Open Science Framework: <https://osf.io/s8zmk/>.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the Departments of Psychology and Educational Psychology at the University of Illinois at Urbana-Champaign via funding from NIH training grant T32-HD055272; National Institute of Child Health and Human Development.

References

- 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.

- Alegre, M., & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language*, 40, 41–61.
- Allen, M., Badecker, W., & Osterhout, L. (2003). Morphological analysis in sentence processing: An ERP study. *Language and Cognitive Processes*, 18(4), 405–430.
- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, 4(4), 439–461.
- Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language*, 37(1), 94–117.
- Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, 45(13), 2883–2901.
- Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., ... Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, 10(2), 344–380.
- Bock, K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, 31(1), 99–127.
- Bock, K., & Eberhard, K. M. (1993). Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes*, 8(1), 57–99.
- Brehm, L., & Bock, K. (2013). What counts in grammatical number agreement? *Cognition*, 128(2), 149–169.
- Brehm, L., & Bock, K. (2017). Referential and lexical forces in number agreement. *Language, Cognition and Neuroscience*, 32(2), 129–146.
- Brehm, L., & Goldrick, M. (2017). Distinguishing discrete and gradient category structure in language: Insights from verb-particle constructions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(10), 1537–1556.
- Bürkner, P. C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28.
- Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes*, 42(2), 205–238.
- Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, 33(4), 497–505.
- Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). London: Academic Press.
- Damian, M. F., Vigliocco, G., & Levelt, W. J. (2001). Effects of semantic context in the naming of pictures and words. *Cognition*, 81(3), B77–B86.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466.
- Davies, M. (2008). *The corpus of contemporary American English: 520 million words, 1990-present*. Retrieved from <http://corpus.byu.edu/coca/>
- Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634), 20120394.
- Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, 69(2), 85–103.
- Eberhard, K. M. (1999). The accessibility of conceptual number to the processes of subject-verb agreement in English. *Journal of Memory and Language*, 41(4), 560–578.
- Footo, R., & Bock, K. (2012). The role of morphology in subject-verb number agreement: A comparison of Mexican and Dominican Spanish. *Language and Cognitive Processes*, 27(3), 429–461.
- Franck, J., Vigliocco, G., Antón-Méndez, I., Collina, S., & Frauenfelder, U. H. (2008). The interplay of syntax and form in sentence production: A cross-linguistic study of form effects on agreement. *Language and Cognitive Processes*, 23(3), 329–374.
- Fyndanis, V., Arcara, G., Christidou, P., & Caplan, D. (2018). Morphosyntactic production and verbal working memory: Evidence from Greek aphasia and healthy aging. *Journal of Speech, Language, and Hearing Research*, 61, 1171–1187.
- Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Attention, Perception, & Psychophysics*, 45(3), 189–195.
- Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes*, 21(1-3), 181–204.
- Haskell, T. R., & MacDonald, M. C. (2003). Conflicting cues and competition in subject-verb agreement. *Journal of Memory and Language*, 48(4), 760–778.
- Haskell, T. R., Thornton, R., & MacDonald, M. C. (2010). Experience and grammatical agreement: Statistical learning shapes number agreement production. *Cognition*, 114(2), 151–164.
- Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, 16(2), 96–101.
- Humphreys, K. R., & Bock, K. (2005). Notional number agreement in English. *Psychonomic Bulletin & Review*, 12(4), 689–695.
- Hussey, E. K., Harbison, J. I., Mishler, A., Teubner-Rhodes, S. E., Velnoskey, K., & Novick, J. M. (2017). Memory and language improvements following cognitive control training. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(1), 23–58.
- Hussey, E. K., & Novick, J. M. (2012). The benefits of executive control training and the implications for language processing. *Frontiers in Psychology*, 3, 158.
- Hussey, E. K., Ward, N., Christianson, K., & Kramer, A. F. (2015). Language and memory improvements following tDCS of left lateral prefrontal cortex. *PLoS ONE*, 10(11), e0141417.
- Jäger, L. A., Engelmann, F., & Vasisht, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, 94, 316–339.
- January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Colocalization of Stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience*, 21, 2434–2444.
- Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 824–843.

- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111(2), 228–238.
- Kay, M. (2019). tidybayes: Tidy data and Geoms for Bayesian models. doi:10.5281/zenodo.1308151 (Retrieved from <http://doi.org/10.5281/zenodo.1308151>), R package version 1.0.4 <http://mjskay.github.io/tidybayes/>
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30(5), 580–602.
- Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language*, 82, 133–149.
- Lenth, R. (2019). emmeans: Estimated marginal means, aka least-squares means. R package version 1.3.3. Retrieved from <https://CRAN.R-project.org/package=emmeans>
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419.
- Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10), 447–454.
- Lorimor, H., Jackson, C. N., Spalek, K., & van Hell, J. G. (2016). The impact of notional number and grammatical gender on number agreement with conjoined noun phrases. *Language, Cognition and Neuroscience*, 31(5), 646–661.
- MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, 24(1), 56–98.
- Mainz, N., Shao, Z., Brysbaert, M., & Meyer, A. S. (2017). Vocabulary knowledge predicts lexical processing: Evidence from a group of participants with diverse educational backgrounds. *Frontiers in Psychology*, 8, 1164.
- Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *Neuroimage*, 59(2), 1859–1869.
- McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48(1), 67–91.
- Mirković, J., & MacDonald, M. C. (2013). When singular and plural are both grammatical: Semantic and morphophonological effects in agreement. *Journal of Memory and Language*, 69(3), 277–298.
- Nicenboim, B., & Vasishth, S. (2018). Models of retrieval in sentence comprehension: A computational evaluation using Bayesian hierarchical modeling. *Journal of Memory and Language*, 99, 1–34.
- Nicenboim, B., Vasishth, S., Engelmann, F., & Suckow, K. (2018). Exploratory and confirmatory analyses in sentence processing: A case study of number interference in German. *Cognitive Science*, 42(Suppl. 4), 1075–1100.
- Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 5(3), 263–281.
- Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, 41(3), 427–456.
- Perception Research Systems. (2007). Paradigm stimulus presentation. Retrieved from <http://www.paradigmexperiments.com>
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329–347.
- Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, 23(7–8), 942–971.
- R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Reifegerste, J., Hauer, F., & Felser, C. (2017). Agreement processing and attraction errors in aging: Evidence from subject-verb agreement in German. *Aging, Neuropsychology, and Cognition*, 24(6), 672–702.
- Schreuder, R., & Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language*, 37(1), 118–139.
- Slevc, L. R., & Martin, R. C. (2016). Syntactic agreement attraction reflects working memory processes. *Journal of Cognitive Psychology*, 28(7), 773–790.
- Smith, G., Franck, J., & Tabor, W. (2018). A self-organizing approach to subject-verb number agreement. *Cognitive Science*, 42(Suppl. 4), 1043–1074.
- Snodgrass, J. G., & Mintzer, M. (1993). Neighborhood effects in visual word recognition: Facilitatory or inhibitory? *Memory & Cognition*, 21(2), 247–266.
- Snyder, H. R., & Munakata, Y. (2008). So many options, so little time: The roles of association and competition in underdetermined responding. *Psychonomic Bulletin & Review*, 15(6), 1083–1088.
- Sorensen, T., Hohenstein, S., & Vasishth, S. (2016). Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology*, 12(3), 175–200.
- Strijkers, K., & Costa, A. (2011). Riding the lexical speedway: A critical review on the time course of lexical selection in speech production. *Frontiers in Psychology*, 2, 356.
- Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, 136(1), 64–81.
- Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language*, 76, 195–215.
- van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, 38(4), 584–589.
- Vandierendonck, A., Loncke, M., Hartsuiker, R. J., & Desmet, T. (2017). The role of executive control in resolving grammatical number conflict in sentence comprehension. *The Quarterly Journal of Experimental Psychology*, 71, 759–778.
- Van Dyke, J. A., & Johns, C. L. (2012). Memory interference as a determinant of language comprehension. *Language and Linguistics Compass*, 6(4), 193–211.
- Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, 131(3), 373–403.
- Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, 65(3), 247–263.
- Veenstra, A., Antoniou, K., Katsos, N., & Kissine, M. (2018). Resisting attraction: Individual differences in executive

- control are associated with subject–verb agreement errors in production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(8), 1242–1253.
- Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition*, 61(3), 261–298.
- Villata, S., Tabor, W., & Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: Evidence from agreement. *Frontiers in Psychology*, 9, 2.
- Vuong, L. C., & Martin, R. C. (2014). Domain-specific executive control and the revision of misinterpretations in sentence comprehension. *Language, Cognition and Neuroscience*, 29, 312–325.
- Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61(2), 206–237.
- Ye, Z., & Zhou, X. (2009). Conflict control during sentence comprehension: fMRI evidence. *NeuroImage*, 48, 280–290.