Attraction Effects for Verbal Gender and Number Are Similar but Not Identical:

Self-Paced Reading Evidence from Modern Standard Arabic

Matthew A. Tucker[*,a], Ali Idrissi[b], and Diogo Almeida[a]

[*]Corresponding Author. Email: `matt.tucker@nyu.edu`

[a]Language, Mind, and Brain Laboratory

Division of Science

New York University Abu Dhabi

P.O. Box 129188

Abu Dhabi

UAE

[b]Department of English Literature and Linguistics

Qatar University

P.O. Box 2713

Doha

State of Qatar

Highlights:

- Agreement attraction errors in comprehension also affect verbal gender agreement morphology independent of other features/cues.

- Agreement attraction occurs even in the presence of rich inflectional morphology.

- Errors for grammatical gender are larger and occur later than errors for grammatical number.

- Arabic does not show attraction effects in acceptable/grammatical sentences.

Manuscript submitted for publication.

1

Attraction Effects for Verbal Gender and Number Are Similar but Not Identical
Self-Paced Reading Evidence from Modern Standard Arabic

Matthew A. Tucker[a,*], Ali Idrissi[b], Diogo Almeida[a]

[a]*Language, Mind, and Brain Laboratory, Division of Science, NYU Abu Dhabi, P.O. Box 129188, Abu Dhabi, United Arab Emirates*
[b]*Department of English Literature and Linguistics, College of Arts and Sciences, Qatar University, P.O. Box 2713, Doha, State of Qatar*

Abstract

Previous work on the comprehension of agreement has shown that, in many languages, incorrectly inflected verbs do not trigger responses typically seen with fully ungrammatical verbs when the preceding sentential context furnishes a possibly matching distractor noun. We report eight studies, three of which are direct replications, designed to test the character and timing of these errors in comprehension along the dimensions of grammatical gender and number in Modern Standard Arabic. A meta–analysis of the results indicate that, despite a robust verbal gender system which interacts with other inflectional features, Arabic readers show agreement attraction effects in reading comprehension for gender and number on verbs given appropriate preceding contexts with mismatching NPs. Moreover, the meta–analysis demonstrates that these two features do not behave identically in that gender effects are larger and occur later relative to number attraction effects. These results challenge models of agreement attraction that predict agreement features to be equipotent. We discuss how models of agreement errors require modifications in order to account for these differential results.

Keywords: agreement errors, self-paced reading, verbal gender, verbal number, meta–analysis

---

*Corresponding Author
  Email addresses:* matt.tucker@nyu.edu (Matthew A. Tucker), ali.idrissi@gmail.com (Ali Idrissi), diogo@nyu.edu (Diogo Almeida)

1. Introduction

Human language, despite being transmitted serially as a string of words, contains a myriad of relationships between words which can obtain at a distance. Subject–verb agreement as in (1) is one such relationship:

(1) a. **The fencers** *are* divided about the best strategy for the World Championships.
     b. **The fencers** on the French National Team that won a major award last year *are* deeply divided about the best strategy for the World Championships.
     c. **The fencers** on the French National Team that won a major award last year by beating the Italian team in a very hotly contested and important match *are* deeply divided about the best strategy for the World Championships.

In the specific case of (1a), the choice of *the fencers* conditions the subsequent choice of *are* in production or the expectation of a plural verb in comprehension. Subject–verb agreement is particularly important in the study of language and its relationship with the performance systems since it not only involves the very basic building blocks of a clause but also because it is a relationship that can obtain at an unbounded serial distance. This is because subjects can theoretically be separated from their verbs by an infinite amount of material yet still require proper agreement — see (1b,c). Despite this potentially infinite linear distance, subjects and verbs are still relatable by dependency because of their structural positions in the clause. This basic fact underscores an important property of the syntax of human languages: despite of their linear externalization, sentences are internally organized in a hierarchical, and not serial, fashion.

Therefore, from the perspective of real–time language production and comprehension, coping with potentially unbounded dependencies such as subject–verb agreement requires attention to the encoding, maintaining, and retrieving of linguistic units from working memory, as well as the monitoring process that oversees whether the correct relationship between the subject and the verb has been completed. It is a remarkable fact, then, that subject–verb agreement errors are not only sometimes observed (both in language production Bock & Miller, 1991 and comprehension Pearlmutter et al., 1999; Wagers et al., 2009), but that they also seem to be at least partially systematic. Known

as AGREEMENT ATTRACTION, a particularly well-studied subset of these errors are commonly seen when a subject co-occurs with a non-subject argument that appears to be the target of the erroneous number agreement, as in the example in (2) from Dillon et al. (2013):[1]

(2)   *The executive* who oversaw **the middle managers** apparently ***were*** dishonest about the company's profits.
      (Dillon et al., 2013)

The characteristic property of this phenomenon is the illusion of acceptability of *prima facie* unacceptable agreement violations — despite the fact that the plural *were* is ungrammatical in (2), many speakers occasionally both accept and produce such utterances. In production studies such as Bock & Miller (1991) or Franck et al. (2002), these errors surface as incorrect verb productions, whereas in comprehension studies such as Pearlmutter et al. (1999) or Tanner et al. (2014), these errors surface as the absence of behavioral or electrophysiological responses typically associated with the perception of ungrammaticality. Because they represent a systematic exception to the idea that processing is faithful to grammar during the production and comprehension of dependencies, these attraction violations have served as a focal point for much theorizing about the nature of both grammatical agreement and dependency processing.

1.1. Failure of Representation or Failure of Process?

One theoretical approach to illusory dependency licensing in the literature conceives of attraction effects as arising as a function of the dynamics of memory encoding and retrieval of agreement–relevant material in memory. This line of research draws on cue–based retrieval theories of language processing (Lewis & Vasishth, 2005) and extends their logic to agreement dependencies (Badecker & Kuminiak, 2007; Badecker & Lewis, 2007; Wagers et al., 2009). This theory, building upon the observation that dependency resolution is subject to retrieval interference, posits that sentence processing contains instances of working memory retrievals which access long-term memory

---

[1]Here the *(correct) subject* appears in italic face, the **attractor/distractor** NP in bold face, and the ***target region*** in both bold and italic.

4

stores in a parallel, cue-based manner. It is therefore a theory in which the probability that memory chunks are retrieved is a function of the similarity of a given chunk to other items in memory as well as the number of dimensions upon which a chunk matches the cues in the goal of the retrieval event. When more than possible retrieval target matches the goal cues, erroneous retrievals of non-subjects can occur. For explicit modeling of agreement in this system, see Badecker & Kuminiak (2007); Badecker & Lewis (2007); Dillon et al. (2013); Wagers et al. (2009); and Tucker et al. (2015), but what all these models have in common is the notion that agreement attraction is a failure of *process* in the memory retrieval system underwriting language use.

A theory which is often contrasted with the cue-based retrieval model is one in which structural representations themselves can be erroneously represented (see Bock & Eberhard, 1993; Eberhard et al., 2005; Franck et al., 2008; Nicol et al., 1997; Pearlmutter et al., 1999 and the discussion in Engelmann et al., 2015; Wagers et al., 2009). In these models, stochastic fallibility in the encoding or maintenance of structural representations leads to misrepresentation of the true subject's features in terms of relevant features of the distractor, resulting in attraction through the normal processes of subject–verb agreement in the proportion of cases where this erroneous representation obtains. One way to formalize this notion is to say that the representations responsible for maintaining features and syntactic constituents in memory allow for targets of agreement to have their features overwritten in the presence of a distractor with mismatching features. Proponents of these models have advanced several distinct mechanisms for achieving this misrepresentation including degradation of structural representation (Eberhard et al., 2005; Staub, 2009), erroneous feature percolation (Nicol et al., 1997), and fallible feature copying (Franck, 2011; Franck et al., 2008). Here we abstract away from considerable detail to jointly consider these models as sharing the notion that agreement attraction is a failure of *representation* in the processing systems subserving language use.

Despite these differences, there is at least one important dimension along which these two families of theories are similar, namely the way in which they incorporate cross-linguistic differences pertaining to the process of subject–verb agreement. Linguistic theory generally takes person, number, and gender features to be equipotent in

5

agreement phenomena (*e.g.,* Chomsky, 1995; Pollock, 1989; Preminger, 2011; though see Béjar, 2003; Béjar & Rezac, 2009 for a different approach), and processing theories which rely directly upon these linguistic notions predict (at least limited) equivalency in attraction effects for each of these features without having to resort to nontrivial extensions. Because misrepresentation theories attribute attraction to normal mechanisms of feature spreading, differences in attraction strength for different features are only predicted if representational considerations constrain spreading, overwriting, or copying. Cue-based models, on the other hand, posit that cues are typically treated equally by the retrieval system. Any observed difference between how different agreement cues are processed would necessitate positing a more complex view of these cues or how they are weighed or retrieved within the memory system.

With this backdrop, it is therefore important to determine whether the basic assumption shared by the two most popular families of theories of agreement errors — namely that all agreement features are equipotent — is in fact supported by the evidence. While number agreement comprehension is relatively well-studied, considerably less work has been conducted to address gender verbal agreement and the extant data do not provide any consensus on the matter in either production or comprehension (which we discuss below). This is an important issue for both representational and cue-based theories, since both would, all else being equal, take verbal gender to be equivalent to verbal number in attraction terms. In a series of eight comprehension studies in Modern Standard Arabic, we directly test these foundational assumptions by comparing the process of subject–verb gender agreement with subject–verb number agreement, in an effort to document the ways in which these two processes are similar (as predicted or assumed by the two existing theoretical approaches to agreement errors) and the ways in which they are different (and therefore the ways in which they would challenge these theories to revise their assumptions). Because a recent meta-analysis (Jäger et al., In Press) about memory interference effects in agreement processing have both indicated that NUMBER AGREEMENT ATTRACTION is generally small in size (22 ms, CrI [9…36 ms]), and in line with the increasing awareness that much research in psychology, including in psycholinguistics, involves underpowered studies and little replication effort (Jäger et al., 2015; Vasishth & Nicenboim, 2016), which may lead to poor reproducibil-

6

ity of results in any given field of study (Button et al., 2013; Collaboration et al., 2015; Sprouse & Almeida, In Press), here we present five self–paced reading studies with large samples ($N > 100$), as well as direct replications of three of these studies, also employing large sample sizes ($N > 100$). This large number of studies, in addition to prior work in Modern Standard Arabic (Tucker et al., 2015), allows us to perform a meta-analysis of both number and gender attraction effects in the language, a practice that has many advantages compared to the use of simple null hypothesis testing (Hoekstra et al., 2006), is considered the gold–standard method for synthesizing the results of many different studies (Cooper et al., 2009; Cumming, 2014; Hunter & Schmidt, 2004), and has recently been fruitfully applied to psycholinguistic work (Jäger et al., In Press; Mahowald et al., 2016).

1.2. Equivalency in Error Studies

Given the preceding discussion, it is important, theoretically, to understand whether attraction for verbal gender agreement is identical to attraction for verbal number agreement. But how does one assess equivalency between gender and number in subject–verb agreement comprehension? Here we identify five distinct ways in which gender and number could be equivalent in regards to comprehension attraction effects:

(3)   a. EXISTENCE: Do both features participate in attraction?

b. SIZE: Do both features yield similar attraction effect magnitudes?

c. GRAMMATICAL ASYMMETRY: Do both features participate in asymmetries based on grammaticality of the verb?

d. MARKEDNESS ASYMMETRY: Do both features participate in asymmetries based on markedness of the agreeing elements?

e. TIMING: Do both features exhibit attraction effects with the same time-course?

An evaluation of the extant literature reveals that the answers to these questions are muddled when it comes to verbal gender. With respect to the existence of attraction effects and their size (3a, b), Lorimor et al. (2008) reported no gender attraction in

production in Russian, whereas Badecker & Kuminiak (2007) and Malko & Slioussar (To Appear) report gender attraction respectively in production in Slovak and in perception in Russian, but without assessing its comparative effect with the one observed for number. Interestingly, verbal gender agreement in Slavic languages, where present, is restricted only to past tense verbs; verbs in other tenses do not show gender agreement at all. While gender can appear on a verb in Slavic, gender in the nominal system is often conflated with case morphology, an issue which leads to ambiguity that has been known to influence agreement attraction rates and sizes (Badecker & Kuminiak, 2007; Hartsuiker et al., 2003; Häussler & Bader, 2009). Deutsch & Dank (2009, 2011) directly compare gender and number attraction effect sizes in a series of Hebrew production studies and find similar rates of attraction which are nevertheless subject to different morphophonological influences, though only inanimate NPs were tested. It therefore seems clear that verbal gender could in principle be subject to attraction, but whether it is of the same size as the effects for verbal number remains to be determined.

In comprehension studies of subject–verb number agreement attraction, it is often (but not always, see Jäger et al., In Press for review) observed that these effects are asymmetric in nature: attraction effects are easily observable in ungrammatical sentences, but less prominently so in grammatical ones (Wagers et al., 2009; though see also Franck et al., 2015). In addition, Tanner et al. (2014) have found that ERPs to English attraction configurations do not show evidence of morphosyntactic error recognition (defined as an observable P600 effect) in grammatical examples. If agreement processing relies on the same structural representations or the same structure of memory cues regardless of the agreement features, then one would predict that a similar asymmetry would be found for gender subject–verb agreement as well. As far as we are aware, no studies directly address the presence of grammaticality asymmetries (3c) between gender and number subject–verb agreement, though examination of the results in Malko & Slioussar (To Appear) show that a grammaticality asymmetry is present with Russian verbal gender. As Wagers et al. (2009) note, comprehension studies assessing the grammaticality asymmetry are particularly important in this regard as (i) the traditional preamble completion task generally used in production studies Bock & Miller (1991) cannot provide evidence about attraction effects in ultimately grammatical pro-

8

ductions (though see Staub, 2009 for a variation argued to be able to do this), and (ii) that cue–based retrieval models seem to be better able to accommodate this effect when compared to misrepresentation models, thereby providing a potential empiric diagnostic that favors one family of theories over the other (Jäger et al., In Press; Tanner et al., 2014; Wagers et al., 2009).

Identity of markedness (3d) is the best-understood of the five dimensions of comparisons along which number and gender subject–verb agreement can be evaluated, and it is the topic of considerable discussion (see Badecker & Kuminiak, 2007; Deutsch & Dank, 2011; and Malko & Slioussar, To Appear). By "markedness," here we mean the asymmetry observed by Bock & Miller (1991) wherein certain feature combinations in pre-verbal material cause more attraction errors than others (as in English, where *The key to the cabinets…* induces many more attraction errors than *The keys to the cabinet…*; see also Eberhard, 1997). The Slovak and Russian studies (Badecker & Kuminiak, 2007; Malko & Slioussar, To Appear) advance the conclusion that markedness in a three-valued system with masculine, feminine, and neuter should be defined in a pairwise fashion such that there is no *gestalt* markedness hierarchy but rather individual relations between masculine, feminine, and neuter. However, this is at odds with results from Russian number, as Malko & Slioussar (To Appear) note, since the results of the latter suggest that number attraction profiles are not prevalent for plural subjects with singular distractors (*i.e.*, there is no markedness asymmetry). In Hebrew, on the other hand, markedness has been shown to affect production of number attraction errors but not gender (Deutsch & Dank, 2011). To our knowledge, there are no studies assessing the phenomenon of markedness in comprehension other than Malko & Slioussar (To Appear), with all other data coming from production studies.

Finally, it is also possible to ask whether or not gender and number attraction effects occur in similar time-courses (3e) — both with respect to global grammaticality effects (Lago et al., 2015; Staub, 2009, 2010) and each other. Given the preponderance of production studies in the existing literature, this question has not been properly addressed to date.

It is also worth noting that much research in the literature on attraction has attempted to examine the locus of morphophonological influences (including markedness) on at-

9

traction rates to one or more of the constituents involved in the attraction process —
the head noun/subject, the local noun/distractor, or the verb (*e.g.*, Dank & Deutsch,
2010; Deutsch & Dank, 2011; Hartsuiker et al., 2003; Malko & Slioussar, To Appear;
Vigliocco et al., 1995). This is a fruitful line of inquiry, but one which is orthogonal to a
more basic concern: are there asymmetries in attraction rates to different combinations
of features on heads, local nouns, and verbs? Whether these effects are attributable to
the influence of the head or local noun *per se* is impossible to evaluate in a binary sys-
tem like that found with English number or Arabic gender. Thus, while it is sensible to
talk about the influence of markedness or morphophonological ambiguity on only the
head or local noun, in trying to evaluate whether there is a basic equivalence among
agreement features, one must first establish whether any asymmetry is present based
on morphophonological or conceptual features before one can examine the importance
of the locus of these features. We therefore retain the term "markedness" from the lit-
erature for these asymmetries but attempt, where possible, to abstract away from tying
the asymmetry to particular morphosyntactic positions.

In the same vein, it is worth noting that grammatical gender does appear as the focus
of a large number of agreement attraction studies, such as Antón-Méndez et al. (2002);
Dank & Deutsch (2010); Deutsch & Dank (2009, 2011); Franck et al. (2008); Vigliocco
& Franck (1999, 2001); Vigliocco & Zilli (1999); and Fuchs et al. (2015), to name just
a few, but these works examine adjective–noun gender co-occurrence restrictions. In
the nominal domain, the combined results of these studies seems to support the idea
that attraction in gender nominal agreement proceeds in similar ways as attraction in
number nominal agreement. We take this point to be well-established in the production
literature but note that there is little reason to treat verbal and nominal agreement as
being, in principle, the same sort of process (for theoretical discussion, see Baker, 2008
and Norris, 2014; for a discussion within the psycholinguistic literature, see *e.g.*, Tanner
et al., 2014). Most crucially for our purposes here, while subject-verb agreement is
potentially unbounded, adjective-noun agreement is by definition extremely local and
not a potentially unbounded dependency.

1.3. The Present Study: Context

The experiments reported here attempt to address several outstanding questions about the equivalency of verbal gender and number with respect to the five dimensions laid out above: whether they both elicit attraction effects of the same size and time course, whether they do so asymmetrically with respect to language-internal marked-ness considerations, and whether they do so asymmetrically with respect to grammaticality. In following this reasoning, our language of study, Modern Standard Arabic (MSA), provides several important *desiderata* for studies of verbal gender (Ryding, 2005): (1) the presence of verbal gender agreement on all verbs in the language (not restricted to a given tense, as in Slavic); (2) the appearance of gender marking on nominals independent of case morphology (also unlike Slavic), allowing the examination of gender independently of the influence of case; (3) a demonstrated number attraction effect in comprehension against which to compare results from gender (Tucker et al., 2015); and (4) a close typological relationship to Hebrew, allowing comparison of our results with the production studies of Deutsch & Dank (2009, 2011), and Dank & Deutsch (2010). Finally, the last five experiments (3, 4A, 4B, 5A, 5B) simultaneously attempt to replicate and expand upon findings from Tucker et al. (2015) that the type of plural marking on attractor NPs matters for agreement attraction effect sizes. These latter five experiments also provide a number contrast to the gender effects reported in Experiments 1, 2A, 2B, 5A and 5B in order to assess similarities and differences in attraction rates, susceptibility to grammatical and markedness asymmetries, as well as time courses.

1.3.1. A note about the presentation of the data

For this series of eight studies, we opt to present the results using estimation of means and confidence intervals, instead of the more traditional *null hypothesis statistical testing* (NHST) framework, in keeping with the goal of conducting a meta–analysis of the results at the end. In this, we follow the advice from a number of statistical reformers (Cumming, 2014 for review), including the Task force on Statistical Inference of the American Psychological Association (Wilkinson, 1999). These reformers argue that use of NHST (i) is severely marred by systematic misinterpretations in prac-

tice (cf. Nickerson, 2000 and Greenland et al., 2016 for review), even by established researchers (Badenes-Ribera et al., 2015, 2016; Haller & Krauss, 2002; Oaks, 1986) and (ii) leads to an overly dichotomous assessment of the evidence from single studies whereby (iii) results are either discarded as false if they fail to pass the significance test (Hoekstra et al., 2006) or (iv) are believed with unreasonable confidence to be true if they pass it (Gigerenzer, 2004), which feeds into (v) wild overestimations of the likelihood of successful replication for a statistically significant result (Cumming, 2008; Cumming & Maillardet, 2006; Francis, 2012; Stanley & Spence, 2014). These shortcomings in the practical use of NHST may lead to publication bias (for e.g., Kühberger et al., 2014), as well as leave researchers ill–equipped to assess evidence cumulatively (Cooper et al., 2009; Francis, 2012; Hedges & Olkin, 1980; Hunter & Schmidt, 2004; Rosenthal & Dimatteo, 2001; Schmidt, 1996).

Given the large number of studies we report here, which include three direct replication attempts, sometimes with apparently conflicting results, we concur with these reformers that a traditional narrative summary of the findings based on counting the number of statistically significant vs non–statistically significant results (i.e., "vote counting", cf. Hedges & Olkin, 1980; Rosenthal & Dimatteo, 2001) would present a biased and misleading picture of the combined evidential value of the studies. This issue is compounded with concerns about statistical power, since our measure of interest, the *attraction effect*, is generally detected through an interaction term in two– or three–way factorial designs, which can have much lower power compared to main effects depending on the nature of the data (e.g., Potvin & Schutz, 2000). Even though we attempted to mitigate this issue by using comparatively large samples sizes (all Ns > 100), it is unclear the level of increase in statistical power that these sample sizes produced.

Therefore, instead of presenting the results of a statistical significance test in each experiment, we will present the results graphically, together with an estimation of the effect sizes of interest (the *attraction effect* and the *grammaticality effect*), and their 95% confidence intervals, calculated via the BCa Bootstrap (with 2000 replications per estimate; cf. Efron, 1987; Kirby & Gerlanc, 2013). However, the results of more traditional statistical tests are presented in the supplementary materials online, for the

interested reader. In addition, in the interest of contributing to the effort of increasing reproducibility and replicability of findings in the cognitive sciences, the full data, as well as the analysis scripts, are publicly available at LINK-TO-DATA-PACKAGE-AT-OSF-OR-FIGSHARE.

1.3.2. General structure of the experiments

In all experiments, we have two effects of interest, namely the GRAMMATICALITY effect, and the ATTRACTION effect. The GRAMMATICALITY effect will be quantified for each participant by summing their average reaction times for the ungrammatical sentences (i.e., those sentences in which the Subject mismatches the Verb in the agreement feature of interest, either NUMBER or GENDER) and subtracting the sum of their average reaction times for the grammatical sentences from it. The ATTRACTION effect will be quantified separately for *grammatical* and *ungrammatical* sentences, since prior work has noticed that ATTRACTION effects have a tendency to occur in *ungrammatical* sentences alone (the so–called *grammaticality asymmetry*, cf. Jäger et al., In Press; Wagers et al., 2009). In every experiment, we code whether the agreement feature of interest on the Attractor NP matches that of the Verb (our MATCH vs NOMATCH conditions). In this coding scheme, the ATTRACTION effect is a subtraction of the average reading time from NOMATCH condition from the MATCH condition, *within each level of* GRAMMATICALITY. Thus, the attraction effect in *ungrammatical* sentences is quantified by the subtraction of the average time of MATCH/UNGRAMMATICAL condition (a sentence which is simply ungrammatical, with no viable attractor NP) from the NO-MATCH/UNGRAMMATICAL condition (a sentence which is ungrammatical, but in the presence of a viable attractor NP). A full example of this coding scheme is shown in Table 1.3.2.

In addition, the direct replications (experiments 2B, 4B and 5B) were conducted several months after the data from the first five experiments had been conducted and statistically analyzed. Because of this, they are presented here as proper direct replications, and not just the same experiment with an increased sample size.

13

Condition	NP1 R1	Comp R2	RCV R3	NP2 R4	Adv R5	V R6	Continuation R7–R$_n$
Match/Gram	The translator	who	helped	the president	often	speaks	....
Match/Ungram	The translator	who	helped	the president	often	speak	...
NoMatch/Gram	The translator	who	helped	the presidents	often	speaks	...
NoMatch/Ungram	The translator	who	helped	the presidents	often	speak	...

Table 1: Example of the condition coding that will be used throughout the experiments, using NUMBER as the agreement feature of interest. Example from Tucker et al. (2015).

2. Experiment 1

Experiment 1 was designed to assess whether attraction for grammatical gender obtains in MSA comprehension using a self-paced reading methodology. In the experimental agreement attraction literature, several grammatical structures are used with some regularity: (i) PP modifiers (Bock & Miller, 1991), (ii) object relative clauses (Bock & Miller, 1991; Wagers et al., 2009), and (iii) infinitival complements (Tucker & Wagers, 2010). In this study, however, we employ subject relative clauses (see, *e.g.,* Bock & Miller, 1991; Dillon et al., 2013) modifying a sentence-initial subject. One Arabic-internal and one Arabic-external consideration each guide the choice of subject relative clauses. Firstly, Wagers et al. (2009) have shown that spillover effects in agreement attraction studies can inadvertently impact measurements at critical verbs when the immediately previous region is manipulated experimentally (see also Jäger et al., In Press). A common solution to the need to manipulate the features/cues of an attractor is to insert an adverb between the attractor and critical verb. In Arabic, adverbs are not commonly found preverbally in Subject — Verb — Object word orders (Tucker, 2011). Adverbs and adverbial PPs are preferred clause-finally, making subject relative clauses an appropriate choice given the availability of a relative-clause final parse of adverbs appearing preverbally. Secondly, number agreement attraction in this configuration has already been studied in MSA (Tucker et al., 2015), allowing for direct comparison of the reaction time profiles of grammatical number and gender processing.

Given the prevailing theoretical and experimental conceptions of agreement, one expects to find that attraction should be possible for gender. In the formal syntactic literature, agreement is often taken to be a uniform process which simultaneously en-

14

compasses the features of gender, number, and person (to wit, the theory of AGR projections originating from Pollock, 1989 and the more modern notion of AGREE from Chomsky, 1995, 2000, 2001; *et seq.*). Furthermore, both misrepresentation models and cue-based retrieval models require added mechanics to differentiate cues for number and gender, meaning that gender should, if isolated properly, behave similarly to number in comprehension. Furthermore, if attraction is a product of the *process* of resolving agreement dependencies, then we do not expect to find attraction profiles in grammatical sentences, in line with the claims of cue-based models. If attraction is due to *fallibility* in representation of gender, on the other hand, we expect to find no differential attraction effect owing to grammaticality.

2.1. Participants

Participants were 104 native speakers of Arabic from the United Arab Emirates University (UAEU) student body with no history of language disorders and self-assessed proficiency with MSA (104 females; mean age 20.4 years).[2] All participants provided informed consent and were compensated monetarily for their time. This and all other studies reported here were approved by the NYU Abu Dhabi Institutional Review Board and the UAEU Ethics Committee.

2.2. Materials & Predictions

In order to assess the possibility of gender attraction in MSA, a set of 48 sentences containing a subject modified by a subject relative clause were constructed based upon the stimuli created for the experiment reported in Tucker et al. (2015). All the sentences were of the structure *NP1 — Complementizer — [Verb — NP2 — Adverb] — **Target Verb** — Continuation*, where NP1 is the grammatically accessible subject and NP2 the attractor NP for agreement realized on the target verb. In MSA, however, there are comparatively few adverbs, and so in some cases adverbial PPs which comprised a sin-

---

[2]The gender composition of our participant sample is due to the nature of instruction at the UAEU — there are separate campuses for male and female students, and all participant testing was conducted on the female campus.

gle orthographic word were used instead. A complete list of stimuli for this experiment appear in 8.4.

MSA has a grammatical case marking system which interacts with orthography in nontrivial ways (Ryding, 2005, 165–204). Given that morphological case marking is known to influence attraction effects for both gender (Badecker & Kuminiak, 2007; Lorimor et al., 2008) and number (Häussler, 2009), NP1 and NP2 were both selected to be morphologically definite (marked with the definite proclitic /al-/) so that they belonged to declension classes that did not involve case marking morphology with orthographically long vowels (Ryding, 2005, 182–204). This is an important *desideratum* because indefinite nouns in Arabic mark morphological accusative case distinct from nominative, allowing participants to potentially disambiguate subjects and objects using this case-marking.

The result of these two constraints is that morphological case on NP1 and NP2 is expressed by short vowel diacritics which are not typically written in Arabic. Our stimuli therefore abstract away from the effects of morphologically overt case marking by not writing these short vowels, a convention which matches everyday written text in the Arab world. In fact, we matched this convention across the entire study: short vowels and other diacritics were only written in our stimuli when they would be written in everyday Arab print media. This is usually because a single vowel would disambiguate two orthographically ambiguous words or be more common by convention (*i.e.,* the *tanwiin* on adverbs/adverbial PPs). For example, the MSA word سابقاً/*saabaqaan* ("previously") was written with the final diacritic to avoid confusion with the string سابقا being read incorrectly as *saabaqaa* ("they.DUAL raced"). An example sentence from these 48 stimuli is shown in (4):

(4)   المترجم الذي ساعد المدير أحياناً يتكلم خمس لغات بفصاحة.

Ɂal-mutarʒim-u   *Ɂallaðii*   saaʕad-a   **Ɂal-mudiir-a**
the-translator-NOM COMP.MASC.SG helped-3.MASG.SG **the-manager-**ACC

*Ɂaħjaanan **ja-takallamu***   xamsata luɣaat-in   bi-fasˤaaħatin.
often   3.SG.MASC-***speaks*** five   languages-ACC with-fluency

"*The translator* who helped **the manager** often ***speaks*** five languages fluently."

In addition to the requirements discussed above, several other constraints were also placed on the creation of stimuli sentences: Firstly, the relative clause verbs were chosen such that they either took a bare NP complement or a PP complement headed by a preposition which is orthographically encliticized to the relative clause direct object in order to ensure that all stimuli had the same number of words up to the main clause target verb. Secondly, Arabic has two distinct morphological tenses which are marked on verbs in part by distinct agreement affixes (Ryding, 2005, 439–44). In order to abstract away from the individual contributions of distinct tense/agreement affixes, the main clause target verbs were counterbalanced for the two tenses, perfect (*e.g.,* تكلم/*takallam*, "he spoke") and imperfect (*e.g.,* يتكلم/*ja-takallam*, "he speaks"). In all the experiments we report here, this tense distinction had no discernible effect on reading times or attraction effects.

With the NP subject and attractor, nouns were chosen which had a masculine stem which could be made feminine solely by addition of the feminine singular nominal suffix /-a/ (orthographic ة-) — in MSA these are easiest to find in the domain of nouns which denote human occupations. While MSA does have nouns which are feminine without the presence of this suffix, restriction to these nouns was employed in order to abstract away from possible differences in the processing of nominal gender owing to whether or not the feminine gender was an inherent property of the stem versus the contribution of an overt suffix (Sicuro Corrêa et al., 2004). Moreover, the choice of an overtly suffixing feminine allows a straightforward comparison between the processing of gender in MSA and suffixal plural morphology in other languages. The result is also a set of stimuli where grammatical gender is morphologically expressed in ways similar to English pluralization with /-s/, for example.

For each experimental sentence, four variants were constructed by systematically varying the grammatical gender of the attractor (NP2) and the main clause verb (target verb). These manipulations are coded as MATCH (does the grammatical subject match the attractor in grammatical gender value?) and GRAMMATICALITY (does the grammatical subject match the verb in grammatical gender?). Note that in this design, NOMATCH conditions are conditions with feminine attractors, since all subjects are masculine. Both relevant NPs remained in the singular throughout the experiment to

17

assess the effect of gender alone. This resulted in four experimental conditions per stimulus; a complete set of four such sentences appears in Table 2.3.

The 48 sets of four sentences were distributed across four lists in a Latin Square design after being combined with 144 grammatical fillers of similar length for a 3:1 filler-to-item ratio. None of the fillers included the relative clause construction used in the experimental stimuli or any construction which drew attention to meaningful alternations in verbal agreement. In the final version of each list, only the experimental sentences contained ungrammaticalities, with 12.5% of the sentences in each list ungrammatical.

In this experiment, two predictions are of interest. First, ungrammatical verbal agreement is widely known to engender slower reading times, and we therefore expect an effect of GRAMMATICALITY at the main clause/target verb region (and possibly in subsequent spillover regions). Additionally, if attraction for grammatical gender in MSA occurs at all, then one also expects to find an additional effect, but how attraction should manifest is different for misrepresentation and cue-based retrieval theories. If cue-based retrieval theories are correct in asserting that attraction is not equivalent for grammatical and ungrammatical sentences, then one expects an interaction effect of GRAMMATICALITY and MATCH at the target verb (or in spillover) owing to a slowdown of smaller magnitude in the NoMATCH/UNGRAM condition as compared to the MATCH/UNGRAM condition. Alternatively, one could view this expectation as an erroneous *facilitation* relative to the ungrammatical baseline in the MATCH/UNGRAM condition. On the other hand, if misrepresentation of gender were the operative theoretical mechanism, then one would expect to find only a main effect of MATCH and no interaction (i.e., attraction effects of similar magnitudes for both grammatical and ungrammatical sentences).

2.3. Procedure

Subjects were seated comfortably up to eight at a time at a table in a quiet room in front of Apple iMac computers running Windows 7 natively via a Boot Camp partition on which the experimental software had been pre-loaded. Sentences were presented using the Linger software (Rhode, 2003) in a self-paced word-by-word moving window

18

Condition	NP1 R1	Comp R2	RCV R3	NP2 R4	Adv R5	V R6	Continuation R7–R_n
MATCH/GRAM	المترجم	الذي	ساعد	المدير	أحياناً	يتكلم	خمس لغات بفصاحة.
	The translator (MASC)	who	helped	the manager (MASC)	often	speaks (MASC)	five languages fluently.
MATCH/UNGRAM	المترجم	الذي	ساعد	المدير	أحياناً	تتكلم	خمس لغات بفصاحة.
	The translator (MASC)	who	helped	the manager (MASC)	often	speaks (FEM)	five languages fluently.
NOMATCH/GRAM	المترجم	الذي	ساعد	المديرة	أحياناً	يتكلم	خمس لغات بفصاحة.
	The translator (MASC)	who	helped	the manager (FEM)	often	speaks (MASC)	five languages fluently.
NOMATCH/UNGRAM	المترجم	الذي	ساعد	المديرة	أحياناً	تتكلم	خمس لغات بفصاحة.
	The translator (MASC)	who	helped	the manager (FEM)	often	speaks (FEM)	five languages fluently.

Table 2: A complete item set for one stimulus in Experiment 1.

paradigm (Just et al., 1982). Each trial began with the display of a screen containing the sentence masked by dashes (including spaces and punctuation). Each time the participant pressed the space bar, a single word was revealed and the previous word re-masked with no look-back allowed. All items were presented in the Courier New Arabic font in 28pt bold type. A yes/no comprehension question followed each sentence, appearing on the screen all at once. Comprehension questions were designed in such a way that the answer could be provided independent of experimental manipulations — no questions asked about the attractor NP or the main clause verb. None of our comprehension questions required lexical elaboration of the item or difficult semantic processing. A majority of the comprehension questions asked about the relative clause verb or the post-critical region continuation. As an example, the item *The student who saw the professor(s) yesterday studied electrical engineering at the university* was followed by the question *Did the student study electrical engineering?* Participants responded via a dual Arabic/English keyboard where the 'f/ب' key was used for "yes (نعم)" and the 'j/ت' key used for "no (لا)." Onscreen feedback was provided for both correct and incorrect answers. Participants were instructed to read at a natural pace ensuring comprehension and were not alerted to the presence of grammatical errors in the stimuli, but they were warned that sentences read out of context might seem pragmatically odd. The order of sentence presentation within each list was randomized for each participant. Four practice items were presented before the start of the experiment, one of which was ungrammatical and three of which were followed by a question.

2.4. Analysis

All data were analyzed in the R statistical software platform (R Core Team, 2015). We adopted as a subject-exclusion criterion answering less than 50% of the comprehension questions correctly.

For reaction time data, only data from sentences in which the comprehension question was answered correctly were included for analysis. Previous work attentive to the contribution of different portions of the reaction time distribution to agreement attraction configurations has shown that the canonical comprehension attraction effects are contained disproportionately in the right tail of reading times in regions where effects

20

⁴⁹⁵ exist (see Lago et al., 2015; Staub, 2009, 2010; and Tucker et al., 2015). Therefore, we
⁴⁹⁶ deliberately chose a conservative method of by-region outlier treatment: Winsorization
⁴⁹⁷ at 1% of the by-region mean (see Ratcliff, 1993 for discussion). No other exclusion
⁴⁹⁸ criteria were used.

⁴⁹⁹ *2.5. Results*

⁵⁰⁰ *2.5.1. Comprehension Question Accuracy*

⁵⁰¹   No participants met the criterion for exclusion due to low comprehension question
⁵⁰² accuracy for this experiment. Overall comprehension question accuracy across all sub-
⁵⁰³ jects was 88.5% for all items, with an accuracy of 90.2% for fillers and 83.4% for ex-
⁵⁰⁴ perimental items. The accuracy for matching {subject, attractor} sentences was 86.8%
⁵⁰⁵ (95% CI = 84.8%-88.6%) with grammatical verbs and 83.8% (95% CI = 81.5%-85.7%)
⁵⁰⁶ with ungrammatical verbs. Accuracy for non-matching {subject, attractor} sentences
⁵⁰⁷ was 86.8% (95% CI = 84.7%-88.6%) with grammatical verbs and 76.3% (95% CI =
⁵⁰⁸ 73.8%-78.7%) with ungrammatical verbs.

⁵⁰⁹ *2.5.2. Self-Paced Reading*

⁵¹⁰   Only the sentences for which the comprehension question was answered correctly
⁵¹¹ were included for subsequent analysis of the self-paced reading data. This resulted
⁵¹² in the exclusion of 12.80% of the raw collected data (across all conditions, regions,
⁵¹³ and participants). Mean reading times for each region and condition in Experiment 1
⁵¹⁴ appear in Figure 1. The GRAMMATICALITY effect and the ATTRACTION effects were
⁵¹⁵ calculated as described in section 1.3.2, and the results are presented in Table 3. There
⁵¹⁶ were substantial GRAMMATICALITY effects in the *Verb* as well the two subsequent
⁵¹⁷ regions (54 ms, 127 ms, and 59 ms, respectively). However, evidence for ATTRACTION
⁵¹⁸ effect was only observed for *ungrammatical* sentences, and in the *Verb+1* region (21
⁵¹⁹ ms). The 95% CI of the latter effect, however, did not technically exclude zero, but its
⁵²⁰ lower boundary was -1 ms ($-.9$ ms to be precise). Grand averages of the raw reading
⁵²¹ times for the critical verb and first and second spillover regions appear in Appendix B.

$N = 104$	**Verb**		**Verb+1**		**Verb+2**	
Attraction Ungrammatical	-4	(-29, 20)	*21*	*(-1, 43)*	-4	(-21, 10)
Attraction Grammatical	-4	(-22, 14)	-4	(-16, 8)	2	(-8, 11)
Grammaticality	**54**	**(20, 93)**	**127**	**(96, 162)**	**59**	**(43, 78)**

Table 3: Results of experiment 1. Mean RT for each effect of interest. 95% Confidence Intervals computed by BCa bootstrap (2000 replications) in parenthesis. Effects in which the CI excludes zero are marked in bold. Effects in which the CI includes zero up to ±1 ms are marked in italic.
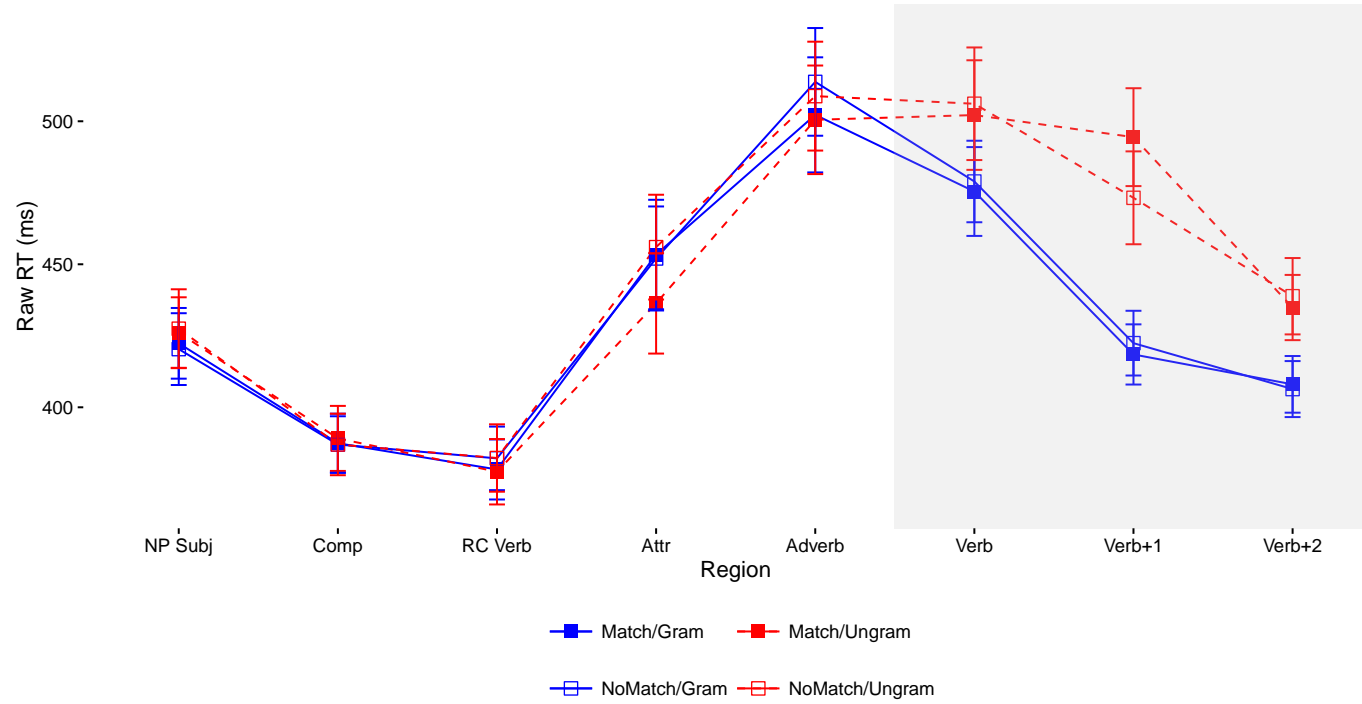
Figure 1: Mean raw reading times from Experiment 1 for all conditions and regions. Error bars represent the standard error of the condition mean across participant averages.

2.6. Discussion

The results of Experiment 1 provide some evidence that that GENDER, like NUM-BER, can be confusable in comprehension, assuming that the reaction time profiles observed here correspond to illusory licensing. Interestingly, the effect size here (21 ms) is virtually identical to the estimate from a recent meta–analysis for NUMBER ATTRAC-TION (22 ms, Jäger et al., In Press). The longer reading times to regions including and following the main clause verb suggest that readers notice verb ungrammaticalities on the whole, spending longer time attempting to resolve the conflicting agreement information. However, relative to the baseline match condition, sentences in which an erroneously feminine verb was preceded by a feminine relative clause object that mismatched the true grammatical subject showed a marginally *reduced* reading time increase. Alternatively, one can view this as a relative *facilitation* of reading times in an otherwise ungrammatical string. Either way, this interaction is the hallmark of agreement attraction effects in comprehension (see Dillon et al., 2013; Lago et al., 2015; Pearlmutter et al., 1999; Wagers et al., 2009; Tucker et al., 2015 and references therein) and plausibly interpretable as illusory licensing of ungrammatical verbs in some cases. Moreover, these effects with gender are not seen in equal measure with *grammatical* verbs. The lack of a reading time difference between masculine and feminine attractors in the grammatical conditions adds to the growing body of literature supporting the idea that attraction effects in self-paced reading comprehension are limited to *ungrammatical* contexts (Lago et al., 2015; Tanner, 2011; Tanner et al., 2014; Wagers et al., 2009; and Tucker et al., 2015).

It is important to emphasize that the finding of attraction for any agreement feature/cue in Arabic is striking given the relative inhospitality of Arabic to misrepresentations in agreement morphology. In our experimental stimuli, for instance, not only are attractor NPs overtly marked with a feminine suffix in the mismatch cases, both the relativizing complementizer *ʔallaðii* (الذي) and the embedded clause verb contained overt morphology matching the correct subject. It seems untenable, therefore, to hold that comprehenders of MSA are more or less susceptible to attraction effects given the prevalence of agreement morphology in the language or a repeated reinforcement of the correct subject during the unfolding of a complex relative clause structure. All of this is

24

true over and above any effect of relative clauses in general (see Bock & Miller, 1991 against Gillespie & Pearlmutter, 2013). This is an important cross-linguistic addition to the conclusions reached by Lago et al. (2015), for instance, that attraction effects are universal in character.

One caveat about the finding of agreement attraction for gender is that, unlike the reading time increases seen with ungrammatical verbs in general, the attraction effect is not present at the target verb in the main clause. Instead, the effect is delayed one region immediately downstream in the spillover. Since the spillover regions were not altered across conditions in a single item set, this difference must be a delayed effect of the gender mismatching agreement morphology encountered in the previous region. It is not uncommon for effects in self-paced reading to appear downstream from the point in the strings where the effect is first possible — in the seven experiments in Wagers et al. (2009), for instance, two of them show results where no effects appear at the critical verb itself (in a structure very similar to the one used here). In fact, a recent meta–analysis (Jäger et al., In Press) found that the NUMBER ATTRACTION effect was present immediately in the verb region in only three studies (including the only one in Arabic, Tucker et al., 2015), while it reliably appeared in the spillover region in eight others.

In conclusion, it seems at least *prima facie* possible that verbal attraction for gender exists in MSA, insofar as sentences containing masculine subjects and feminine attractors show the reading time correlates of attraction. However, this is only one-half of the attraction effect profile seen for number in languages such as, *e.g.,* English. The other component to this effect is an asymmetry owing to *markedness* — attraction effects on reaction times or in productions are often found in languages when the erroneous verbal morphology is the marked version more than when it is in the unmarked version (Eberhard, 1997), but nothing in Experiment 1 has shown that this is true for MSA. As discussed in the Introduction, this is an important dimension of similarity upon which to assess the similarity of gender and number attraction. Experiments 2A and 2B, involving the manipulation of subject gender, were designed to address this question.

25

3. Experiments 2A and 2B

Our goal in the second experiment was to assess whether the evidence of attraction effects for Arabic gender we obtained in experiment 1 is replicable and, if so, whether gender attraction effects would pattern along markedness lines the way other agreement features/cues have been observed to in other languages. At least three papers (Badecker & Kuminiak, 2007; Badecker & Lewis, 2007; Malko & Slioussar, To Appear) have all shown that gender attraction can in principle follow language-internal markedness hierarchies with attraction effects sensitive to whether the verb appears in the marked or unmarked version. These findings are at odds, however, with findings from Hebrew, where markedness effects do not appear to obtain in production (Dank & Deutsch, 2010). Moreover, only one study (Malko & Slioussar, To Appear) has assessed this phenomenon in comprehension, reporting one experiment on the three-way gender system of modern Russian.

In MSA — a language with a two-valued system including masculine and feminine nouns — the marked grammatical gender is arguably feminine given that on many nouns, feminine gender is overtly marked with a suffix. Furthermore, conjunctions containing both masculine and feminine nouns invariably resolve to the masculine plural (Ryding, 2005). We therefore expect to find that gender attraction effect profiles would appear more often in reading times when the true subject is masculine and the attractor feminine, rather than the other way around, if markedness effects obtain as in English number, where ungrammatical plural verbs are more acceptable with plural attractors than ungrammatical singular verbs with singular attractors. However, it is equally possible that no markedness asymmetry obtains, as in Hebrew (Dank & Deutsch, 2010), and we would then expect no difference between masculine-feminine conditions and feminine-masculine conditions. This latter result would be challenging given the predictions of both misrepresentation and cue-based retrieval models, assuming that grammatical gender features are subject to the same kinds of markedness distinctions applicable to grammatical number. In view of the importance of establishing the presence or absence of the markedness asymmetry for gender in our data, we present the original study (2A) followed by its direct replication (2B) with a different set of participants.

3.1. Participants

Participants in the study 2A were 128 native speakers of Arabic from the UAEU student community with no history of language disorder and self-assessed proficiency in MSA (128 females; mean age 20.4 years). Participants in the study 2B were 202 native speakers of Arabic from the UAEU student community with no history of language disorder and self-assessed proficiency in MSA (202 females; mean age XX years). All participants provided informed consent and were compensated for their participation in this study and, in experiment 2A, an additional unrelated study. Participants were orally asked whether they had participated in Experiment 1, and, in the case of experiment 2B, whether they had participated in experiment 2A as well, and were excluded from these experiments if they answered affirmatively.

3.2. Materials & Predictions

In order to directly assess the impact of markedness on gender attraction effects in MSA, the 48 item sets from Experiment 1 were altered to allow the main clause subject NP to also appear with the feminine suffix -a/ŝ-. Where pragmatics required, the continuations were altered to allow for sensible interpretations across different genders of subject NPs. All other constraints on the creation of stimuli used in Experiment 1 were followed in this experiment as well, resulting in items which were identical to the items used in Experiment 1 save for these specific changes.

Using each of the 48 sentences as a standard, seven additional variants were constructed by systematically varying the grammatical gender of both the main clause subject and relative clause object NP as well as the main clause verb (the target verb). All feminine NPs were created by attaching the feminine suffix -a/ŝ- to the NP used in the equivalent masculine conditions. All NPs which were the target of experimental manipulations were in the singular number and grammatically animate. We elected to use animate nouns despite the inclusion of notional gender of the referent in order to facilitate comparisons to Experiments 3, 4A, 4B, 5A and 5B and the experiment from Tucker et al. (2015); this additionally adds a new body of evidence to the production data from inanimates furnished by Dank & Deutsch (2010); and Deutsch & Dank (2009, 2011). The items obtained by this choice also match English number marking in the nominal

27

domain extremely closely: the marked alternative (here feminine, in English plural) is expressed with a single orthographic character suffix (-ة in Arabic and -*s* in English). The result is eight conditions per experimental sentence in a $2 \times 2 \times 2$ factorial design crossing SUBJECT GENDER, GRAMMATICALITY, and MATCH.

It should also be noted that complementizers in MSA agree with the NP they modify in both grammatical number and grammatical gender (Ryding, 2005, 322), meaning that conditions with a feminine subject also contain a feminine singular definite complementizer (*ʔallatii*/التي), in contrast to the masculine singular definite complementizer (*ʔallaðii*/الذي) found in masculine subject conditions. Additionally, whenever the subject NP was feminine, the relative clause verb also appeared in the feminine, so that the only possible agreement attraction effects occur on the main clause/target verb. A complete item set for one experimental sentence appears in Table 3.2.

The 48 sets of eight sentences were distributed across eight lists in a Latin Square design after being combined with 144 grammatical fillers of a similar length for a 3:1 filler-to-item ratio. None of the fillers used in Experiment 1 were used for this experiment, and none of the fillers included the relative clause construction used in the experimental stimuli or any construction which drew attention to meaningful alternations in verbal agreement. In the final version of each list, only the experimental sentences contained ungrammaticalities, with 12.5% of the sentences in each list ungrammatical.

In these experiments, one expects a replication of the effects found in Experiment 1. In particular, one expects an effect of GRAMMATICALITY at the target verb (and/or possibly into adjacent spillover) region. While the results from Experiment 1 certainly lead one to expect an ATTRACTION effect in Experiments 2A and 2B, what form that effect should take depends on the expectations one has about the role of markedness in gender attraction. If, following Badecker & Lewis (2007) and Badecker & Kuminiak (2007), markedness applies to gender in identical ways as it applies to number, then one expects to find an interaction of SUBJECT GENDER and the ATTRACTION effect. Moreover, assuming the *grammaticality asymmetry* holds, then this ATTRACTION effect will only be observed in *ungrammatical* sentences. Therefore, in order for us to observe a *markedness asymmetry* effect here, we would need to observe an attraction effect that appears only, or at least more strongly, in sentences with MASC subjects

Condition	NP1 R1	Comp R2	RCV R3	NP2 R4	Adv R5	V R6	Continuation R7–R$_n$
Masc/Match/Gram	المهندس	الذي	استقبل	العالم	بالصدفة	يعمل	على ابتكار جديد.
	The engineer (MASC)	who	met	the scientist (MASC)	by chance	is working (MASC)	on a new invention.
Masc/Match/Ungram	المهندس	الذي	استقبل	العالم	بالصدفة	تعمل	على ابتكار جديد.
	The engineer (MASC)	who	met	the scientist (MASC)	by chance	is working (FEM)	on a new invention.
Masc/NoMatch/Gram	المهندس	الذي	استقبل	العالمة	بالصدفة	يعمل	على ابتكار جديد.
	The engineer (MASC)	who	met	the scientist (FEM)	by chance	is working (MASC)	on a new invention.
Masc/NoMatch/Ungram	المهندس	الذي	استقبل	العالمة	بالصدفة	تعمل	على ابتكار جديد.
	The engineer (MASC)	who	met	the scientist (FEM)	by chance	is working (FEM)	on a new invention.
Fem/NoMatch/Gram	المهندسة	التي	استقبلت	العالم	بالصدفة	تعمل	على ابتكار جديد.
	The engineer (FEM)	who	met	the scientist (MASC)	by chance	is working (FEM)	on a new invention.
Fem/NoMatch/Ungram	المهندسة	التي	استقبلت	العالم	بالصدفة	يعمل	على ابتكار جديد.
	The engineer (FEM)	who	met	the scientist (MASC)	by chance	is working (MASC)	on a new invention.
Fem/Match/Gram	المهندسة	التي	استقبلت	العالمة	بالصدفة	تعمل	على ابتكار جديد.
	The engineer (FEM)	who	met	the scientist (FEM)	by chance	is working (FEM)	on a new invention.
Fem/Match/Ungram	المهندسة	التي	استقبلت	العالم	بالصدفة	يعمل	على ابتكار جديد.
	The engineer (FEM)	who	met	the scientist (FEM)	by chance	is working (MASC)	on a new invention.

Table 4: A complete item set for one stimulus in Experiment 2.

than in sentences with FEM subjects. On the other hand, if markedness affects different agreement cues differentially, one expects to find similar ATTRACTION effects along the SUBJECT GENDER levels.

3.3. Procedure

The procedure for Experiments 2A and 2B were identical to that employed for Experiment 1, save for the difference that participants in 2A were asked to participate in a second, unrelated experiment upon completion of the self-paced reading experiment reported here.

3.4. Analysis

Comprehension question accuracy data in Experiments 2A and 2B were analyzed identically to the analysis for Experiment 1. For the self-paced reading data, all of the analysis was the same as Experiment 1 save for the addition of the additional experimental manipulation of SUBJECT GENDER. Thus, the effects of interest (GRAMMATICALITY and ATTRACTION) will still be computed as described in section 1.3.2, except that they will be calculated along the levels of SUBJECT GENDER (i.e., separately for sentences with MASC subjects, as in experiment 1, and sentences with FEM subjects.

3.5. Results

3.5.1. Comprehension Question Accuracy

In experiment 2A, three participants failed to meet the comprehension question accuracy criterion and were excluded from this and all further analysis. Overall comprehension question accuracy for this experiment was 86.7%, with an accuracy of 87.7% for fillers and 83.7% for experimental items. Since there was little variation in accuracy across experimental conditions, the accuracy data is not shown here (but they are available in the supplementary materials).

In experiment 2B, only one participant failed to meet the comprehension question accuracy criterion and were excluded from this and all further analysis. MATT CAN YOU RUN THE ACCURACY DATA FOR 2B? WE JUST NEED THE DATA FOR THE SUPPLEMENTARY MATERIALS.

	Subject	Verb		Verb+1		Verb+2	
2A: *N* = 125							
Attraction Ungrammatical	Masc	-25	(-48, -2)	26	(-4, 63)	7	(-6, 21)
	Fem	5	(-17, 29)	14	(-4, 32)	8	(-7, 23)
Attraction Grammatical	Masc	-1	(-26, 22)	-5	(-20, 8)	0	(-13, 12)
	Fem	-15	(-33, 2)	*-12*	*(-30, 1)*	-18	(-32, -6)
Grammaticality	Masc	30	(-7, 71)	**119**	**(82, 163)**	**42**	**(20, 64)**
	Fem	-8	(-42, 25)	**25**	**(2, 49)**	**28**	**(9, 49)**
2B: *N* = 201							
Attraction Ungrammatical	Masc	-7	(-26, 12)	**35**	**(17, 57)**	**21**	(6, 38)
	Fem	-8	(-26, 8)	*-16*	*(-40, 1)*	4	(-7, 14)
Attraction Grammatical	Masc	5	(-10, 23)	1	(-13, 14)	-2	(-12, 9)
	Fem	12	(-3, 29)	10	(-2, 24)	-7	(-18, 3)
Grammaticality	Masc	**43**	**(14, 74)**	**73**	**(51, 99)**	**52**	**(34, 72)**
	Fem	9	(-15, 36)	**59**	**(33, 89)**	**33**	**(16, 49)**

Table 5: Results of experiment 2. Mean RT for each effect of interest. 95% Confidence Intervals computed by BCa bootstrap (2000 replications) in parenthesis. Effects in which the CI excludes zero are marked in bold. Effects in which the CI includes zero up to ±1 ms are marked in italic.

3.5.2. Self-Paced Reading

Only the sentences for which the comprehension question was answered correctly were included for subsequent analysis of the self-paced reading data. This resulted in the exclusion of 14.56% of the raw collected data (across all conditions, regions, and participants) in Experiment 2A, and XXX% of the data in Experiment 2B. Mean reading times for each region and condition in Experiments 2A and 2B appear in Figure 2.

In experiment 2A, the grammaticality effect for sentences with MASC subjects appeared strongly only in the *Verb+1* (119 ms) and *Verb+2* (42 ms) regions, although there was a numerical trend in the *Verb* (30 ms) region as well. Reliable GRAMMATICALITY effects were also observed for sentences with FEM subjects in the same *Verb+1* and *Verb+2* regions (25 ms and 28 ms, respectively). As for the attraction effect, we find a numerical trend only in grammatical sentences in the *Verb+1* and, to a lesser extent, *Verb+2* regions, even though in none of these regions the 95% CI excludes zero. The effect size for the attraction effect is larger for ungrammatical sentences with MASC

31

Figure 2: Mean raw reading times from Experiment 2 for all conditions and regions by SUBJECT GENDER. Error bars represent the standard error of the condition mean across participant averages.

subjects (26 ms) than for FEM subjects (14 ms). In addition, for the grammatical sentences, we observe a trend towards a "reverse" attraction effect in the *Verb+1* region (−12 ms).

In experiment 2B, we find reliable grammaticality effects starting in the *Verb* and continuing into the two subsequent regions for sentences with MASC subjects (43 ms, 73 ms, and 52 ms, respectively). The grammaticality effect in sentences with FEM subjects were reliable starting in the *Verb+1* region and continuing into the *Verb+2* region (59 and 33 ms respectively). The attraction effect in experiment 2B was only reliably observed in ungrammatical sentences, and within this group, only in sentences with MASC subjects. It starts in the *Verb+1* region (35 ms) and continues into the subsequent region (21 ms). Contrary to the results of experiment 2A, ungrammatical sentences with FEM subjects had a "reversed attraction effect" in the *Verb+1* region (−16 ms).

3.6. Discussion

When it comes to the question of whether the MARKEDNESS of agreement features modulates the attraction effects, the answer from experiments 2A and 2B is tentatively positive. In experiment 2A, FEM headed sentences showed a numerical attraction effect of 14 ms (even though the 95% CI did not exclude zero) in the *Verb+1* region. Even taking this result at face value, it is weaker than the ones observed in the same regions for MASC headed sentences across all three experiments. However, this numerical attraction effect was not replicated in experiment 2B, where it in fact became an almost reliable "reverse" attraction effect of −16 ms. This is unlike the results observed in MASC headed sentences across the three experiments, which showed remarkable consistency in effect sizes in that post–verbal region. Also worthy of mention is the fact that the size of the grammaticality effect itself seems to be modulated by the MARKEDNESS of the agreement features involved: the results of experiments 2A and 2B for MASC headed sentences roughly replicate the range of effect sizes observed in experiment 1 (54 ms, 127 ms and 59 ms, for *Verb* and its two subsequent regions respectively). In comparison, the grammaticality effects for FEM headed sentences in experiments 2A and 2B were noticeably smaller, and never appeared at the *Verb* region.

As for the *grammaticality asymmetry*, the combined results of experiments 1, 2A

33

and 2B show that the attraction effect, when it occurs, it seems to do so in ungrammatical sentences only, mirroring the findings for what has been observed for NUMBER in languages like English (Wagers et al., 2009).

Regarding the gender attraction effect, the results of Experiments 2A and 2B replicate the evidence observed in Experiment 1, with similar effect sizes. The observed effect sizes for gender attraction in MASC headed sentences were 21 ms and 26 ms and 35 ms in the *Verb+1* region across the three experiments, although only in 2B did the 95% CI exclude zero (it included −1 in Experiment 1 and −4 in Experiment 2A). Experiment 2B also showed reliable gender attraction effects of similar magnitudes in the *Verb+2* region (21 ms), but compared to the results of experiments 1 (−4 ms) and 2A (7 ms), this result looks more like an outlier.

What seems to be consistent across all three experiment so far is the gender attraction effect occurring in the spillover region of the verb. In Experiment 2A neither the grammaticality effect nor the attraction effect appeared until the region immediately following the critical verb. However, in Experiment 2B, as in Experiment 1, the grammaticality effect did appear at the critical verb region (like in Experiment 1), whereas the gender attraction effect only appeared in the region immediately following the critical verb (again like in Experiment 1). In fact, in Experiment 2B, the attraction effect was also visible in the second region following the verb.

In an interim conclusion, the combined results from the first three experiments suggest that gender attraction does seem to occur in Modern Standard Arabic. But is this the same as number attraction in the language? The only study which addresses this question is reported in Tucker et al. (2015). In that paper, the authors show that the attraction effect also does occur for number in MSA, but had nothing to say about the featural asymmetry issue. Experiments 3, 4A and 4B, as well as 5A and 5B try to address these issues, and also clarify some unanticipated results Tucker and colleagues obtained vis-à-vis the number attraction effect in Arabic.

4. Experiment 3

In order to examine the similarities and differences between gender and number attraction in MSA, one must examine whether the markedness asymmetry is present in Arabic number attraction — an effect left untested in the comprehension study by Tucker et al. (2015). However, testing number independent of gender in Arabic requires making a choice about which genders to include while independently manipulating number values. Since gender is orthogonal to number in MSA number agreement paradigms, the simplest option would be to simply counterbalance masculine and feminine verbs across experimental items. However, the one existing study on MSA number attraction in comprehension, Tucker et al. (2015), presents findings concerning the interplay of nominal gender and morphophonological effects on plural formation which make this counterbalancing possibly undesirable. Since any experiment which *a priori* restricted itself to one of two available genders in a language would need to be justified, we first examine the findings from Tucker et al. (2015) in some detail with an experiment designed to replicate and extend those findings.

We begin with a items subgroup issue in the study of Tucker et al. (2015). In that work, the authors leave unresolved a peculiar difference in agreement attraction effect sizes owing to the morphosyntactic nature of the NPs involved. Arabic allows for two different strategies of plural formation: SOUND/SUFFIXING plurals and BROKEN/ABLAUTING plurals. The former take their plurals with a regular, shape-invariant suffix (in that study, -*aat*/ات-), whereas the latter mark plurality by a change in the the vowel and syllabic structure of the singular noun. In the traditional descriptive work on Arabic, this collection of vowels and prosodic structure is known variously as the CV-template, skeleton, or pattern. The vast majority of words in Arabic can be decomposed into a prosodic template and root consisting of 2-4 consonants, as (5) exemplifies for the root $\sqrt{\text{drs}}$:

(5) Words Containing $\sqrt{\text{drs}}$:   (Wehr, 1976, 321)

      a. *darasa*/درس — "he studied/learned"

      b. *darrasa*/درّس — "he taught/caused to learn"

      c. *dars*/درس — "lesson/chapter"

802   d. *diraasa*/دراسة — "study/written"

803   e. *darraas*/درّاس — "student"

804   f. *madrasa*/مدرسة — "school"

805   g. *mudarris*/مدرس — "teacher/instructor"

806 Typologically, Arabic is unique in the high number of broken/ablauting plurals rel-
807 ative to other languages which utilize alteration of the CV-template — indeed, they
808 are arguably more frequent than suffixing/sound plurals insofar as many of the high-
809 frequency nouns in the language take broken plurals. Here, examining just English
810 would lead to a different conclusion, such as that reached by Bock & Eberhard (1993),
811 who demonstrate that attractors with irregular plurals in English do not condition dif-
812 ferent attraction rates in production than those with regular plurals.

813   As Ryding (2005) and Tucker et al. (2015) note, masculine animate nouns tend to
814 take broken plurals and feminine animate nouns tend to take sound plurals. In Tucker
815 et al. (2015), the authors demonstrated that the size of the number agreement attraction
816 effect in MSA is modulated by whether the NPs in the pre-critical region are feminine
817 and the attractor takes a plural with a regular suffix ("sound" plurals in the Arabic litera-
818 ture) or masculine and the attractor takes a plural by alteration of the CV-template/ablaut
819 ("broken" plurals). Specifically, they showed that broken plural attractors cause smaller
820 intrusion effect sizes at ungrammatical verbs than sound plural attractors do. This effect
821 can be seen in the difference between the top and bottom panels of Figure 3 (Tucker
822 et al., 2015, Fig.2). Whereas masculine/broken plural attractor sentences involve only
823 a modest attraction effect, feminine/sound plural attractor sentences involve a much
824 larger attraction effect, with the attraction condition nearly identical to grammatical
825 sentences. Given that all the subjects in this experiment were singular, Tucker and col-
826 leagues reason that this might be due to the salience of morphological plural marking on
827 the attractor insofar as sound plurals contain a morphological or orthographic unit (the
828 suffix) which is clearly associated with plurality, whereas comprehension of a broken
829 plural *qua* plural requires decomposition of a word into its root and CV-template.

830   However, one issue that study does not address is whether there might be differenti-
831 ations to be made inside the class of broken plurals such that the distinction in attraction
832 effect sizes is not due to broken plurals *per se*, but instead is due to more general factors
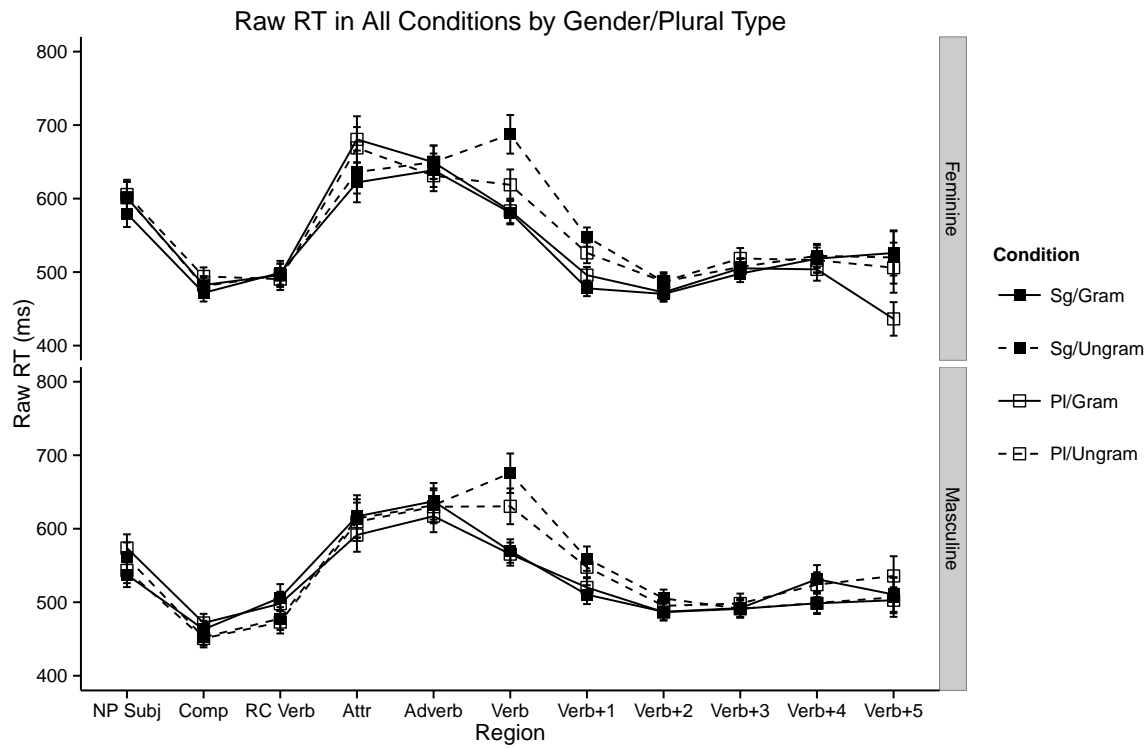
Figure 3: Mean raw reading times from Tucker et al. (2015) segregated by subject/attractor gender. All subjects are singular and Sg/Pl in the condition label refers to attractor number. Error bars represent standard error of the mean computed over subject averages.

known to influence processing. One such property is AMBIGUITY of the morphological marking. As hinted at above, one of the distinctions between sound and broken plurals is that sound plural suffixes *unambiguously* mark plural number, whereas template alterations are commonplace across Arabic and serve to mark many morphological distinctions. Whether morphophonological properties of the attractor plays a role in modulating attraction rates is currently an open question at present: Whereas Vigliocco et al. (1995) and Malko & Slioussar (To Appear) find that they do not, Badecker & Kuminiak (2007) and Dank & Deutsch (2010) find that they do, and the experiments in Hartsuiker et al. (2003) suggest that they do, but at smaller rates than those of the head noun and only for certain kinds of ambiguity (*i.e.,* grammatical case). We thus wish to rule out whether ambiguity of morphophonology on the attraction might be confounding the decreased magnitude in the broken plural attraction effect reported in Tucker et al. (2015).

Furthermore, Tucker et al. (2015) leave open whether the difference between sound and broken plural attractors is a categorical or gradient one: both the idea that broken plurals do not engender any attraction as well as the idea that they engender considerably smaller rates of attraction are compatible with their results. Here, we design an experiment which aims to clear up both these outstanding issues from Tucker et al. (2015) while simultaneously re-examining the timing of number agreement attraction. In order to do this, we exploit a fortunate property of Arabic broken plurals wherein some CV-templates underwriting broken plurals are used exclusively to mark plural number on nouns and some are not. For example, the CV-template associated with the plural noun الصوص/*lus$^\Omega$uus$^\Omega$*, "thieves" — $C_1uC_2uuC_3$ — is also found in singular nouns, such as the deverbal nominalization دخول/*duxuul*, "entering (n.)" and is therefore morphologically ambiguous with respect to number marking. This can be contrasted with a different template — such as $C_1uC_2aC_3$aaʔ as in the noun علماء/*ʕulamaaʔ*, "scientists" — which is found only with plural nouns and can be considered morphologically unambiguous with respect to number. We therefore designed an experiment which tested only masculine attractors taking broken plurals and varied whether the template of those broken plurals is ambiguous or not. The result is a higher-powered replication of the masculine half of the study in Tucker et al. (2015) (with twice as many items) and a

38

further investigation of the role of ambiguity in MSA number agreement attraction.

4.1. Participants

Participants were 110 native speakers of Arabic from the UAEU community with no history of language or other cognitive disorders and self-assessed proficiency in MSA (110 females; mean age 21.1 years). All participants provided informed consent and were compensated for approximately 45 minutes of time.

4.2. Materials & Predictions

In order to jointly assess the reliability of a lack of an attraction effect for number in masculine broken plurals from Tucker et al. (2015) and the contribution (if any) of template ambiguity, 48 sentences were constructed of the form *NP1 — Complementizer — [Verb — NP2 — Adverb] —* **Target Verb** *— Continuation*, exactly as in the previous two experiments and in Tucker et al. (2015). This is twice the number of items with masculine pre-critical NPs compared to the subgroup in Tucker et al. (2015), where only 24 such items appeared. In this experiment, however, both *NP1* and *NP2* were specified as masculine grammatically and took their plural form in a broken pattern and not with a suffix. Additionally, broken plurals were classified into two categories — AMBIGUOUS and UNAMBIGUOUS plurals. Plural ambiguity was assigned based on the prosodic/CV-template pattern that the plural contained. Templates were considered ambiguous if the second author and a collection of other native speaker consultants could easily think of singular nouns which appeared in that same CV-pattern and unambiguous otherwise. A complete list of the templates and classifications used in the construction of the stimuli for this experiment appear in Table .1. In order to keep the duration of the experiment manageable, the ambiguity of NP2 was manipulated across the 48 sentences. The result was 24 items with NP2s that took ambiguous plurals and 24 items with NP2s that took unambiguous plurals. All other constraints on the creation of stimuli in Experiments 1 and 2 were followed, where applicable to number instead of grammatical gender. A complete list of sentences for this experiment appears in 8.4.

The 48 sentences were then individually converted into four conditions by systematically varying the grammatical number (singular, plural) of both NP2 and the target

39

AMBIGUOUS	UNAMBIGUOUS
$C_1aC_2aC_3a$	$C_1uC_2aC_3aa?$
$C_1aC_2iiC_3$	$?aC_1aaC_2iC_3a$
$C_1iC_2aaC_3$	$C_1awaaC_2iC_3$
$C_1iC_2C_3aan$	$?aC_1C_2aaC_3$
$C_1uC_2C_3aan$	$?aC_1C_2iC_3aa?$
$C_1aC_2C_3aa$	
$C_1uC_2C_2aaC_3$	
$C_1uC_2uuC_3$	

Table 6: Templates and ambiguity assignments for broken plural templates in Experiment 3.

verb. The resulting collection of four conditions for each of the 48 sentences comprised a $2 \times 2 \times 2$ factorial design crossing MATCH (yes, no) and GRAMMATICALITY (grammatical, ungrammatical) and a between-items manipulation of AMBIGUITY. The 2×2 subset collapsing over ambiguity is therefore an indentical design to Experiment 1. However, in this study, all the NOMATCH conditions contained a singular NP1 and a plural NP2, and ungrammatical verbs were always plural. A complete item set for one of the experimental sentences appears in Table 4.2.

These 48 sets of four sentences were distributed across four lists in a Latin Square design and combined with 144 grammatical fillers for a 3:1 filler:item ratio where 12.5% of the items were ungrammatical. None of the fillers used in Experiments 1 or 2 were used for this experiment, and fillers varied in length from four to fifteen words long. None of the fillers contained the relative clause construction at the core of the experimental sentences.

Given that Tucker et al. (2015) report a diminished attraction effect in masculine items with broken plurals, one would expect to find only a GRAMMATICALITY effect in this experiment with potentially no ATTRACTION effect, though one could feasibly expect to see a numerical trend toward attraction which is not very large in magnitude. If ambiguity of number marking is relevant for the effect reported by Tucker and colleagues, then we additionally would expect an effect of AMBIGUITY interacting with the size of the ATTRACTION effect, meaning that attraction is modulated by the level of AMBIGUITY. Were that to obtain, whether or not GRAMMATICALITY is also part

40

Condition	NP1 R1	Comp R2	RCV R3	NP2 R4	Adv R5	V R6	Continuation R7–R_n
MATCH/GRAM	الطفل The child (SG)	الذي who	رأى watched	الساحر the magician (SG)	بانبهار with amazement	صفق applauded (SG)	بشدة خلال العرض. hysterically during the show.
MATCH/UNGRAM	الطفل The child (SG)	الذي who	رأى watched	الساحر the magician (SG)	بانبهار with amazement	صفقوا applauded (PL)	بشدة خلال العرض. hysterically during the show.
NOMATCH/GRAM	الطفل The child (SG)	الذي who	رأى watched	السحرة the magicians (PL)	بانبهار with amazement	صفق applauded (SG)	بشدة خلال العرض. hysterically during the show.
NOMATCH/UNGRAM	الطفل The child (SG)	الذي who	رأى watched	السحرة the magicians (PL)	بانبهار with amazement	صفقوا applauded (PL)	بشدة خلال العرض. hysterically during the show.

Table 7: A complete item set for one stimulus in Experiment 3.

of the interaction would be dependent upon the choice among misrepresentation and cue-based retrieval models, exactly as in Experiments 1, 2A and 2B.

4.3. Procedure

The procedure for Experiment 3 was exactly the same as the procedure for Experiments 2A and 2B.

4.4. Analysis

Comprehension question accuracy data for Experiment 3 was analyzed identically to the analysis of experiments 1, 2A and 2B. For the self-paced reading data, all of the analysis was the same as Experiment 1 save for the addition of the additional experimental manipulation of PLURAL TEMPLATE AMBIGUITY of the attractor NP. Thus, the effects of interest (GRAMMATICALITY and ATTRACTION) will still be computed as described in section 1.3.2, except that they will be calculated along the levels of PLURAL TEMPLATE AMBIGUITY (i.e., separately for sentences containing attractors carrying AMBIGUOUS vs UNAMBIGUOUS plural templates).

4.5. Results

4.5.1. Comprehension Question Accuracy

None of the participants in this experiment met the criteria for exclusion based on global comprehension question accuracy, and so all were included in the subsequent analyses. Overall comprehension question accuracy for this experiment was 88.8% with accuracy rates of 86.8% for fillers and 89.5% for experimental items. Accuracy rates to matching attractor sentences were 88.6% (CI = 86.8-90.3%) to grammatical sentences and 87.2% (CI = 85.2%-88.9%) to ungrammatical sentences. Accuracy for non-matching attractors was 87.2% (CI = 85.3-89.0%) to grammatical sentences and 84.1% (CI = 82.0-86.0%) to ungrammatical sentences.

4.5.2. Self-Paced Reading

Only sentences for which the comprehension question was answered accurately were included in the subsequent reading time analysis. This resulted in the exclusion of approximately 13.01% of the raw collected data (across all conditions, participants,

42

$N = 110$	**Verb**		**Verb+1**		**Verb+2**	
Ambiguous						
Attraction Ungrammatical	17	(-15, 50)	15	(-9, 38)	12	(-6, 32)
Attraction Grammatical	5	(-16, 26)	-15	(-35, 4)	-7	(-23, 9)
Grammaticality	**110**	**(67, 159)**	**80**	**(46, 116)**	**39**	**(13, 67)**
Unambiguous						
Attraction Ungrammatical	-13	(-51, 22)	-3	(-29, 31)	-1	(-19, 17)
Attraction Grammatical	4	(-24, 28)	4	(-16, 22)	14	(-2, 32)
Grammaticality	**154**	**(99, 216)**	**88**	**(50, 124)**	**39**	**(13, 67)**

Table 8: Results of experiment 3. Mean RT for each effect of interest. 95% Confidence Intervals computed by BCa bootstrap (2000 replications) in parenthesis. Effects in which the CI excludes zero are marked in bold. Effects in which the CI includes zero up to ±1 ms are marked in italic.

and items). Mean reading times across participant averages for each region are shown in Figure 4. Table 8 shows the results for critical regions of interest.

The only reliable results observed here were the GRAMMATICALITY effects, which were found in the *Verb* and its two subsequent regions for sentences containing attractors carrying both AMBIGUOUS and UNAMBIGUOUS plural templates.

4.6. Discussion

The results of Experiment 3 serve as a replication of one-half of the experiment reported in Tucker et al. (2015), insofar as it contained items with masculine NPs and attractors that take broken plurals. In this experiment, we also fail to find any reliable evidence of attraction effects in reading times. The only effects that were numerically compatible with NUMBER attraction were the ones from sentences that had AMBIGUOUS attractors in the three critical regions, although in none of them did the 95% CIs exclude 0. Because GRAMMATICALITY differences are being noticed by participants regardless of the attractor type (leading to slowdowns in reading time), it is clear that participants are attending to the agreement morphology. However, the absence of reliable NUMBER attraction effects means that agreement attraction is either not occurring or incredibly small. More conservatively, one might simply maintain that there is an important distinction to be made between the feminine sound plurals examined in Tucker
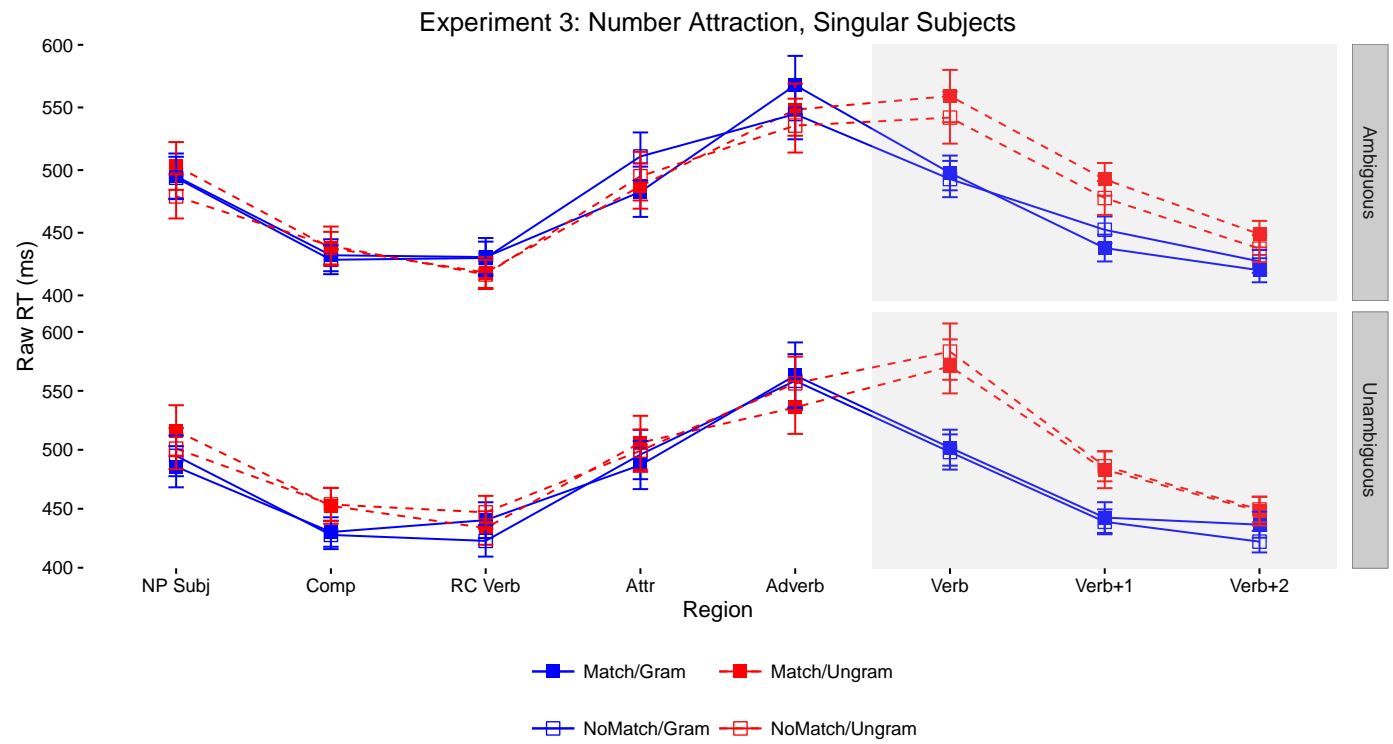
43

Figure 4: Mean raw reading times from Experiment 3 for all conditions and regions by attractor AMBIGUITY. Error bars represent the standard error of the condition mean across participant averages.

et al. (2015) and the masculine broken plurals re-examined here, one which must be taken into account when considering appropriate items for comparing attraction across gender and number features.

However, Experiment 3 can also provide an additional piece of information concerning what this difference might not be attributed to. In this experiment we find no evidence that the morphological ambiguity of the CV-template of the attractor drives this difference between masculines and feminines in Tucker et al. (2015). In fact, to the extent that there is evidence of NUMBER attraction effects in Experiment 3, it comes from the sentences containing AMBIGUOUS attractors, and not from the UNAMBIGU-OUS ones, which is the exact opposite pattern of results one would have expected if ambiguity of the plural template was the causal factor leading to small or inexistent attraction effects for broken plurals in Tucker et al. (2015). As outlined above, it was reasonable to wonder whether this could be the case, given the considerations that the ease with which participants recover morphological number information could underly broken versus sound plural differences. However, we find no evidence that this is actually occurring and therefore find converging evidence with that reported by, for instance, Vigliocco et al. (1995) and Malko & Slioussar (To Appear), that morphological ambiguity of the attractor not relating to case morphology plays little or no role in modulating attraction rates.

In summary, in order to directly compare the attraction effects of gender and number features in MSA, it was first necessary to ensure that the plural-type asymmetry from Tucker et al. (2015) was replicable. Here we find converging evidence that broken plural attractors either fail to elicit NUMBER attraction effects, or do so at a much smaller rate than do sound plurals. With this in mind, we now turn to a domain in which agreement attraction effects for number *are* expected in MSA: sentences with sound plural attractors (the other subgroup from Tucker et al., 2015) in order to directly compare the results of Experiments 1, 2A and 2B with similar effects for number.
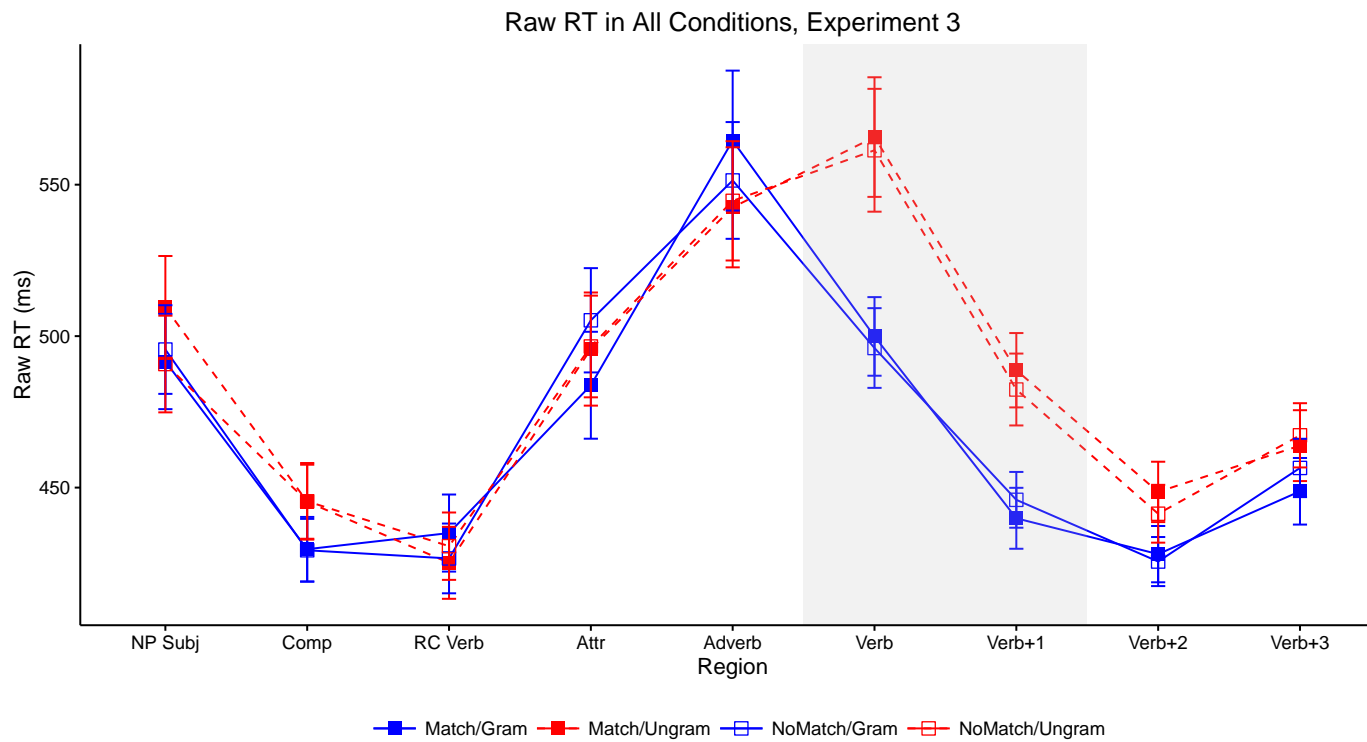
Figure 5: Mean raw reading times from Experiment 3 for all conditions and regions collapsed across attractor ambiguities. Error bars represent the standard error of the condition mean across participant averages.

5. Experiments 4A and 4B

While Experiment 3 seems to confirm the claim that MSA number agreement attraction is not present when the pre-critical region contains masculine NPs and/or broken plural attractors, there remain several open questions about the nature of number agreement attraction in MSA given the results from Tucker et al. (2015) and the first four experiments reported here. First, while it has been claimed above that gender attraction effects mirror agreement attraction effects in directionality and potentially markedness as well, this latter property has not been evaluated for Arabic number agreement in any fashion. The predictions are clear: given that English number attraction only gives rise to attraction RT profiles when the unmarked singular (*i.e., is*) is replaced by the marked plural (*i.e., are*), one could expect that attraction proceeds in the same way in MSA. Conversely, one could expect, in line with the predictions of both representation and cue-based models and the results of Experiments 2A and 2B, that number and gender behave identically in not displaying markedness asymmetries in MSA. Furthermore, given that English and Arabic belong to distinct and somewhat disparate language families where different notions of markedness are could be at play, it is important to examine whether plural-to-singular attractions give rise to attraction RT profiles in MSA, as well.

Finally, the exact experimental design used by Tucker et al. (2015) was, as the authors themselves admit, not designed to observe the true strength of agreement attraction effects after acknowledging a difference between masculine broken/ablauting and feminine sound/suffixing animate plural attractors. This differential effect was an unexpected subgroup effect which should be examined more closely. In Experiment 3 we provided converging evidence that ablaut plurals in MSA do not show agreement attraction, which means that attraction in that language should be solely a function of suffixing plurals. To these ends, we designed an experiment exactly like Experiments 2A and 2B, but which utilized only the feminine/sound plural attractor subgroup of items from Tucker et al. (2015). The result is an experiment designed to replicate the presence of attraction for number cues at the verb while simultaneously testing for the presence or absence of a markedness asymmetry in MSA feminine number agreement

47

attraction effects. Given the importance of these results, in addition to the original study
(4A) we conducted a direct replication (4B), as it was done in Experiment 2.

5.1. Participants

Participants in experiment 4A were 112 native speakers of Arabic from the UAEU
community (112 females; mean age 20.6 years). Participants in experiment 4B were
218 native speakers of Arabic from the UAEU community (218 females; mean age
XX.X years). Participants reported proficiency in MSA and no history of language or
other cognitive disorders. All participants provided informed consent and were compensated for their time.

5.2. Materials & Predictions

In order to assess the effects of markedness in MSA number attraction as well as
replicate the findings of Tucker et al. (2015) with respect to feminine attractors, 54
sentences were constructed of the form *NP1 — Complementizer — [Verb — NP2 —
Adverb] — **Target Verb** — Continuation*, exactly as in the previous three experiments.
However, in this experiment both *NP1* and *NP2* were constrained to be grammatically
feminine nouns bearing the feminine suffix -*a*/ة-. Given that these nouns had singulars
ending in -*a*/ة-, their plurals were all suffixal, ending in -*aat*/ات-. This choice was made
for two reasons: (1) it allowed for higher-powered replication of the subset of results
from Tucker et al. (2015) that involved feminine sound plural attractors (with 54 items
compared to 24 in Tucker et al., 2015) and (2) if Tucker and colleagues' hypothesis that
suffixing attractors provide for greater attraction rates, then these feminines items provide the greatest opportunity to observe attraction with erroneous unmarked feminine
singular verbs. All other constraints applied to items in Experiment 3 and in Tucker
et al. (2015) were followed, where possible.

The 54 sentences were then individually converted into eight conditions by systematically varying the grammatical number (singular or plural) of the word in the *NP1*,
NP2, and the *Verb*. The result was a collection of eight variants organized in a $2 \times 2
\times 2$ factorial design crossing SUBJECT NUMBER (singular, plural), MATCH (yes, no),
and GRAMMATICALITY (grammatical, ungrammatical). A complete item set for one

48

of the experimental sentences appears in Table 5.5.1 and a complete list of experimental sentences appears in 8.4.

These 54 sets of eight sentences were distributed across eight lists in a Latin Square design and combined with 144 fillers for a filler-to-item ratio of 2.67:1. The fillers were randomly selected from the collection of fillers used in Experiments 1–3, none of which contained the construction used in the experimental items (subject relative clauses attached to a subject) and varied in length from four to sixteen words long. All the fillers were grammatical with a total of 13.6% of the sentences ungrammatical in any given list.

If the results from the subset of items in Tucker et al. (2015) bearing feminine sound plural attractors replicate, then one expects to find a GRAMMATICALITY effect beginning at the main clause/target verb along with NUMBER ATTRACTION effects. These effects may spill over into the post–verbal regions but, given the effects in the previous study by Tucker and colleagues, one expects to find that the number attraction effect begins and is largest at the critical verb region itself.

5.3. Procedure

The procedure followed for Experiment 4 was exactly the same as the procedure for Experiments 1, 2A, 2B and 3.

5.4. Analysis

Comprehension question accuracy for Experiments 4A and 4B were analyzed identically to the comprehension question accuracy analysis in Experiments 1–3. For the self-paced reading data, raw reading times were analyzed exactly as in Experiments 2A and 2B, save for the substiution of SUBJECT GENDER for SUBJECT NUMBER.

5.5. Results

5.5.1. Comprehension Question Accuracy

In Experiment 4A, one subject met the criteria for exclusion due to low accuracy based upon global comprehension question scores; she was therefore excluded from the subsequent analyses. Overall comprehension question accuracy for this experiment was 89.6% with accuracy rates of 89.4% for fillers and 89.7% for experimental items.

49

Condition	NP1 R1	Comp R2	RCV R3	NP2 R4	Adv R5	V R6	Continuation R7–R$_n$
Sg/Match/Gram	المدربة The coach (SG)	التي who	اهتمت was interested	باللاعبة in.the player (SG)	جداً very	اشتغلت worked (SG)	في الأكاديمية الوطنية للمبارزة. at the National Fencing Academy.
Sg/Match/Ungram	المدربة The coach (SG)	التي who	اهتمت was interested	باللاعبة in.the player (SG)	جداً very	اشتغلن worked (PL)	في الأكاديمية الوطنية للمبارزة. at the National Fencing Academy.
Sg/NoMatch/Gram	المدربة The coach (SG)	التي who	اهتمت was interested	باللاعبات in.the players (PL)	جداً very	اشتغلت worked (SG)	في الأكاديمية الوطنية للمبارزة. at the National Fencing Academy.
Sg/NoMatch/Ungram	المدربة The coach (SG)	التي who	اهتمت was interested	باللاعبات in.the players (PL)	جداً very	اشتغلن worked (PL)	في الأكاديمية الوطنية للمبارزة. at the National Fencing Academy.
Pl/NoMatch/Gram	المدربات The coaches (PL)	اللواتي who	اهتمن were interested	باللاعبة in.the player (SG)	جداً very	اشتغلن worked (PL)	في الأكاديمية الوطنية للمبارزة. at the National Fencing Academy.
Pl/NoMatch/Ungram	المدربات The coaches (PL)	اللواتي who	اهتمن were interested	باللاعبة in.the player (SG)	جداً very	اشتغلت worked (SG)	في الأكاديمية الوطنية للمبارزة. at the National Fencing Academy.
Pl/Match/Gram	المدربات The coaches (PL)	اللواتي who	اهتمن were interested	باللاعبات in.the players (PL)	جداً very	اشتغلن worked (PL)	في الأكاديمية الوطنية للمبارزة. at the National Fencing Academy.
Pl/Match/Ungram	المدربات The coaches (PL)	اللواتي who	اهتمن were interested	باللاعبات in.the players (PL)	جداً very	اشتغلت worked (SG)	في الأكاديمية الوطنية للمبارزة. at the National Fencing Academy.

Table 9: A complete item set for one stimulus in Experiment 4. Note that NP1, NP2, RCV and V are all morphologically feminine.

	Subject	Verb		Verb+1		Verb+2	
4A: *N* = 111							
Attraction Ungrammatical	Singular	**32**	**(10, 56)**	**19**	**(2, 40)**	-4	(-17, 9)
	Plural	*-19*	*(-39, 1)*	0	(-13, 13)	-3	(-15, 8)
Attraction Grammatical	Singular	-12	(-28, 7)	-3	(-14, 10)	2	(-9, 14)
	Plural	*21*	*(-1, 46)*	-1	(-21, 15)	9	(-2, 21)
Grammaticality	Singular	**55**	**(21, 100)**	**72**	**(50, 96)**	**33**	**(15, 55)**
	Plural	-18	(-50, 13)	1	(-22, 21)	2	(-16, 18)
4B: *N* = 21?							
Attraction Ungrammatical	Singular	-3	(-16, 10)	6	(-6, 19)	-1	(-10, 7)
	Plural	1	(-10, 12)	0	(-8, 10)	7	*(-1, 16)*
Attraction Grammatical	Singular	6	(-6, 18)	3	(-5, 12)	-1	(-9, 6)
	Plural	-5	(-18, 7)	-6	(-15, 4)	-3	(-11, 5)
Grammaticality	Singular	**37**	**(17, 58)**	**54**	**(39, 72)**	**31**	**(18, 44)**
	Plural	-6	(-26, 14)	-5	(-19, 9)	*11*	*(0, 24)*

Table 10: Results of experiment 4. Mean RT for each effect of interest. 95% Confidence Intervals computed by BCa bootstrap (2000 replications) in parenthesis. Effects in which the CI excludes zero are marked in bold. Effects in which the CI includes zero up to ±1 ms are marked in italic.

In Experiment 4B, one subject met the criteria for exclusion due to low accuracy based upon global comprehension question scores; she was therefore excluded from the subsequent analyses. NEED ACCURACY RATES HERE TOO!

5.5.2. Self-Paced Reading

Only sentences for which the comprehension questions were answered correctly were included in the reading time analysis. This resulted in the exclusion of approximately 10.69% of the raw data acquired from the experimental sentences (across all conditions, participants, and items) in experiment 4A, and XX% in experiment 4B. Mean reading times across participant averages for all conditions by subject number appear in Figure 6. Table 10 shows the results for critical regions of interest.

In experiments 4A and 4B alike, reliable GRAMMATICALITY effects were only observed in SINGULAR headed sentences, and they were found in the three critical regions.

Regarding the presence of NUMBER ATTRACTION effects, in experiment 4A we find reliable effects in the *Verb* (32 ms) and *Verb+1* (19 ms) regions, but only for *un-*
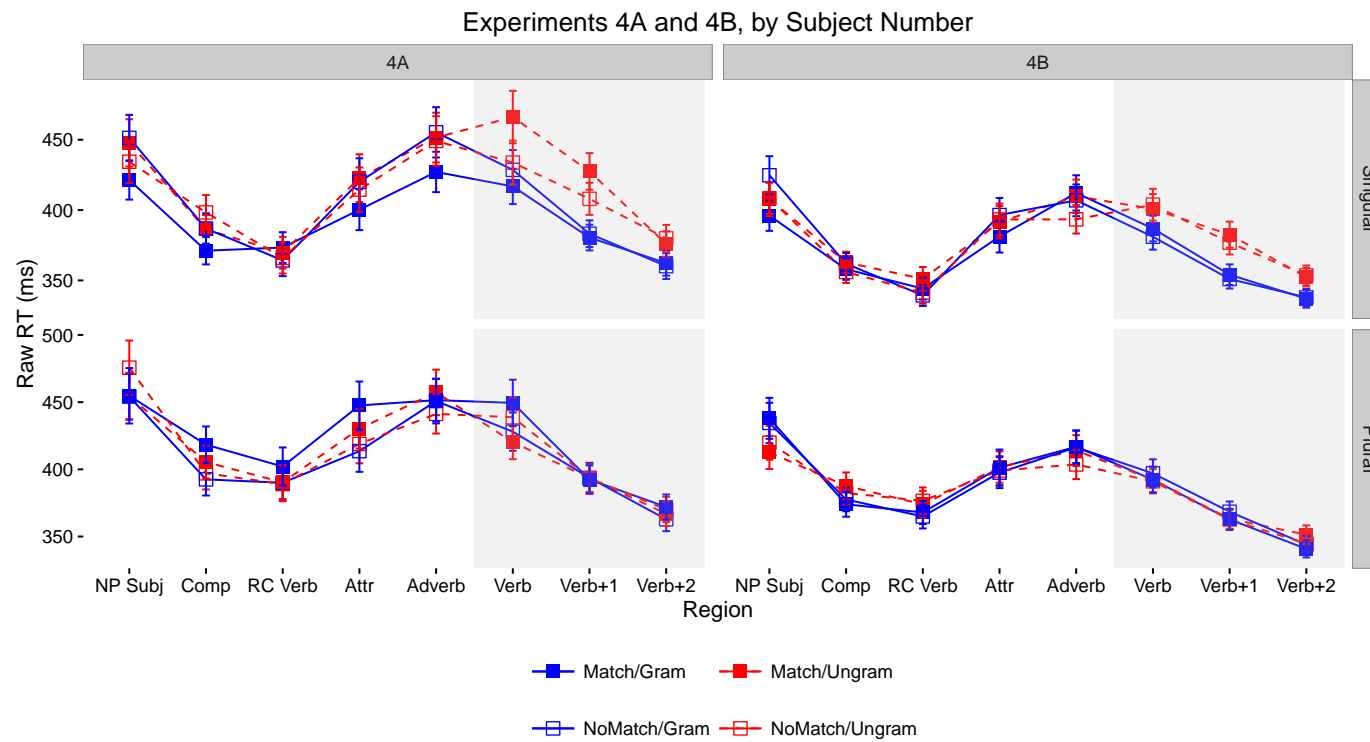
Figure 6: Mean raw reading times from Experiment 4 for all conditions and regions by subject number. Error bars represent the standard error of the condition mean across participant averages.

grammatical SINGULAR headed sentences; no reliable effects were observed when the subject was PLURAL or the sentence was *grammatical*.

In experiment 4B, there was no reliable evidence of NUMBER ATTRACTION effects, in either *ungrammatical* or *grammatical* sentences.

5.6. Discussion

The results of Experiment 4A largely replicate the results found by Tucker et al. (2015) for the feminine suffixing plural subgroup of items in their experiment. Specifically, we can observe here that participants are able to recognize grammaticality manipulations early — upon being presented with the ungrammatical verb. Also like in all experiments in this study, the grammaticality effect is also found in post–verbal regions.

As to the question of whether or not the subgroup effect owing to suffixing feminine plurals in Tucker et al. (2015) can be relied upon, the answer from experiment 4A seems to be an affirmative one. The NUMBER ATTRACTION effect which appears at the verb for singular subject sentences is a direct analogue of the attraction effect in English and a replication of the previous results reported by Tucker and colleagues. Moreover, this effect is largest at the verb region, though it continues into the immediately postverbal spillover region. Moreover, the results for experiment 4A also show other properties normally associated with NUMBER ATTRACTION in other languages: the GRAMMATICALITY ASYMMETRY (effect only found in ungrammatical sentences), and the MARKEDNESS ASYMMETRY (attraction occurs from singular to plurals, but not the reverse). Finally, the NUMBER ATTRACTION effect size observed in 4A (32 ms in *Verb* and 19 ms in *Verb+1* matches the range of gender attraction effects observed in experiments 1, 2A and 2B (21 ms, 26 ms and 35 ms in the *Verb+1* region), as well as the point estimate found in a recent meta–analysis on NUMBER ATTRACTION effects (22 ms, Jäger et al., In Press).

For all these reasons it is extremely perplexing that the results of experiment 4B completely fail to replicate the attraction effect observed in Tucker et al. (2015) and in experiment 4A, even though a grammaticality effect is observed at the verb and all post–verbal critical regions. Given that experiment 4B had a sample size of almost twice the size as that of experiment 4A and of Tucker et al. (2015), this creates a conundrum: on

the one hand, we have two relatively high–powered experiments replicating each other and the results observed in other languages, but on the other we have a third experiment that is even better powered than the previous two, but which fails to replicate them. It is interesting to note that this was not the case for gender attraction in experiments 1, 2A and 2B, which obtained largely similar results amongst themselves. This discordance in the empirical findings about number attraction will be better adjudicated in the follow up experiments (5A and 5B) and the subsequent meta–analysis.

In summary, the results from Experiment 4A confirm the notion that number agreement attraction in MSA is present at erroneous verbs in the feminine morphological paradigm, given the presence of suffixing distractors. Moreover, this effect is timed similarly to other number attraction results insofar as it peaks at the critical verb and decays quickly thereafter. This generalization, however, is substantially challenged by the results of experiment 4B, where no number attraction was observed, even though a grammaticality effect in the verb and its spillover regions was observed in sentences with singular subjects. However, the results of experiments 4A and 4B alike fail to provide evidence that attraction in grammatical sentences occurs in MSA.

Taken together, the combined results of experiments 1–4 document apparent differences between number and gender attraction, namely the differential effects in timing (at *Verb* and perhaps *Verb+1* for NUMBER, but at *Verb+1* and perhaps *Verb+2* regions for GENDER).

However, given how noisy self–paced reading results can be, it is important to see if these differences occur within the same experiment, for the same set of participants. This is the primary goal of experiments 5A and 5B. In addition, given the conflicting results about number attraction itself (observed in Tucker et al. (2015) and experiment 4A, but not in 4B), the results of experiments 5A and 5B may also help clarify the status of number attraction in MSA.

6. Experiment 5

The results of experiments 1, 2A and 2B thus far paint a consistent picture about the nature of GENDER attraction effects: They (i) exhibit GRAMMATICALITY ASYMME-

TRY (i.e., only occur for *ungrammatical sentences*), (ii) they also exhibit the MARKED-NESS ASYMMETRY (i.e., reliably occur from masculine to feminine, but not the other way round), and (iii) systematically occur *after* the *Verb* region, even though a GRAM-MATICALITY effect is often detectable at the *Verb* region itself.

The picture that emerges from Tucker et al. (2015) and experiments 3, 4A and 4B about NUMBER attraction, on the other hand, is a little more mixed: *when it occurs*, it (i) exhibits GRAMMATICALITY ASYMMETRY (i.e., only occur for *ungrammatical sentences*), (ii) also exhibits the MARKEDNESS ASYMMETRY (i.e., reliably occurs from singular to plurals, but not the other way round), and (iii) systematically occurs *at* the *Verb* region (with potential spillover to the post–verbal region), as well as it (iv) tends to occur only when the attractor is a SUFFIXING/SOUND PLURAL.

These differences in timing, and perhaps reliability, observed between AGREE-MENT ATTRACTION for NUMBER and GENDER have so far only been observed across different experiments, with different samples of participants. Therefore, it is important to see if the differences would hold in a fully within–participants design. That is the goal of experiment 5A. Given the importance of these findings, we again conduct a direct replication study (5B), with a different sample of participants.

6.1. Participants

Participants in experiment 5A were 200 native speakers of Arabic from the UAEU community (200 females; mean age 20.6 years). Participants in experiment 5B were another 100 native speakers of Arabic from the UAEU community (100 females; mean age XX.X years). Participants reported being proficient in MSA and having no history of language or other cognitive disorders. All participants provided informed consent and were compensated for approximately one hour of their time.

6.2. Materials & Predictions

MATT I NEED YOU TO DESCRIBE THE MATERIALS HERE

6.3. Analysis

Comprehension question accuracy for Experiments 5A and 5B were analyzed identically to the comprehension question accuracy analysis in Experiments 1–4. For the

Condition	NP1 R1	Comp R2	RCV R3	NP2 R4	Adv R5	V R6	Continuation R7–R$_n$
Sg/Match/Gram	المدربة The coach (SG)	التي who	اهتمت was interested	باللاعبة in.the player (SG)	جداً very	اشتغلت worked (SG)	في الأكاديمية الوطنية للمبارزة. at the National Fencing Academy.
Sg/Match/Ungram	المدربة The coach (SG)	التي who	اهتمت was interested	باللاعبة in.the player (SG)	جداً very	اشتغلن worked (PL)	في الأكاديمية الوطنية للمبارزة. at the National Fencing Academy.
Sg/NoMatch/Gram	المدربة The coach (SG)	التي who	اهتمت was interested	باللاعبات in.the players (PL)	جداً very	اشتغلت worked (SG)	في الأكاديمية الوطنية للمبارزة. at the National Fencing Academy.
Sg/NoMatch/Ungram	المدربة The coach (SG)	التي who	اهتمت was interested	باللاعبات in.the players (PL)	جداً very	اشتغلن worked (PL)	في الأكاديمية الوطنية للمبارزة. at the National Fencing Academy.
Pl/NoMatch/Gram	المدربات The coaches (PL)	اللواتي who	اهتمن were interested	باللاعبة in.the player (SG)	جداً very	اشتغلن worked (PL)	في الأكاديمية الوطنية للمبارزة. at the National Fencing Academy.
Pl/NoMatch/Ungram	المدربات The coaches (PL)	اللواتي who	اهتمن were interested	باللاعبة in.the player (SG)	جداً very	اشتغلت worked (SG)	في الأكاديمية الوطنية للمبارزة. at the National Fencing Academy.
Pl/Match/Gram	المدربات The coaches (PL)	اللواتي who	اهتمن were interested	باللاعبات in.the players (PL)	جداً very	اشتغلن worked (PL)	في الأكاديمية الوطنية للمبارزة. at the National Fencing Academy.
Pl/Match/Ungram	المدربات The coaches (PL)	اللواتي who	اهتمن were interested	باللاعبات in.the players (PL)	جداً very	اشتغلت worked (SG)	في الأكاديمية الوطنية للمبارزة. at the National Fencing Academy.

Table 11: A complete item set for one stimulus in Experiment 5. Note that NP1, NP2, RCV and V are all morphologically feminine in the number manipulation and singular in the gender manipulation.

	Subject	Verb		Verb+1		Verb+2	
5A: *N* = 200							
Attraction Ungrammatical	Gender	2	(-11, 15)	*10*	*(-1, 23)*	**11**	**(3, 21)**
	Number	14	(-2, 29)	10	(-3, 23)	0	(-11, 10)
Attraction Grammatical	Gender	-2	(-11, 10)	-2	(-10, 6)	1	(-7, 9)
	Number	8	(-5, 21)	-4	(-14, 5)	6	(-4, 15)
Grammaticality	Gender	9	(-10, 27)	**57**	**(42, 76)**	**43**	**(29, 58)**
	Number	**35**	**(13, 60)**	**38**	**(21, 56)**	10	(-5, 24)
5B: *N* = 100							
Attraction Ungrammatical	Gender	-4	(-26, 21)	15	(-2, 33)	**16**	**(3, 28)**
	Number	-25	(-58, 3)	-5	(-22, 14)	-1	(-14, 11)
Attraction Grammatical	Gender	7	(-18, 30)	-14	(-34, 4)	6	(-5, 19)
	Number	-29	(-53, -6)	-9	(-25, 5)	2	(-11, 16)
Grammaticality	Gender	9	(-25, 41)	**42**	**(19, 67)**	**26**	**(5, 47)**
	Number	**58**	**(24, 95)**	**35**	**(15, 55)**	1	(-22, 22)

Table 12: Results of experiment 5. Mean RT for each effect of interest. 95% Confidence Intervals computed by BCa bootstrap (2000 replications) in parenthesis. Effects in which the CI excludes zero are marked in bold. Effects in which the CI includes zero up to ±1 ms are marked in italic.

self-paced reading data, raw reading times were analyzed exactly as in Experiments 4A and 4B, save for the substitution of SUBJECT NUMBER for SUBJECT PHI–FEATURE.

6.4. Results

6.4.1. Comprehension Question Accuracy

MATT ANOTHER TASK FOR YOU

6.4.2. Self-Paced Reading

Only sentences for which the comprehension questions were answered correctly were included in the reading time analysis. This resulted in the exclusion of approximately XX% of the raw data acquired from the experimental sentences (across all conditions, participants, and items) in experiment 5A, and XX% in experiment 5B. Mean reading times across participant averages for all conditions by subject number appear in Figure 7. Table 12 shows the results for critical regions of interest.

In experiments 5A and 5B alike, a reliable GRAMMATICALITY effect emerged in the *Verb* region for the NUMBER manipulation which continued into the *Verb+1* region,
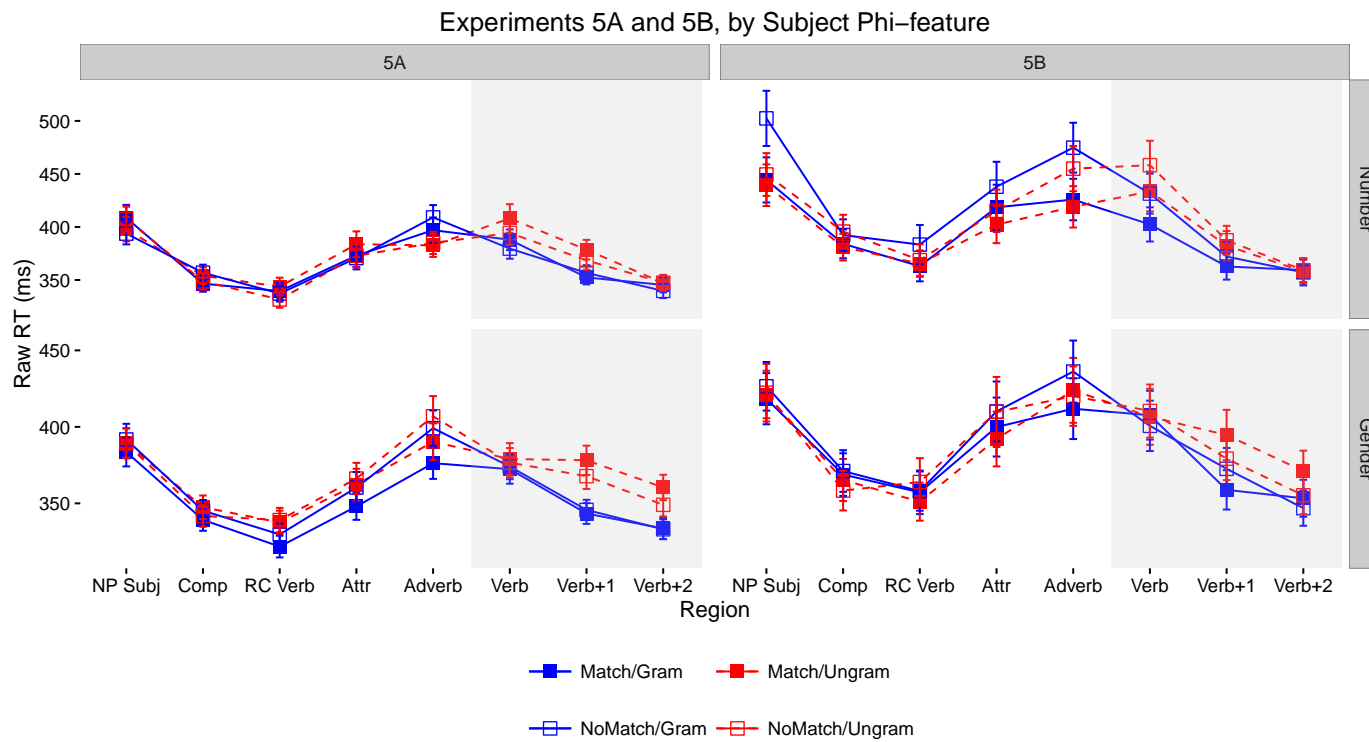
Figure 7: Mean raw reading times from Experiment 4 for all conditions and regions by subject number. Error bars represent the standard error of the condition mean across participant averages.

whereas a reliable GRAMMATICALITY effect for GENDER emerged only in the *Verb+1* region and continued into the *Verb+2* region.

When it comes to the GENDER attraction effects in ungrammatical sentences, they were numerically observed in experiments 5A and 5B at the *Verb+1* region (10 ms and 15 ms), but in neither case the 95% CI technically excluded zero (its lower bound included −1 ms in 5A and −2 ms in 5B). They were, however, reliably observed in the *Verb+2* region (11 ms in 5A and 16 ms in 5B). There was no clear indication of GENDER attraction effects in grammatical sentences.

The results for NUMBER attraction effects in ungrammatical sentences was mixed. They were numerically observed in experiment 5A in both *Verb* and *Verb+1* region (14 ms and 10 ms respectively), but in neither case the 95% CI excluded zero (lower bound included −2 at the *Verb* region and −3 at the *Verb+1* region). They were, however, not even numerically observed in 5B (they were "reversed" in all critical regions). There was no clear indication of NUMBER attraction effects in grammatical sentences.

6.5. Discussion

Experiments 5A and 5B provide further support for the notion that GENDER also participates in illusory agreement, and that it exhibits the GRAMMATICAL ASYMME-TRY that has been described for NUMBER in other languages. The effect sizes in experiments 5A and 5B were nonetheless smaller than the ones that had been observed until now in the *Verb+1* region: 21 ms, 26 ms and 35 ms in experiments 1, 2A and 2B respectively, but only 10 ms in 5A and 15 ms in 5B. However, experiments 5A and 5B provided reliable evidence for GENDER attraction effects in region *Verb+2*, which thus far had only been observed in experiment 2B: −4 ms, 7 ms in experiments 1 and 2A, but 21 ms in 2B, 11 ms in 5A and 16 ms in 5B. This indicates that the GENDER attraction effect may spillover into the next critical region once it emerges. More importantly, in these five experiments, the GENDER attraction effect has reliably appeared *after* the *Verb* region, even when the GRAMMATICALITY effect appeared at the *Verb*.

When it comes to the NUMBER attraction effect, experiment 5A and 5B give conflicting results, much like experiments 4A and 4B. Experiment 5A provides a very similar pattern of results compared to experiment 4A, albeit with effects sizes of half of

59

the size, and with 95% CIs that do include zero, even if by little. Therefore, when combined with the results of Tucker et al. (2015) and experiment 4A, we observe evidence of a traditional NUMBER attraction effect. Moreover, this effect appears to occur immediately at the *Verb* region in all three experiments, occasionally spilling over into the subsequent critical region. More importantly, to the extent that we observe evidence of both NUMBER and GENDER attraction effects in experiment 5A, they occur in different regions (*Verb* and *Verb+1* for NUMBER, *Verb+1* and *Verb+2* for GENDER), as strongly suggested by the results of the experiments in which each feature was individually manipulated.

However, the above remarks should be tempered with questions of how to interpret the results of Experiment 5B, which joins experiment 4B as another direct replication in which no evidence of a NUMBER attraction effect is observed, even though a clear GRAMMATICALITY effect is. There are basically two ways of interpreting this apparent discrepancy: either the NUMBER attraction effect in Arabic is much smaller and/or less reliable than it is in other languages, or MSA, unlike other languages that have been tested, does not really accommodate illusory licensing of number agreement (which would imply that the results of experiments 4A and 5A, as well as those of Tucker et al. (2015) were type I errors). In order to adjudicate between these two alternatives, a meta–analysis will be conducted. Crucially, regardless of the results of the meta–analysis, it is clear that this would be another dimension in which the process of NUMBER would be different from the process of GENDER agreement in MSA.

7. Meta–analysis

In order to help make sense of the large number of results reported in the preceding eight experiments, we resort to a *meta–analysis* (Cooper et al., 2009; Cumming, 2014; Hunter & Schmidt, 2004; Rosenthal & Dimatteo, 2001). In this kind of analysis, we combine the results of multiple experiments testing the same hypothesis into a single joint summary that provides a less biased and better statistically grounded view of the cumulative evidence than just counting whether or not particular experiments exhibited or failed to exhibit the predicted pattern of results. This latter point is extremely

60

important, given the challenges a researcher faces when trying to combine the results from many different experiments dealing with the same hypothesis: on the one hand, researchers have been shown to both hold unrealistically high expectations of replication rates in cases where they assume or know the hypothesis under test to be true (Francis, 2012; Stanley & Spence, 2014) and be overconfident about the prospects of replication if they observe a statistically significant result (Gigerenzer, 2004; Haller & Krauss, 2002; Hoekstra et al., 2006; Oaks, 1986). In addition, researchers also often irrationally dismiss as false results that fail to show statistical significance (Hoekstra et al., 2006; Schmidt, 1996). Given these propensities, it is hard to imagine that researchers are in general well–equipped to conduct an unbiased review of many different findings about a hypothesis when several of them are in apparent conflict. On the other hand, the natural impulse of simply tallying "positive" versus "negative" results (i.e., "vote counting") is also, as a summary procedure, rife with statistical problems: not only has it low power, but its power actually *decreases*, tending to zero, as the number of results being evaluated increases (Hedges & Olkin, 1980).

Here, we opt to conduct a *fixed effects* meta–analysis (cf. Cooper et al., 2009) as opposed to a *random effects* alternative, for a few reasons. At a conceptual level, our goal is to primarily summarize the results of the eight experiments reported here, and not necessarily extrapolate from them. Relatedly, the eight experiments reported are either direct replications or extremely similar to each other in terms of their design, procedure, experimental materials, but also in terms of the population being tested — all students from the same university, tested within a period of twelve months. This also matches the conceptual assumptions of the *fixed effects* meta–analysis when compared to the *random effects* alternative. Moreover, because of the extreme similarity between studies and their related samples, it is unclear that the results of the meta–analysis would be generalizable on a *statistical basis* to other language populations that are not included in the meta–analysis. Finally, even though we report eight studies, that is a rather low number for a meta–analysis, and the *fixed-effect* model has a power advantage (Rosenthal & Dimatteo, 2001) compared to the alternative.

Given our research questions, we are interested in comparing the *attraction effects* for NUMBER and GENDER, and how they may vary as a function of their timing, ef-

61

fect size and susceptibility to the *grammatical* asymmetry and the *markedness asymmetry*. Therefore, we conduct eight meta–analysis, each on the three critical regions we have been focusing on: *Verb*, *Verb+1*, and *Verb+2*. Each analysis is focused on a specific agreement feature (NUMBER VS GENDER), a specific grammaticality level (GRAMMATICAL VS UNGRAMMATICAL sentences) and markedness status (SINGULAR VS PLURAL for NUMBER; MASCULINE VS FEMININE for GENDER). In each analysis, the studies were weighed by the inverse of their variance. All analyses were performed using the *metafor* package in the R programming language (Viechtbauer, 2010).

7.1. Meta–analysis of GENDER

Figure 8 displays the meta–analysis for GENDER attraction using unmarked (MASCULINE) subjects in ungrammatical versus grammatical sentences. The results are straightforward: there is a clear *grammaticality asymmetry* in that GENDER attraction only occurs in ungrammatical sentences. Moreover, GENDER attraction seems to occur in the two regions *after* the verb. The point–estimate effect size of the effect was 17 ms for the *Verb+1* region and 11 ms for the *Verb+2* region, both with 95% parametric CIs excluding zero.

Figure 9 displays the meta–analysis for GENDER attraction using marked (FEMININE) subjects in ungrammatical versus grammatical sentences. Unlike what has been shown for sentences with unmarked subjects, there is no clear GENDER attraction effect for sentences with marked subjects, and therefore there cannot be evidence for a *grammaticality asymmetry*. The only other notable effect is a "reverse" GENDER attraction effect for grammatical sentences in the *Verb+2* region.

7.1.1. Discussion

The meta–analysis shows clear evidence of a GENDER attraction effect that is susceptible to the *grammatical asymmetry* and likely to the *markedness asymmetry* as well. This effect is estimated to emerge only in the post–verbal regions, never in the *Verb* region itself.

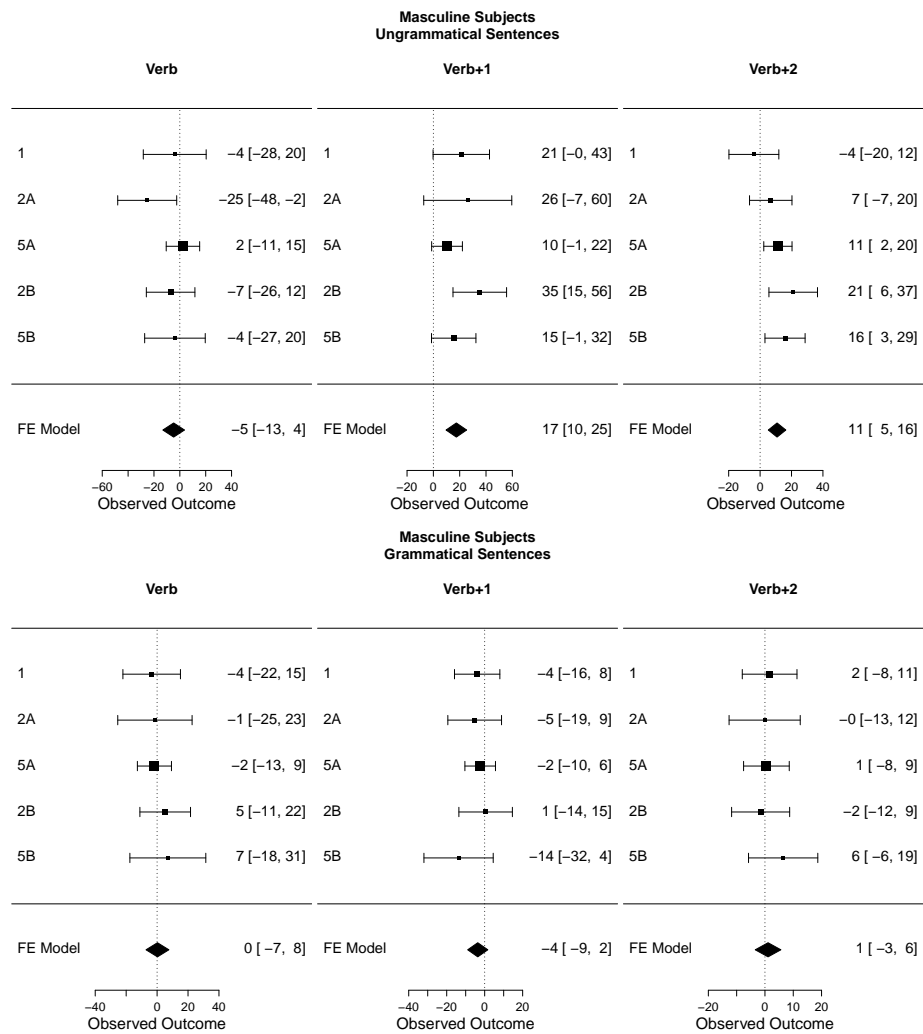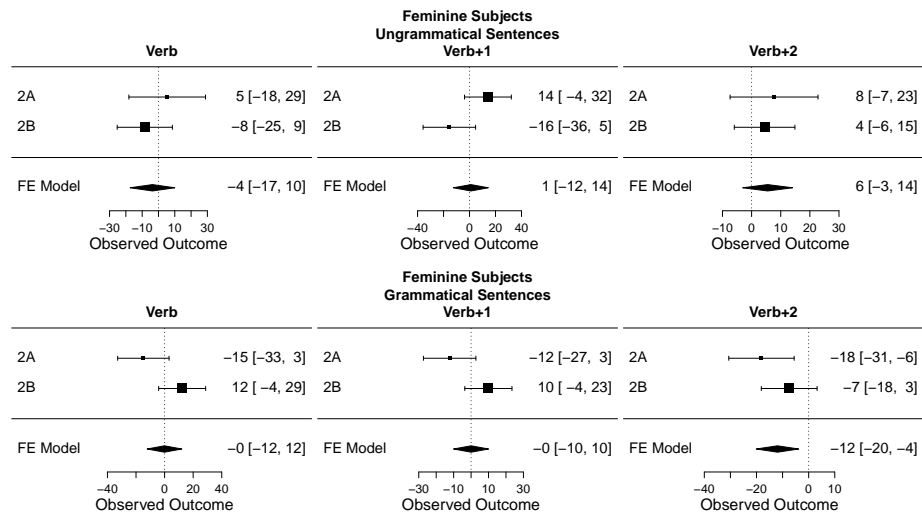Figure 8: Gender attraction effect in Ungrammatical and Grammatical sentences: Meta analysis for masculine subjects.

Feminine Subjects
Ungrammatical Sentences

Verb		Verb+1		Verb+2	
2A	5 [−18, 29]	2A	14 [−4, 32]	2A	8 [−7, 23]
2B	−8 [−25, 9]	2B	−16 [−36, 5]	2B	4 [−6, 15]
FE Model	−4 [−17, 10]	FE Model	1 [−12, 14]	FE Model	6 [−3, 14]

Feminine Subjects
Grammatical Sentences

Verb		Verb+1		Verb+2	
2A	−15 [−33, 3]	2A	−12 [−27, 3]	2A	−18 [−31, −6]
2B	12 [−4, 29]	2B	10 [−4, 23]	2B	−7 [−18, 3]
FE Model	−0 [−12, 12]	FE Model	−0 [−10, 10]	FE Model	−12 [−20, −4]

Figure 9: Gender attraction effect in Ungrammatical and Grammatical sentences: Meta analysis for feminine subjects.

7.2. Meta–analysis of NUMBER

For the meta–analysis for NUMBER attraction effect, we also included the results of Tucker et al. (2015), broken down by their subgroup analysis of SOUND/SUFFIXING plurals versus BROKEN/ABLAUTING plurals. The raw data from Tucker et al. (2015) was subjected to the same pre–processing steps as the other eight experiments.

Figure 10 displays the meta–analysis for NUMBER attraction using unmarked (SINGULAR) subjects in ungrammatical versus grammatical sentences. The results show a clear *grammaticality asymmetry* in that NUMBER attraction only occurs in ungrammatical sentences. Moreover, NUMBER attraction seems to occur immediately at the *Verb* region as well as its spillover region. The point–estimate effect sizes of the effect were 8 ms for the *Verb* region and 9 ms for the *Verb+1* region, both with 95% parametric CIs excluding zero, with the caveat that the lower bound of the attraction effect in the *Verb* region was .3 ms.

Figure 11 displays the meta–analysis for NUMBER attraction using marked (PLURAL) subjects in ungrammatical versus grammatical sentences. Unlike what has been shown for sentences with unmarked subjects, there is no clear NUMBER attraction effect for sentences with marked subjects. Thus, there cannot be evidence for a *grammaticality*

asymmetry either.

7.2.1. Discussion

The meta–analysis shows evidence of a NUMBER attraction effect that is susceptible to the *grammatical asymmetry* and likely to the *markedness asymmetry* as well. This effect is estimated to emerge immediately at the verb regions and spills over into the first post–verbal region. Compared to the GENDER attraction effect, the NUMBER attraction effect size is considerably smaller when the regions where each effect first emerges is compared (17 ms for GENDER vs 8 ms for NUMBER).

8. General Discussion

The results of the eight experiments and the meta–analysis reported here confirm the notion that errors in agreement dependency comprehension are, at their core, universal in scope. Despite the universality of the errors, however, the studies reported here have uncovered some important differences between number and gender agreement in comprehension which have ramifications for theories of agreement attraction.

8.1. Dimensions of Similarity

As the results of these eight studies show, whether or not one concludes that gender and number are subject to the same conditions of illusory licensing depends on which dimension one assesses similarity upon. Here we conclude that gender and number are largely qualitatively similar in their attraction profiles. Quantitatively, however, it appears that these two features attract differently.

Existence. The first and perhaps most obvious way in which gender and number can be similarly involved in attraction is the basic fact that both these features give rise to attraction RT profiles in the comprehension of verbs. In Experiments 1, 2A and 2B, the RT profiles at and immediately following the critical verbs include a facilitation to NO-MATCH/UNGRAM conditions relative to the large reading time spike seen in response to MATCH/UNGRAM conditions. This is the classic attraction profile in comprehension observed for number in Arabic in Tucker et al. (2015) as well as experiments 4A and
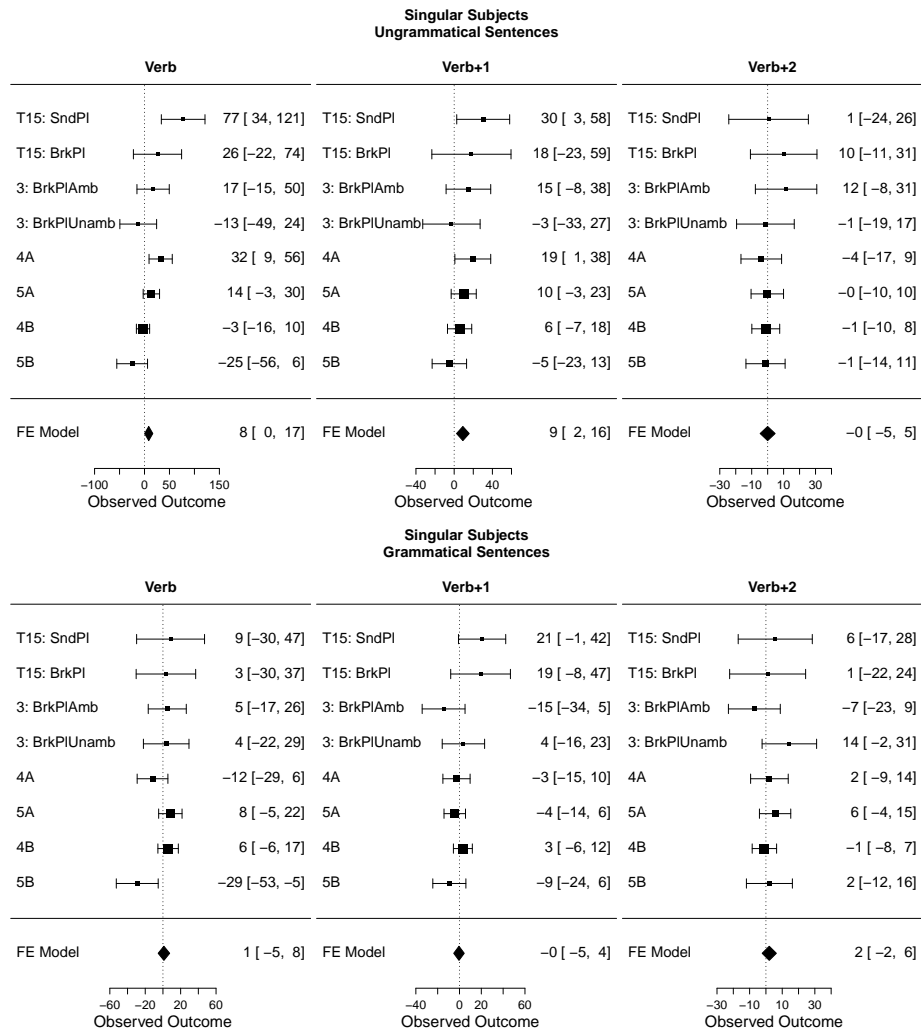
Figure 10: Number attraction effect in Ungrammatical and Grammatical sentences: Meta analysis for singular subjects.
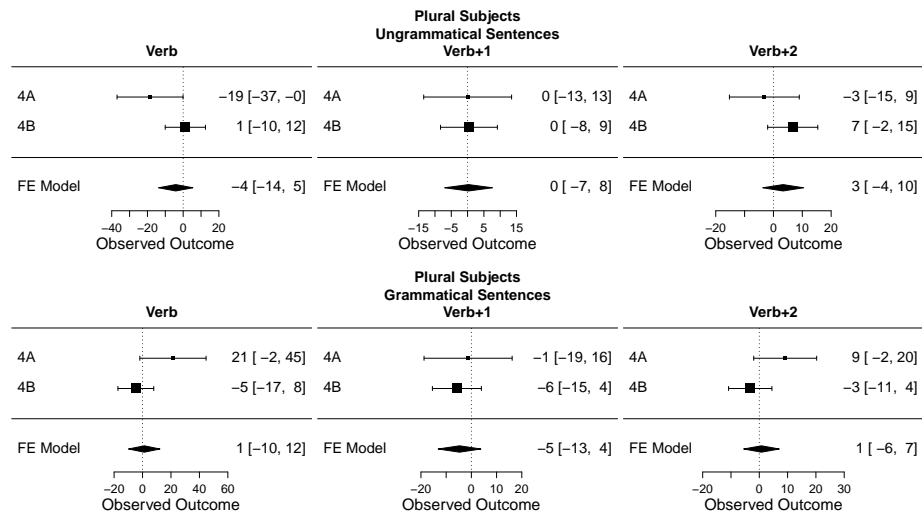
Figure 11: Number attraction effect in Ungrammatical and Grammatical sentences: Meta analysis for plural subjects.

4B, and in many other languages (Dillon et al., 2013; Malko & Slioussar, To Appear; Pearlmutter, 2000; Pearlmutter et al., 1999; Wagers et al., 2009; *i.a.*). This is an important conclusion despite its obviousness given that no major theory of attraction effects could, in principle or without alteration, ensure that grammatical number is subject to attraction effects in verbal comprehension but grammatical gender does not.

Grammatical Asymmetries. Another important dimension along which attraction for gender and number emerges as identical in our studies is the asymmetry of the attraction effects with respect to the grammaticality status of the verb. In all eight experiments reported here, attraction RT profiles, if they are present, are present only in ungrammatical sentences. *Modulo* experiments 3, 4B and 5B, where no NUMBER attraction appears to be present at all, throughout all other experiments a difference in the MATCH versus NO-MATCH conditions emerges only when the verb is grammatically unacceptable. While there is some contention about the generality of this finding (see Franck et al., 2015), here we can add five more experiments as well as a within–language meta–analysis to the list of those which do not observe attraction effects in grammatical sentences (*e.g.*, Dillon et al., 2013; Tanner et al., 2014; Wagers et al., 2009). As noted in the in-

troduction, one empirical point of distinction between competing theories of attraction effects has to do with the equivalency of attraction effects in both grammatical and ungrammatical sentences — process theories are arguably better-equipped to handle these asymmetries than representation theories, a point to which we return below.

Markedness Asymmetries. Another way in which gender and number emerge as similar across our experiments has to do with the presence of the asymmetry that we have been calling markedness-based. In MSA, plural number is marked (in the sense of Trubetskoy, 1939/1958) relative to singular and feminine gender is marked relative to masculine. If gender and number are equivalent along the markedness dimension and in line with the markedness results reported for English (Bock & Miller, 1991; Eberhard, 1997), one would expect that attraction RT profiles are present and/or strongest for singular subjects with plural attractors and masculine subjects with feminine attractors. In contrast, one would expect attraction RT profiles to be absent or greatly reduced for plural subjects with singular attractors and feminine subjects with masculine attractors. Even though we have less data, and therefore less confidence in this conclusion, the meta–analysis shows little evidence for attraction effects when subjects carry an unmarked agreement feature, either for ungrammatical or grammatical sentences, while the evidence of attraction when the subjects carry a marked agreement feature is much stronger.

RT Effect Size. Another important dimension along which to assess the similarity of attraction effects is the dimension of effect size. *A priori*, one could imagine two distinct quantities which define the quantity to be examined: the number of attraction incidents and the amount of reading time attraction change. Since this study involved only reading time, we have no direct way to assess the former, as individual trials do not provide such information given the latin square design (ensuring no subject saw all the relevant conditions). Here, the meta–analysis results are reasonably clear: the effect size for NUMBER attraction and GENDER attraction does seem to be different. In the region where they first emerge, the former is half the size of the latter (8 ms vs 17 ms), but they seem to align in their respective spillover regions (9 ms vs 11 ms).

68

Interestingly, the estimated effect size for GENDER attraction is close to the one estimated in a recent meta–analysis of NUMBER attraction effects (Jäger et al., In Press): 17 ms here and 22 ms there. On the other, our estimated effect size for NUMBER attraction is much smaller (8 ms) than these two, and in fact would fall outside of the Credible Interval provided by Jäger et al. (In Press).

Effect Timing. Finally, it is worth considering whether the studies reported here provide any evidence for similarities or differences in timing in the appearance of agreement attraction effects. This is especially topical given the recent observations by Lago et al. (2015) that attraction effects can, in principle, appear after grammaticality effects in self-paced reading data. The question therefore arises as to whether gender and number show the appearance of attraction effects after grammaticality effects, and whether these profiles are the same or different.

Although the nature of the self-paced reading methodology employed in this study is suboptimal to fully resolve this issue, our results are nonetheless replicable enough to strongly suggest that the time–courses of attraction effects are different between GENDER and NUMBER (see also Figures 8 and 10). The former emerges reliably at the *Verb+1* region, while the latter emerges systematically at the *Verb* region whenever it is found. Interestingly, in three out of five experiments the GENDER attraction effect occurs in the region following the one where the grammaticality effect occurs.

8.2. Implications for Representing Features and Cues

Given the importance of representational commitments to both major kinds of theories of agreement attraction, it is crucial to consider whether our results could be accounted for in ways neutral to processing theories by way of changes to the ways that linguistic features are used in processing or mapped onto cues for memory retrieval. Here we consider two approaches to featural representation: (1) an approach which localizes the difference in the valency of feature representation (*i.e.,* Fuchs et al., 2015) and (2) one which localizes the difference in the location of gender information in grammar and processing (*i.e.,* Deutsch & Dank, 2011).

One approach to asymmetries between gender and number would be to assert that these features are simply represented differently in grammar or processing. For in-

stance, one could follow the approach of Fuchs et al. (2015) and assert that agreement features which show markedness asymmetries are PRIVATIVE — they are represented only in the marked value and not present otherwise. Features which do not show markedness contrasts are instead EQUIPOLLENT — they are represented by the presence of features regardless of markedness. Fuchs et al. (2015), use this idea to represent the differential activity of gender and number in Spanish agreement attraction, and one could extend it to Arabic by positing that gender is bivalent ([± MASC]) whereas number is privative ([PL] or ∅). From this assumption one could tie either misrepresentation or cue-based retrieval models to this featural specification.

The problem with this approach is that it is not sufficiently supported by the distributional properties of the MSA grammar. For one, equipollent featural representations are typically used to encode three-way contrasts, which gender is not in Arabic — there is no neuter gender in MSA. While this is not an insurmountable representational issue, it does mean that the only evidence for equipollent gender in MSA would be the very markedness patterns that must be explained. A larger issue, however, has to do with number. Grammatical number in MSA is not a two-way system, but instead a three-way system, including a morphological DUAL which is used for sets of cardinality two (Ryding, 2005). Three-way distinctions are more difficult to encode in privative feature systems since privative representations are meant to encode two-way contrasts. What is needed to properly assess this question is a comparison of our results concerning singular and plural number with similar data concerning the dual in MSA.

A different approach to these issues would be to assert that gender and number are represented in different components of the processing system. For instance, Deutsch & Dank (2011) suggest that one could capture an identical pattern to our results but for Hebrew gender and number production data by assuming that gender is an inherent property of the lexical lemma and not part of the morpho-phonological properties of the word (see also Sicuro Corrêa et al., 2004). Grammatical number, on the other hand, is not an inherent property of the lemma, since any given lemma can be either singular or plural. Since the computation of number on nominals is part of the morpho-phonological process translating a lemma into a spoken word, it can be subject to principles of morpho-phonological markedness (see Deutsch & Dank, 2011 for details on a

70

particular implementation of this idea in the Marking and Morphing model of Eberhard et al., 2005).

This approach certainly has some conceptual and empirical intuitiveness given that grammatical gender is not typically meaningful in the same way as grammatical number and that the approach was designed to account for a similar set of facts in a closely related language — Modern Hebrew. However, while this approach is very well-suited to gating the presence or absence of attraction based upon markedness, it is incapable of attenuating or strengthening attraction effects in similar dimensions. Our results show that gender and number attraction effects are not simply different in quality, they are different in quantity, as well. In fact, one can step back and see that *any* attempt to explain our results based upon the representational structure or geometry of the features involved will be incapable of explaining the quantitative results we have observed in this study.

8.3. *Implications for Theories of Attraction*

Given that a simple representational change is not sufficient for explaining the differential effects that we observe for agreement attraction with gender and number, we now return to the two major classes of theories discussed in the introduction in light of these results. While both kinds of theories require nontrivial changes to their architectures to account for differences between gender and number, we ultimately suggest that cue-based retrieval theories require less drastic modifications (*i.e.*, such as those proposed in Engelmann et al., 2015). What is over-arching to both discussions, however, is a need for a shift in the empirical domain of investigation for agreement attraction studies in particular and illusory dependency licensing studies in general: whereas current work has derived much of its insights from studies of the qualitative profiles of number in Indo-European languages, we believe that much insight can be gained by examining typologically diverse languages/features as well as the *quantitative* patterns of attraction in several comprehension methodologies.

71

8.3.1. (Mis)representation Theories

Our results present two major challenges for misrepresentation theories broadly speaking: (1) the differential quantitative strength of gender and number attraction and (2) the absence of agreement attraction RT profiles in grammatical sentences. Both of these challenges stem from a similar prediction common to representational theories: since theories that attribute attraction effects to failures of representation take the agreement process itself to be undisturbed when attraction occurs, they predict parity of attraction effects across identically represented subject NPs. What causes attraction in, *e.g.*, the theories of Eberhard et al. (2005); Franck et al. (2002); Nicol et al. (1997); Vigliocco & Nicol (1998) is a process by which structural representations of the subject are malleable enough to allow features of the attractor to be copied erroneously to the verb by the normal processes of subject-verb agreement. It is a corollary of this assumption that attraction should occur in equal measure in structurally identical subject NPs (Wagers et al., 2009).

But this is not what we observe for gender attraction. Our results suggest a smaller quantitative profile of attraction for number in MSA than for gender. Given that our experiments involved structurally identical subject and attractor NPs across all experiments, these results cannot be explained by reference to different structural configurations leaking attractor features in different strengths. Number attraction appears diminished in strength relative to number when compared directly in a subject relative clause configuration in both cases.

Here one could appeal to the quantitative Marking and Morphing Model of Eberhard et al. (2005) to attempt to derive these effects from our use of animate human-denoting NPs in all four experiments. In the Marking and Morphing Model, one of the ways that structural representations are malleable is that top-level phrases are assigned featural strengths based upon a function of the features of their contained constituents in addition to their own feature values. An NP such as *The key to the cabinets* therefore contains some residual plurality given the plural denotation and marking of *cabinets*. However, with animate human-denoting NPs where grammatical gender is arguably semantically contentful, models such as the Marking and Morphing model should predict *stronger*

attraction for gender given the clear morphological marking and semantic content of feminine gender in our MSA stimuli.

More broadly, however, both quantitative and qualitative misrepresentation models struggle with the lack of attraction consistently observed in our studies in grammatical sentences. As Wagers et al. (2009) have argued, these models cannot predict anything other than parity in the rates of attraction, since the malleable or leaky representation of subjects occurs blind to what happens at the verb. Here it is not even enough to dispute the qualitative appearance of grammatical agreement attraction (*pace* Franck et al., 2015), as we have shown a large quantitative difference which cannot be accounted for under misrepresentation approaches. The only misrepresentation approach which could account for these sorts of effects is the degraded memory representation model of Staub (2009, 2010), though this model too needs modifications to successfully predict differential quantitative strengths of attraction for number and gender.

8.3.2. Cue-based Retrieval Theories

Cue-based retrieval theories, on the other hand, deal much more successfully with the lack of attraction in grammatical sentences. In these models (such as those deriving from Lewis & Vasishth, 2005 and Badecker & Lewis, 2007), attraction occurs when cue-mismatches between subjects and attractors lead to the erroneous retrieval of the attractor during a working memory retrieval event triggered by the verb. There are two distinct ways to concretize this idea: either the retrieval event occurs in all instances or it only occurs upon the presentation of ungrammatical verbs. In either case, however, grammatical attraction is not predicted. In the first case, the complete cue match between subject and grammatical verb causes the probability of attractor retrieval to plummet relative to ungrammatical sentences. In the latter case, one simply constructs the model not to consider attraction in grammatical sentences by fiat.

What is less easily representable in these theories is the lack of quantitative symmetry between gender and number attraction in our results. Cue-based retrieval models are dependent upon the cue structure posited in the model. Setting the issues of markedness discussed in the previous section aside, it is difficult to see how cue structures can be posited that simultaneously cause attraction and also do so at different strengths. In the

73

model of Lewis & Vasishth (2005), for example, the strength of a cue can only be modulated as a function of the number of other cues in memory, not the intrinsic properties of the cue itself. Thus, gender could be relatively strong relative to number, but only if number uniformly occurred as a cue in contexts where more cues were available in the system in general. Obviously, this is not a feasible assumption for MSA, where gender and number always co-occur (Ryding, 2005).

However, cue–based retrieval models are the subject of much active research and are constantly evolving. Engelmann et al. (2015), for instance, have proposed two new mechanisms based upon a literature review in order to account for unrelated effects in the literature, DISTRACTOR PROMINENCE (a quantitative adjustment giving more activation to attractors as functions of their position and discourse prominence) and CUE CONFUSABILITY (the ability for cues to be related to features quantitatively instead of categorically). The latter of these ideas is an intriguing notion, though the specific formulation of the idea in Engelmann et al. (2015) would not accommodate our results. Further work is needed to see if the cue–based retrieval models can be enriched in such ways to predict quantitatively different effects based on grammatically equipotent linguistic features but it does seem clear what this work would look like: an expanded theory of cue confusability which allows cues to be differentially weighted in isolation, as well as differentially confusible with one another.

Finally, it is worth noting that neither misrepresentation nor cue-based retrieval models could account for differences in timing of gender and number attraction effects. Attraction is a *verbal* process, meaning that the representations and processes responsible for these effects should be keyed at the verb, not later. As we observed, it is possible that our evidence hints at the delayed appearance of attraction for gender relative to attraction for number. Self-paced reading methodologies commonly involve spillover effects with no clear theoretical explanation, but even when taking these into account, the combined data from our eight experiments strongly suggests a *Verb* locus for the NUMBER attraction effect and a *Verb+1* locus for the GENDER attraction effect.

74

8.4. Conclusions

We have demonstrated that verbal gender agreement attraction occurs in comprehension. Moreover, these results obtain in an inflectionally rich language in relative clause configurations where attraction should be smaller in effect, all else equal. We have also demonstrated that attraction for gender and number is not identical in Arabic. Quantitatively, we demonstrated that agreement attraction for gender is stronger relative to number attraction but occurs later in time. We also added additional evidence to the body of work suggesting that comprehension attraction effects do not occur in grammatical sentences, for gender or number. These results were shown to be largely more compatible with cue-based retrieval models over misrepresentation models insofar as the former are capable of accounting for grammaticality asymmetries and require fewer alterations to account for quantitative differences among agreement features. Finally, we suggested that much progress can be made in theorizing about attraction by moving from qualitative work on grammatical number to quantitative work on other features and languages.

Complete Materials – Experiments 1–2

A.1. The translator who helped the manager occasionally speaks five languages fluently.

المترجم الذي ساعد المدير أحياناً يتكلم خمس لغات بفصاحة.

A.2. The student who saw the professor yesterday studied electrical engineering at the university.

الطالب الذي رأى الأستاذ بالأمسِ درس الهندسة الكهربائية في الجامعة.

A.3. The engineer who met the scientist by chance is working on a new invention.

المهندس الذي استقبل العالِم بالصدفةِ يعمل على ابتكار جديد.

A.4. The cook who scolded the waiter forcefully works in an expensive restaurant during the summer.

الطباخ الذي وبّخ النادل بشدةٍ يشتغل في مطعم غالٍ خلال الصيف.

A.5. The analyst who advised the minister intelligently discusses the Palestinian issue in depth.

المحلل الذي نصح الوزير بذكاءٍ يتناول القضية الفلسطينية بعمق.

A.6. The child who saw the prince before visits the royal family each week.

الطفل الذي رأى الأمير سلفاً يزور العائلة الملكية كل أسبوع.

A.7. The teacher who taught the child dedicatedly attended the graduation party of the students.

المعلم الذي علّم الطفل بتفانٍ حضر حفل تخرج الطلاب.

A.8. The consultant who warned the president yesterday found a solution for the financial problem.

المستشار الذي حذّر الرئيس بالأمسِ وجد الحل للمشكلة المالية.

A.9. The driver who accompanied the ambassador regularly works seven days a week.

السائق الذي رافق السفير بانتظامٍ يعمل سبعة أيام في الأسبوع

A.10. The jailor who tortured the prisoner constantly cleans the cells nightly.

السجّان الذي عذّب السجين باستمرارٍ ينظف الزنازين كل ليلة.

A.11. The broadcaster who talked to the activist yesterday trained in a famous company.

المذيع الذي كلم الناشط بالأمس تدرب في شركة مشهورة.

A.12. The employee who helped the colleague humbly gained the confidence of colleagues at work.

الموظف الذي ساعد الزميل بتواضع استطاع كسب ثقة باقي الزملاء في العمل.

A.13. The singer who challenged the poet arrogantly has lost for not showing up on time.

المنشد الذي تحدى الشاعر بغرور خسر لعدم حضوره في الوقت المحدد.

A.14. The beginner who questioned the expert daily has acquired good experience.

المبتدئ الذي سأل الخبير يومياً اكتسب خبرة جيدة.

A.15. The man who hosted the friend with pleasure slept in the basement of the house.

الرجل الذي استضاف الصديق بسرور نام في الطابق السفلي من المنزل.

A.16. The patient who consulted the doctor yesterday returned home satisfied.

المريض الذي استشار الطبيب بالأمس عاد الى المنزل مطمئناً.

A.17. The teacher who met the writer happily likes reading about literature.

المعلم الذي استقبل الأديب بسعادةٍ يحب القراءة عن الأدب.

A.18. The manager who phoned the partner in the morning plans to expand the branches of the company.

المدير الذي هاتف الشريك بالصباح يخطط لتوسيع فروع الشركة.

A.19. The chef who invited the guest in the evening masters preparing various delicious dishes.

الطاهي الذي دعا النزيل بالمساء يتقن إعداد أصناف لذيذة من الطعام.

A.20. The child who watched the magician with amazement applauded hysterically during the show.

الطفل الذي رأى الساحر بانبهار صفق بشدة خلال العرض.

A.21. The young man who helped the wounded man in the morning spends every week reading.

الشاب الذي ساعد الجريح صباحاً يقضي كل الأسبوع في المطالعة.

A.22. The doctor who healed the captive quickly is joining the national guard (army).

الطبيب الذي عالج الأسير بسرعةٍ ينخرط في فرقة الدفاع المدني.

A.23. The ruler who imprisoned the criminal previously practiced justice and equality among the people.

الحاكم الذي سجن المجرم سابقاً مارس العدل والمساواة على الشعب.

A.24. The coach who trained the partner efficiently possesses more than 10-years of experience.

المدرب الذي درّب الرفيق بإتقان يمتلك خبرة تفوق العشر سنين.

A.25. The Imam who advised the neighbor wisely lives in the local neighborhood.

الإمام الذي نصح الجار بحكمةٍ يعيش في المنطقة المجاورة.

A.26. The policeman who questioned the murderer seriously asked the pedestrians the reasons for the crime.

الشرطي الذي استجوب القاتل بجديّةٍ سأل المشاة عن سبب الجريمة.

A.27. The lawyer who called the inheritor in the morning discussed the issue of the inheritance distribution.

المحامي الذي استدعى الوريث بالصباح ناقش موضوع تقسيم الإرث.

A.28. The teacher who taught the student well worked as a host in television.

الأستاذ الذي علّم الطالب جيداً عمل كمذيع في التلفزيون.

A.29. The immigrant who spoke with the visitor for a long time feels nostalgic for the country always.

المهاجر الذي حدّث الزائر طويلاً يشعر بالحنين للوطن دائماً.

A.30. The journalist who interviewed the leader persistently publishes the article in the national newspaper.

الصحافي الذي استجوب الحاكم بالحاح نشر المقال في الجريدة الوطنية.

A.31. The pilot who greeted the airline attendant warmly asked many questions during the journey.

الطيار الذي حيّا المضيف بحرارةٍ سأل أسئلة كثيرة خلال الرحلة.

A.32. The man who talked to the monk intelligently works in the post office.

الرجل الذي كلم الراهب بفطنةٍ يشتغل في مكتب البريد.

A.33. The policeman who arrested the driver quickly helps the pedestrians in crossing the street.

الشرطي الذي اعتقل السائق بسرعةٍ يساعد المارة على عبور الشارع.

A.34. The lawyer who startled the witness cunningly stopped the accusation of his client in the court.

المحامي الذي أربك الشاهد بدهاءٍ أبعد التهم عن موكله في المحكمة.

A.35. The worker who helped the soldier yesterday drives a large truck for the company.

العامل الذي ساعد الجندي بالأمسِ يَسوق شاحنة كبيرة لفائدة الشركة.

A.36. The artist who served the king devotedly gave a portrait to the ambassador of the United States.

الفنان الذي خدم الملك بتفانٍ تبرع بلوحة لسفير الولايات المتحدة.

A.37. The journalist who hosted the star brilliantly raised very embarrassing questions.

الإعلامي الذي استضاف النجم بتألّق طرح أسئلة محرجة جداً.

A.38. The nurse who is treating the patient carefully studies at the university hospital.

الممرض الذي عالج المريض بعنايةٍ يدرس في مستشفى الجامعة.

A.39. The seller who thanked the customer enthusiastically was happy with the large profit.

البائع الذي شكر الزبون بحماسٍ فرح بالفائدة الكبيرة.

A.40. The coach who was very interested in the player worked at the National Fencing Academy.

المدرب الذي اهتم باللاعب جداً اشتغل في الأكاديمية الوطنية للمبارزة.

A.41. The soldier who met the policewoman yesterday loved the atmosphere at the air base.

الجندي الذي قابل الشرطي بالأمسِ أحبّ العمل في القاعدة الجوية.

A.42. The singer who met the dancer previously sings with the city orchestra.

المطرب الذي استقبل الراقص سابقاً يغني مع أوركسترا المدينة.

A.43. The producer who enthusiastically chose the actress produces a film every month.

المخرج الذي اختار الممثل بشغفٍ ينتج فيلما كل شهر.

A.44. The maid who helped the caregiver earnestly cleans the rooms seven days a week.

الخادم الذي ساعد المربي بجدٍّ ينظف الغرف سبعة أيام في الأسبوع.

A.45. The announcer who invited the historian nicely presented a program on TV.

المذيع الذي دعا المؤرّخ بلطفٍ قدم برنامجاً في التلفزيون.

A.46. The assistant who served the pharmacist loyally collected all of the old reports.

المساعِد الذي خدم الصيدلانيّ بإخلاصٍ جمع كل التقارير القديمة.

A.47. The journalist who interviewed the winner by chance writes in many newspapers.

الصحفي الذي قابل الفائز صدفةً يكتب في صحف كثيرة.

A.48. The bedouin who visited the farmer at night lives in the middle of the desert.

البدوي الذي زار المُزارع ليلاً يسكن في وسط الصحراء.

Complete Materials — Experiment 3

AMBIGUOUS	UNAMBIGUOUS
$C_1aC_2aC_3a$	$C_1uC_2aC_3aa?$
$C_1aC_2iiC_3$	$?aC_1aaC_2iC_3a$
$C_1iC_2aaC_3$	$C_1awaaC_2iC_3$
$C_1iC_2C_3aan$	$?aC_1C_2aaC_3$
$C_1uC_2C_3aan$	$?aC_1C_2iC_3aa?$
$C_1aC_2C_3aa$	
$C_1uC_2C_2aaC_3$	
$C_1uC_2uuC_3$	

Table .1: Templates and ambiguity assignments for broken plural templates in Experiment 3.

B.1. The child who watched the magician with amazement applauded hysterically during the show.

(السحرة) الطفل الذي رأى الساحر بانبهار صفق بشدة خلال العرض.

B.2. The investigator who grabbed the robber at night installs listening devices every-where.

(الخونة) المحقق الذي أمسك الخائن بالليل يضع أجهزة تنصت في كل مكان.

B.3. The young man who helped the wounded man in the morning every week reading.

(الجرحى) الشاب الذي ساعد الجريح صباحاً يقضي كل الأسبوع في المطالعة.

B.4. The doctor who healed the captive quickly is joining the national guard (army).

(الأسرى) الطبيب الذي عالج الأسير بسرعةٍ ينخرط في فرقة الدفاع المدني.

B.5. The ruler who freed the slave in the past practiced justice and equality among the people.

(العبيد) الحاكم الذي أعتق العبد قديمًا مارس العدل والمساواة على الشعب.

B.6. The driver who transported the pilgrim at noon drove the car very quickly.

(الحجيج) السائق الذي أخذ الحاج بالظهيرةٍ قاد السيارة بسرعة فائقة.

B.7. The carpenter who employed the man regularly made wonderful furniture for the exhibition.

(الرجال) النجار الذي شغّل الرجل بانتظام أعدّ قطع أثاث رائعة للمعرض.

B.8. The coach who trained the partner efficiently possesses more than 10-years of experience.

(الرفاق) المدرب الذي درّب الرفيق بإتقان يمتلك خبرة تفوق العشر سنين.

B.9. The tv-host who presented the bridegroom skillfully met with the minister at the New Year's party.

(العرسان) المذيع الذي قدّم العريس ببراعةٍ قابل الوزير في حفل رأس السنة.

B.10. The Imam who advised the neighbor wisely lives in the local neighborhood.

(الجيران) الإمام الذي نصح الجار بحكمةٍ يعيش في المنطقة المجاورة.

B.11. The policeman who questioned the murderer harshly asked the pedestrians about the reasons for the crime.

(القتلة) الشرطي الذي استجوب القاتل بجديّةٍ سأل المشاة عن سبب الجريمة.

B.12. The lawyer who called the inheritor in the morning discussed the issue of the inheritance distribution.

(الورثة) المحامي الذي استدعى الوريث بالصباح ناقش موضوع تقسيم الإرث.

B.13. The teacher who taught the student well worked as a host in television.

(الطلاب) الأستاذ الذي علّم الطالب جيداً عمل كمذيع في التلفزيون.

B.14. The businessman who trained the workers frequently learned English in the UK.

(العمال) التاجر الذي درّب العامل كثيراً تعلم اللغة الانجليزية في بريطانيا.

B.15. The (football) player who admonished the referee angrily won the prize of best player.

(الحكام) اللاعب الذي عارض الحَكم بغضبٍ فاز بجائزة أفضل لاعب.

B.16. The immigrant who spoke with the visitor for a long time feels nostalgic for the country always.

(الزوار) المهاجر الذي حدّث الزائر طويلاً يشعر بالحنين للوطن دائماً.

B.17. The journalist who interviewed the leader persistently publishes the article in the national newspaper.

(الحكام) الصحافي الذي استجوب الحاكم بالحاح نشر المقال في الجريدة الوطنية.

B.18. The pilot who greeted the knight warmly asked many questions to ask during the journey.

(الفرسان) الطيار الذي حيّا الفارس بحرارةٍ سأل أسئلة كثيرة خلال الرحلة.

B.19. The man who talked to the monk intelligently works in the post office.

(الرهبان) الرجل الذي كلم الراهب بفطنةٍ يشتغل في مكتب البريد.

B.20. The policeman who arrested the thief quickly helps the pedestrians in crossing the street.

(اللصوص) الشرطي الذي اعتقل اللص بسرعةٍ يساعد المارة على عبور الشارع.

B.21. The lawyer who startled the witness cunningly stopped the accusation of his client in the court.

(الشهود) المحامي الذي أربك الشاهد بدهاءٍ أبعد التهم عن موكله في المحكمة.

B.22. The worker who helped the soldier yesterday drives a large truck for the company.

(الجنود) العامل الذي ساعد الجندي بالأمسِ يَسوق شاحنة كبيرة لفائدة الشركة.

B.23. The artist who served the king devotedly gave a portrait to the ambassador of the United States.

(الملوك) الفنان الذي خدم الملك بتفانٍ تبرع بلوحة لسفير الولايات المتحدة.

B.24. The journalist who hosted the star brilliantly raised very embarrassing questions.

(النجوم) الإعلامي الذي استضاف النجم بتألق طرح أسئلة محرجة جداً.

B.25. The translator who worked for the manager occasionally speaks five languages fluently.

(المدراء) المترجم الذي ساعد المدير أحياناً يتكلم خمس لغات بفصاحة.

B.26. The student who saw the professor yesterday studied electrical engineering at the university.

(الأساتذة) الطالب الذي رأى الأستاذ بالأمسِ درس الهندسة الكهربائية في الجامعة.

B.27. The engineer who met the scientist by chance is working on a new invention.

(العلماء) المهندس الذي استقبل العالمِ بالصدفةِ يعمل على ابتكار جديد.

B.28. The cook who scolded the waiter forcefully works in an expensive restaurant during the summer.

(النوادل) الطباخ الذي وبّخ النادل بشدةٍ يشتغل في مطعم غالٍ خلال الصيف.

B.29. The analyst who advised the minister intelligently discusses the Palestinian issue in depth.

(الوزراء) المحلل الذي نصح الوزير بذكاءٍ يتناول القضية الفلسطينية بعمق.

B.30. The child who saw the prince before visits the royal family each week.

(الأمراء) الطفل الذي رأى الأمير سلفاً يزور العائلة الملكية كل أسبوع.

B.31. The teacher who taught the child dedicatedly attended the graduation party of the students.

(الأطفال) المعلم الذي علّم الطفل بتفانٍ حضر حفل تخرج الطلاب.

B.32. The criminal who attacked the boy viciously breaks through the checkpoint every night.

(الأولاد) المجرم الذي هاجم الولد بشراسةٍ يخترق نقطة التفتيش كل ليلة.

B.33. The consultant who warned the president yesterday found a solution for the financial problem.

(الرؤساء) المستشار الذي حذّر الرئيس بالأمسِ وجد الحل للمشكلة المالية.

B.34. The driver who accompanied the ambassador regularly works seven days a week.

(السفراء) السائق الذي رافق السفير بانتظامٍ يعمل سبعة أيام في الأسبوع.

B.35. The jailor who tortured the prisoner constantly cleans the cells nightly.

(السجناء) السجّان الذي عذّب السجين باستمرارٍ ينظف الزنازين كل ليلة.

B.36. The broadcaster who talked to the activist yesterday trained in a famous company.

(النشطاء) المذيع الذي كلم الناشط بالأمس تدرب في شركة مشهورة.

B.37. The employee who helped the colleague humbly gained the confidence of colleagues at work.

(الزملاء) الموظف الذي ساعد الزميل بتواضع استطاع كسب ثقة باقي الزملاء في العمل.

B.38. The singer who challenged the poet arrogantly has lost for not showing up on time.

(الشعراء) المنشد الذي تحدى الشاعر بغرور خسر لعدم حضوره في الوقت المحدد.

B.39. The man who consulted the forgiver yesterday wants retribution of sin/guilt.

(الشفعاء) الرجل الذي استشار الشفيع البارحة يريد التكفير عن الذنب.

B.40. The old man who has mischievously insulted the scholar strives to create problems.

(الفقهاء) العجوز الذي أهان الفقيه بخبث يسعى الى افتعال المشاكل.

B.41. The beginner who questioned the expert daily has acquired good experience.

(الخُبراء) المبتدئ الذي سأل الخبير يومياً اكتسب خبرة جيدة.

B.42. The man who hosted the friend with pleasure slept in the basement of the house.

(الأصدقاء) الرجل الذي استضاف الصديق بسرور نام في الطابق السفلي من المنزل.

B.43. The patient who consulted the doctor yesterday returned home satisfied.

(الأطباء) المريض الذي استشار الطبيب بالأمس عاد الى المنزل مطمئن.

B.44. The teacher who met the writer happily likes reading about literature.

(الأدباء) المعلم الذي استقبل الأديب بسعادةٍ يحب القراءة عن الأدب.

B.45. The representative who talked to the Khalif yesterday works hard to get a promotion.

(الخلفاء) النائب الذي كلم الخليفة بالأمس يجتهد في العمل للحصول على ترقية.

B.46. The president who hurriedly called the ally tries to reign the situation.

(الحلفاء) الرئيس الذي استدعى الحليف باستعجال يحاول التحكم بزمام الأمور.

B.47. The manager who phoned the partner in the morning plans to expand the branches of the company.

(الشركاء) المدير الذي هاتف الشريك بالصباح يخطط لتوسيع فروع الشركة.

B.48. The chef who invited the guest in the evening masters preparing various delicious dishes.

(النزلاء) الطاهي الذي دعا النزيل بالمساء يتقن إعداد أصناف لذيذة من الطعام.

Complete Materials — Experiment 4

C.1. The nurse who is treating the patient carefully studies at the university hospital.

الممرضة التي عالجت المريضة بعنايةٍ تدرس في مستشفى الجامعة.

C.2. The queen who looked after the princess recently appears in public every week.

الملكة التي اهتمت بالأميرة حديثاً تظهر في العلن كل أسبوع.

C.3. The seller who thanked the customer enthusiastically was happy with the large profit.

البائعة التي شكرت الزبونة بحماسٍ فرحت بالفائدة الكبيرة.

C.4. The novelist who mentioned the maid of honor accurately sells many books to the public.

الكاتبة التي ذكرت الوصيفة بدقةٍ تبيع كتبا كثيرة للجمهور.

C.5. The coach who was very interested in the player worked at the National Fencing Academy.

المدربة التي اهتمت باللاعبة جداً اشتغلت في الأكاديمية الوطنية للمبارزة.

C.6. The midwife who cared for the girl repeatedly volunteers at the university hospital.

القابلة التي اعتنت بالفتاة تكراراً تتطوع في مستشفى الجامعة.

C.7. The soldier (fem.) who met the policewoman yesterday loved the atmosphere at the air base.

الجندية التي قابلت الشرطية بالأمسِ أحبّت العمل في القاعدة الجوية .

C.8. The singer who met the dancer previously sings with the city orchestra.

المطربة التي استقبلت الراقصة سابقاً تغني مع أوركسترا المدينة.

C.9. The nanny who cared for the schoolgirl affectionately traveled to a new country.

المربية التي ربّت الطالبة بحنانٍ سافرت إلى بلدٍ جديد.

C.10. The producer who enthusiastically chose the actress produces a film every month.

المخرجة التي اختارت الممثلة بشغفٍ تنتج فيلما كل شهر.

C.11. The photographer who photographed the witch artistically published the photos in a new book.

المصورة التي صورت الساحرة بتفننٍ نشرت الصور في كتاب جديد.

C.12. The maid who helped the nanny earnestly cleans the rooms seven days a week.

الخادمة التي ساعدت المربية بجدٍّ تنظف الغرف سبعة أيام في الأسبوع.

C.13. The announcer who invited the historian nicely presented a program on TV.

المذيعة التي دعت المؤرّخة بلطفٍ قدمت برنامجاً في التلفاز.

C.14. The model who met the accountant repeatedly owns a lot of expensive clothes.

العارضة التي التقت بالمحاسبة تكراراً تملك كثيرا من الملابس الغالية.

C.15. The assistant who served the pharmacist loyally collected all of the old reports.

المساعِدة التي خدمت الصيدلانيّة بإخلاصٍ جمعت كل التقارير القديمة.

C.16. The journalist who interviewed the winner by chance writes in many newspapers.

الصحفية التي قابلت الفائزة صدفةٍ تكتب في صحف كثيرة.

C.17. The bedouin who visited the farmer at night goes to the middle of the desert.

البدوية التي زارت المزارعة ليلاً تذهب إلى وسط الصحراء.

C.18. The doctor who treated the girl recently discovered a cure for the terrible disease.

الطبيبة التي عالجت الطفلة مؤخراً اكتشفت شفاء للمرض الرهيب.

C.19. The artist who corresponded with the publisher eagerly desired a new contract.

الفنانة التي راسلت الناشرة بشغفٍ رغبت في عَقد جديد.

C.20. The student who admired the poet greatly read many poems last year.

التلميذة التي أُعجبت بالشاعرة بشدةٍ قرأت قصائد كثيرة العام الماضي.

C.21. The director who contacted the author during the day supervises many large projects.

المديرة التي اتصلت بالمؤلفة نهاراً تشرف على كثيرٍ من المشاريع الكبيرة.

C.22. The dean who summoned the professor angrily observed a problem in the university departments.

العميدة التي استدعت الأستاذة بغضبٍ لاحظت خللاً في أَقسام الجامعة.

C.23. The musician who accompanied the singer professionally played with the national music group.

الموسيقيّة التي رافقت المغنية بمهنيّةٍ عزفت مع الفرقة الوطنية للموسيقى.

C.24. The ambassador who hosted the delegate yearly spoke at the United Nations.

السفيرة التي استضافت المندوبة سنوياً تحدثت في الأمم المتحدة.

C.25. The grandmother who met the neighbor suddenly talked about the neighborhood issues.

الجدّة التي صادفت الجارة فجأةً تحاورت عن أمور الحي.

C.26. The student who met the manager yesterday got high grades in the remaining subjects.

الطالبة التي قابلت المديرة البارحة نالت درجات عالية في المواد المتبقية.

C.27. The accountant who talked to the employee harshly suffered from social problems.

المحاسبة التي حادثت الموظفة بصرامة عانت من مشاكل اجتماعية.

C.28. The study abroad student who thanked the official a lot studied at one of the best international universities.

المبتعثة التي شكرت المسؤولة بكثرة درست باحدى أرقى الجامعات الدولية.

C.29. The graduate who talked to the lecturer happily works for extra hours at the library.

الخريجة التي كلمت المحاضرة بسعادة تعمل ساعات اضافية في المكتبة.

C.30. The painter who excitedly interviewed the director painted wonderful paintings

الرسامة التي حاورت المخرجة بإثارة رسمت لوحات فنية رائعة.

C.31. The chef who lived next to the trader for a long time practices a cooking career skillfully.

الطاهية التي جاورت التاجرة مطولاً تمارس مهنة الطبخ بمهارة.

C.32. The visitor who talked to the guide in the morning gave a lecture about how to manage time.

الزائرة التي حدثت المرشدة صباحاً القت محاضرة عن كيفية تنظيم الوقت.

C.33. The teacher who visited the doctor yesterday masters speaking in Arabic and English.

المعلمة التي زارت الطبيبة بالأمس تُجيد التحدث باللغة العربية و الإنجليزية.

C.34. The lawyer who accused the guilty person angrily is trying to find the way to the truth.

المحامية التي اتهمت المذنبة بغضبٍ تحاول الوصول إلى الحقيقة.

C.35. The engineer who met the colleague daily aspires to get a job at a prominent company.

المهندسة التي قابلت الزميلة يومياً تسعى للحصول على وظيفة في شركة مرموقة.

C.36. The farmer who blamed the young lady yesterday loves working at the farm near the park.

المزارعة التي لامت الشابة بالأمسِ تحب العمل في المزرعة المجاورة للحديقة.

C.37. The beginner who helped the boss in the morning was hired for the military company.

المبتدئة التي ساعدت الرئيسة صباحاً توظفت في الشركة العسكرية.

C.38. The actress who met the interviewer in the past resigned from the acting career recently.

الممثلة التي قابلت المذيعة بالماضي استقالت من مهنة التمثيل مؤخراً.

C.39. The guard who talked to the pupil in the morning goes home late every day.

الحارسة التي كلمت التلميذة صباحاً تذهب إلى المنزل في ساعة متاخرة كل يوم.

C.40. The worker who gently scolded the maid cares about helping the needy.

العاملة التي وبخت الخادمة برفق تهتم بمساعدة المحتاجين.

C.41. The analyst who patiently waited for the reporter is trying to educate people about the importance of a clean environment.

المحللة التي انتظرت المراسلة بصبر تسعى لتوعية الشعب على أهمية نظافة البيئة.

C.42. The reporter who spoke to the plaintiff adeptly interviews the president at international conferences.

المراسلة التي سألت المدعية بنباهة تقابل رئيس الدولة في المؤتمرات العالمية.

C.43. The magician who talked to the lady quickly worked at the theater near the village.

الساحرة التي كلمت السيدة بسرعة عملت في المسرح المجاور للقرية.

C.44. The employee who accompanied the visitor in the morning stays at work until late.

الموظفة التي رافقت الزائرة بالصباح تبقى حتى ساعة متاخرة في العمل.

C.45. The novelist who pleasantly shook hands with the designer writes international and local novels about literature.

الروائية التي صافحت المصممة بسرور تكتب روايات عالمية و محلية في الادب.

C.46. The researcher who calmly called the detective provides money for charity society

الباحثة التي هاتفت المحققة بهدوء توفر المال للجمعيات الخيرية.

C.47. The coordinator who helped the guard devotedly was in the school courtyard.

المنسقة التي عاونت الحارسة بتفان تواجدت في ساحة المدرسة.

C.48. The candidate who pleasantly thanked the participant took part in the electoral campaign.

المرشحة التي شكرت المشتركة بسرور شاركت في الحملة الانتخابية.

C.49. The judge who decisively questioned the thief ruled fairly among people.

القاضية التي سألت السارقة بحزم حكمت بالعدل بين الناس.

C.50. The immigrant who answered the inspector anxiously faced difficulties at the check point.

المهاجرة التي أجابت المفتشة بقلق واجهت صعوبات عند نقطة التفتيش.

C.51. The young girl who helped the grandmother at night works at bakery for sweets.

الفتاة التي ساعدت الجدة بالليل تعمل في مخبز الحلويات.

C.52. The tourist who met the driver on the road loves traveling to different countries.

السائحة التي صادفت السائقة بالطريق تحب السفر إلى بلدان مختلفة.

C.53. The princess who intelligently answered the journalist owns many huge palaces.

الأميرة التي أجابت الصحافية بذكاء تمتلك عدة قصور كبيرة.

C.54. The client who consulted the lawyer nervously practiced painting for a few years.

الموكلة التي استشارت المحامية بتوتر مارست مهنة الرسم لعدة سنوات.

Appendix A. Complete Materials — Experiment 5

D.1. The nurse who is treating the patient carefully studies at the university hospital.

الممرضة التي عالجت المريضة بعنايةٍ تدرس في مستشفى الجامعة.

D.2. The king who looked after the prince recently appears in public every week.

الملكة التي اهتمت بالأميرة حديثاً تظهر في العلن كل أسبوع.

D.3. The seller who thanked the customer enthusiastically was happy with the large profit.

البائعة التي شكرت الزبونة بحماسٍ فرحت بالفائدة الكبيرة.

D.4. The novelist who mentioned the servant accurately sells many books to the public.

الكاتبة التي ذكرت الوصيفة بدقةٍ تبيع كتبا كثيرة للجمهور.

D.5. The coach who was very interested in the player worked at the National Fencing Academy.

المدربة التي اهتمت باللاعبة جداً اشتغلت في الأكاديمية الوطنية للمبارزة.

D.6. The obstetrician who cared for the child repeatedly volunteers at the university hospital.

القابلة التي اعتنت بالطفلة تكراراً تتطوع في مستشفى الجامعة.

D.7. The soldier who met the policeperson yesterday loved the atmosphere at the air base.

الجندية التي قابلت الشرطية بالأمسِ أحبّت العمل في القاعدة الجوية.

D.8. The singer who met the dancer previously sings with the city orchestra.

المطربة التي استقبلت الراقصة سابقاً تغني مع أوركسترا المدينة.

D.9. The prompter who cared for the student affectionately traveled to a new country.

الملقّنة التي ربّت الطالبة بحنانٍ سافرت إلى بلدٍ جديد.

D.10. The producer who enthusiastically chose the actor produces a film every month.

المخرجة التي اختارت الممثلة بشغفٍ تنتج فيلما كل شهر.

D.11. The photographer who photographed the magician artistically published the photos in a new book.

المصورة التي صورت الساحرة بتفننٍ نشرت الصور في كتاب جديد.

D.12. The servant who helped the sponsor earnestly cleans the rooms seven days a week.

الخادمة التي ساعدت الكفيلة بجدٍّ تنظف الغرف سبعة أيام في الأسبوع.

D.13. The announcer who invited the historian nicely presented a program on TV.

المذيعة التي دعت المؤرّخة بلطفٍ قدمت برنامجاً في التلفاز.

D.14. The translator who met the accountant repeatedly owns a lot of expensive clothes.

المترجمة التي التقت بالمحاسِبة تكراراً تملك كثيرا من الملابس الغالية.

D.15. The assistant who served the pharmacist loyally collected all of the old reports.

المساعِدة التي خدمت الصيدلانيّة بإخلاصٍ جمعت كل التقارير القديمة.

D.16. The journalist who interviewed the winner by chance writes in many newspapers.

الصحفية التي قابلت الفائزة صدفةٍ تكتب في صحف كثيرة.

D.17. The bedouin who visited the farmer at night goes to the middle of the desert.

البدوية التي زارت المزارعة ليلاً تذهب إلى وسط الصحراء.

D.18. The doctor who treated the boy recently discovered a cure for the terrible disease.

الطبيبة التي عالجت الصبية مؤخراً اكتشفت شفاء للمرض الرهيب.

D.19. The artist who corresponded with the publisher eagerly desired a new contract.

الفنانة التي راسلت الناشرة بشغفٍ رغبت في عَقد جديد.

D.20. The student who admired the poet greatly read many poems last year.

التلميذة التي أُعجبت بالشاعرة بشدةٍ قرأت قصائد كثيرة العام الماضي.

D.21. The director who contacted the author during the day supervises many large projects.

المديرة التي اتصلت بالمؤلفة نهاراً تشرف على كثيرٍ من المشاريع الكبيرة.

D.22. The dean who summoned the professor angrily observed a problem in the university departments.

العميدة التي استدعت الأستاذة بغضبٍ لاحظت خللاً في أقسام الجامعة.

D.23. The musician who accompanied the singer professionally played with the national music group.

الموسيقيّة التي رافقت المغنية بمهنيّةٍ عزفت مع الفرقة الوطنية للموسيقى.

D.24. The ambassador who hosted the delegate yearly spoke at the United Nations.

السفيرة التي استضافت المندوبة سنوياً تحدثت في الأمم المتحدة.

D.25. The grandparent who met the neighbor suddenly talked about the neighborhood issues.

الجدّة التي صادفت الجارة فجأةً تحاورت عن أمور الحي.

D.26. The student who met the manager yesterday got high grades in the remaining subjects.

الطالبة التي قابلت المديرة البارحة نالت درجات عالية في المواد المتبقية.

D.27. The accountant who talked to the employee harshly suffered from social problems.

المحاسبة التي حادثت الموظفة بصرامة عانت من مشاكل اجتماعية.

D.28. The volunteer who thanked the official a lot studied at one of the best international universities.

المتطوعة التي شكرت المسؤولة بكثرة درست باحدى أرقى الجامعات الدولية.

D.29. The graduate who talked to the lecturer happily works for extra hours at the library.

الخريجة التي كلمت المحاضِرة بسعادة تعمل ساعات اضافية في المكتبة.

D.30. The painter who excitedly interviewed the director painted wonderful paintings

الرسامة التي حاورت المخرجة بإثارة رسمت لوحات فنية رائعة.

D.31. The chef who lived next to the trader for a long time practices a cooking career skillfully.

الطاهية التي جاورت التاجرة مطولاً تمارس مهنة الطبخ بمهارة.

D.32. The visitor who talked to the guide in the morning gave a lecture about how to manage time.

الزائرة التي حدثت المرشدة صباحاً القت محاضرة عن كيفية تنظيم الوقت.

D.33. The teacher who visited the doctor yesterday masters speaking in Arabic and English.

المعلمة التي زارت الطبيبة بالأمس تُجيد التحدث باللغة العربية و الإنجليزية.

D.34. The lawyer who accused the guilty person angrily is trying to find the way to the truth.

المحامية التي اتهمت المذنبة بغضبٍ تحاول الوصول إلى الحقيقة.

D.35. The engineer who met the colleague daily aspires to get a job at a prominent company.

المهندسة التي قابلت الزميلة يومياً تسعى للحصول على وظيفة في شركة مرموقة.

D.36. The farmer who blamed the young person yesterday loves working at the farm near the park.

المزارعة التي لامت الشابة بالأمسِ تحبّ العمل في المزرعة المجاورة للحديقة.

D.37. The beginner who helped the boss in the morning was hired for the military company.

المبتدئة التي ساعدت الرئيسة صباحاً توظفت في الشركة العسكرية.

D.38. The actress who met the interviewer in the past resigned from the acting career recently.

الممثلة التي قابلت المذيعة بالماضي استقالت من مهنة التمثيل مؤخراً.

D.39. The guard who talked to the pupil in the morning goes home late every day.

الحارسة التي كلمت التلميذة صباحاً تذهب إلى المنزل في ساعة متاخرة كل يوم.

D.40. The worker who gently scolded the maid cares about helping the needy.

العاملة التي وبخت الخادمة برفق تهتم بمساعدة المحتاجين.

D.41. The analyst who patiently waited for the reporter is trying to educate people about the importance of a clean environment.

المحللة التي انتظرت المراسلة بصبر تسعى لتوعية الشعب على أهمية نظافة البيئة.

D.42. The reporter who spoke to the plaintiff adeptly interviews the president at international conferences.

المراسلة التي سألت المدعية بنباهة تقابل رئيس الدولة في المؤتمرات العالمية.

D.43. The magician who talked to the master/lady quickly worked at the theater near the village.

الساحرة التي كلمت السيدة بسرعة عملت في المسرح المجاور للقرية.

D.44. The employee who accompanied the visitor in the morning stays at work until late.

الموظفة التي رافقت الزائرة بالصباح تبقى حتى ساعة متاخرة في العمل.

D.45. The novelist who pleasantly shook hands with the designer writes international and local novels about literature.

الروائية التي صافحت المصممة بسرور تكتب روايات عالمية و محلية في الادب.

D.46. The researcher who calmly called the detective provides money for charity society

الباحثة التي هاتفت المحققة بهدوء توفر المال للجمعيات الخيرية.

D.47. The coordinator who helped the guard devotedly was in the school courtyard.

المنسقة التي عاونت الحارسة بتفانٍ تواجدت في ساحة المدرسة.

D.48. The candidate who pleasantly thanked the participant took part in the electoral campaign.

المرشحة التي شكرت المشتركة بسرور شاركت في الحملة الانتخابية.

D.49. The judge who decisively questioned the thief ruled fairly among people.

القاضية التي سألت السارقة بحزم حكمت بالعدل بين الناس.

D.50. The immigrant who answered the inspector anxiously faced difficulties at the check point.

المهاجرة التي أجابت المفتشة بقلق واجهت صعوبات عند نقطة التفتيش.

D.51. The pharmacist who helped the grandmother at night works at bakery for sweets.

الصيدلانيّة التي ساعدت الجّدة بالليل تعمل في مخبز الحلويات.

D.52. The tourist who met the driver on the road loves traveling to different countries.

السائحة التي صادفت السائقة بالطريق تحبّ السفر إلى بلدان مختلفة.

D.53. The prince who intelligently answered the journalist owns many huge palaces.

الأميرة التي أجابت الصحافية بذكاء تمتلك عدة قصور كبيرة.

D.54. The client who consulted the lawyer nervously practiced painting for a few years.

الموكلة التي استشارت المحامية بتوتر مارست مهنة الرسم لعدة سنوات.

Appendix B. Supplemental Data & Model Results

This appendix contains tables of grand average raw reading times and mixed-effects model results for regions not immediately germane to the main claims of the paper for experiments 1–4. In all cases, these values are reported identically to the analysis procedures outlined in the *Analysis* section in the main text for each experiment.

Condition	Mean	SD
Subject Region		
MATCH/GRAM	422	126
MATCH/UNGRAM	426	126
NOMATCH/GRAM	420	128
NOMATCH/UNGRAM	427	141
Complementizer Region		
MATCH/GRAM	387	105
MATCH/UNGRAM	389	116
NOMATCH/GRAM	387	101
NOMATCH/UNGRAM	387	110
Relative Clause Verb Region		
MATCH/GRAM	378	107
MATCH/UNGRAM	377	117
NOMATCH/GRAM	382	113
NOMATCH/UNGRAM	382	120
Attractor Region		
MATCH/GRAM	453	195
MATCH/UNGRAM	436	179
NOMATCH/GRAM	452	185
NOMATCH/UNGRAM	456	187
Adverb Region		
MATCH/GRAM	502	205
MATCH/UNGRAM	501	194
NOMATCH/GRAM	514	192
NOMATCH/UNGRAM	509	194
Second Spillover Region		
MATCH/GRAM	408	101
MATCH/UNGRAM	435	116
NOMATCH/GRAM	406	100
NOMATCH/UNGRAM	439	136

Table B.1: Raw condition grand avergage reading times across participant means in milliseconds for additional regions in Experiment 1.

Coefficient	$\hat{\beta}$	SE	t
Relative Clause Verb Region			
Intercept	**331.98**	**14.52**	**22.86**
Match:No	5.82	6.01	0.97
Gram:Ungram	2.66	6.07	0.44
Item Order	**-0.58**	**0.04**	**-14.50**
Length	**6.55**	**2.30**	**2.84**
Previous Region RT	**0.18**	**0.01**	**19.25**
Match:No × Gram:Ungram	-2.35	8.69	-0.27
Attractor Region			
Intercept	**328.06**	**52.26**	**6.28**
Match:No	**-29.02**	**13.76**	**-2.11**
Gram:Ungram	-17.73	11.49	-1.54
Item Order	**-0.98**	**0.08**	**-12.95**
Length	25.18	7.91	3.18
Previous Region RT	**0.17**	**0.02**	**10.78**
Match:No × Gram:Ungram	23.66	16.45	1.44
Adverb Region			
Intercept	**599.34**	**53.64**	**11.17**
Match:No	8.35	11.99	0.70
Gram:Ungram	-1.84	12.11	-0.15
Item Order	**-1.24**	**0.08**	**-15.49**
Length	0.99	7.94	0.12
Previous Region RT	**0.04**	**0.01**	**3.42**
Match:No × Gram:Ungram	2.18	17.34	0.13
Second Spillover Region			
Intercept	**433.27**	**15.51**	**27.93**
Match:No	-2.44	5.75	-0.42
Gram:Ungram	**26.51**	**5.84**	**4.54**
Item Order	**-0.74**	**0.04**	**-19.27**
Length	**5.41**	**1.98**	**2.74**
Previous Region RT	**0.05**	**0.01**	**7.12**
Match:No × Gram:Ungram	5.28	8.32	0.63

Table B.2: Mixed effects regression coefficients for other regions in Experiment 1. Significant coefficients ($|t| > 2$) are in bold and marginal coefficients ($|t| > 1.65$) are in italics.

Condition	Mean	SD
Subject Region		
MASC/MATCH/GRAM	402	131
MASC/MATCH/UNGRAM	411	158
MASC/NOMATCH/GRAM	406	148
MASC/NOMATCH/UNGRAM	415	147
FEM/NOMATCH/UNGRAM	422	162
FEM/NOMATCH/GRAM	428	187
FEM/MATCH/UNGRAM	414	151
FEM/MATCH/GRAM	428	179
Complementizer Region		
MASC/MATCH/GRAM	368	107
MASC/MATCH/UNGRAM	360	103
MASC/NOMATCH/GRAM	365	125
MASC/NOMATCH/UNGRAM	372	129
FEM/NOMATCH/UNGRAM	381	122
FEM/NOMATCH/GRAM	386	127
FEM/MATCH/UNGRAM	375	113
FEM/MATCH/GRAM	381	128
Relative Clause Verb Region		
MASC/MATCH/GRAM	354	119
MASC/MATCH/UNGRAM	356	125
MASC/NOMATCH/GRAM	357	118
MASC/NOMATCH/UNGRAM	363	129
FEM/NOMATCH/UNGRAM	360	122
FEM/NOMATCH/GRAM	364	123
FEM/MATCH/UNGRAM	363	128
FEM/MATCH/GRAM	374	139
Attractor Region		
MASC/MATCH/GRAM	369	141
MASC/MATCH/UNGRAM	390	161
MASC/NOMATCH/GRAM	404	174
MASC/NOMATCH/UNGRAM	410	171
FEM/NOMATCH/UNGRAM	404	153
FEM/NOMATCH/GRAM	401	141
FEM/MATCH/UNGRAM	407	176
FEM/MATCH/GRAM	415	175
Adverb Region		
MASC/MATCH/GRAM	432	176
MASC/MATCH/UNGRAM	434	191
MASC/NOMATCH/GRAM	449	215
MASC/NOMATCH/UNGRAM	458	205
FEM/NOMATCH/UNGRAM	424	160
FEM/NOMATCH/GRAM	426	169
FEM/MATCH/UNGRAM	449	194
FEM/MATCH/GRAM	452	181
Second Spillover Region		
MASC/MATCH/GRAM	365	104
MASC/MATCH/UNGRAM	393	135
MASC/NOMATCH/GRAM	368	105
MASC/NOMATCH/UNGRAM	384	122
FEM/NOMATCH/UNGRAM	383	122
FEM/NOMATCH/GRAM	382	117
FEM/MATCH/UNGRAM	395	141
FEM/MATCH/GRAM	365	103

Table B.3: Raw condition grand avergage reading times across participant means in milliseconds for additional regions in Experiment 2.

Coefficient	$\hat{\beta}$	SE	t
Relative Clause Verb Region			
Intercept	**328.45**	**14.67**	**22.39**
Subj:Fᴇᴍ	11.39	8.05	1.41
Match:Nᴏ	1.81	7.86	0.23
Grammaticality:Uɴɢʀᴀᴍ	1.02	7.85	0.13
Item Order	**-0.52**	**0.04**	**-14.26**
Length	**4.94**	**2.28**	**2.16**
Previous Region RT	**0.15**	**0.01**	**17.66**
Subj:Fᴇᴍ × Match:Nᴏ	-13.32	11.05	-1.21
Subj:Fᴇᴍ × Gram:Uɴɢʀᴀᴍ	-10.29	11.08	-0.93
Match:Nᴏ × Gram:Uɴɢʀᴀᴍ	-1.08	11.10	-0.10
Subj:Fᴇᴍ × Match:Nᴏ × Gram:Uɴɢʀᴀᴍ	10.19	15.64	0.65
Attractor Region			
Intercept	**305.26**	**35.77**	**8.53**
Subj:Fᴇᴍ	**28.15**	**11.44**	**2.46**
Match:Nᴏ	**24.11**	**11.50**	**2.10**
Grammaticality:Uɴɢʀᴀᴍ	16.56	10.20	1.62
Item Order	**-0.78**	**0.05**	**-16.64**
Length	*9.73*	*5.32*	*1.83*
Previous Region RT	**0.23**	**0.01**	**18.04**
Subj:Fᴇᴍ × Match:Nᴏ	-26.58	17.80	-1.49
Subj:Fᴇᴍ × Gram:Uɴɢʀᴀᴍ	-17.19	14.39	-1.19
Match:Nᴏ × Gram:Uɴɢʀᴀᴍ	-18.19	14.42	-1.26
Subj:Fᴇᴍ × Match:Nᴏ × Gram:Uɴɢʀᴀᴍ	23.91	20.32	1.18
Adverb Region			
Intercept	**390.77**	**27.92**	**14.00**
Subj:Fᴇᴍ	10.26	11.90	0.86
Match:Nᴏ	8.04	11.96	0.67
Grammaticality:Uɴɢʀᴀᴍ	-4.38	11.94	-0.37
Item Order	**-1.04**	**0.06**	**-18.78**
Length	**14.55**	**3.51**	**4.15**
Previous Region RT	**0.13**	**0.01**	**11.41**
*Subj:*Fᴇᴍ × *Match:*Nᴏ	*-33.04*	*16.80*	*-1.97*
Subj:Fᴇᴍ × Gram:Uɴɢʀᴀᴍ	6.14	16.85	0.36
Match:Nᴏ × Gram:Uɴɢʀᴀᴍ	13.64	16.88	0.81
Subj:Fᴇᴍ × Match:Nᴏ × Gram:Uɴɢʀᴀᴍ	-13.56	23.79	-0.57
Second Spillover Region			
Intercept	**347.19**	**13.05**	**26.60**
Subj:Fᴇᴍ	-6.94	6.38	-1.09
Match:Nᴏ	0.75	6.41	0.12
Grammaticality:Uɴɢʀᴀᴍ	**16.31**	**6.46**	**2.53**
Item Order	**-0.57**	**0.03**	**-18.84**
Length	3.97	1.60	2.48
Previous Region RT	**0.15**	**0.01**	**16.56**
*Subj:*Fᴇᴍ × *Match:*Nᴏ	*16.60*	*9.01*	*1.84*
Subj:Fᴇᴍ × Gram:Uɴɢʀᴀᴍ	12.74	9.05	1.41
Match:Nᴏ × Gram:Uɴɢʀᴀᴍ	-6.94	9.07	-0.77
Subj:Fᴇᴍ × Match:Nᴏ × Gram:Uɴɢʀᴀᴍ	-17.71	12.76	-1.39

Table B.4: Mixed effects regression coefficients for other regions in Experiment 2. Significant coefficients ($|t| > 2$) are in bold and marginal coefficients ($|t| > 1.65$) are in italics.

Condition	Mean	SD
Subject Region		
MATCH/GRAM	490	187
MATCH/UNGRAM	510	213
NOMATCH/GRAM	494	178
NOMATCH/UNGRAM	490	184
Complementizer Region		
MATCH/GRAM	431	132
MATCH/UNGRAM	445	148
NOMATCH/GRAM	428	123
NOMATCH/UNGRAM	447	155
Relative Clause Verb Region		
MATCH/GRAM	435	159
MATCH/UNGRAM	426	139
NOMATCH/GRAM	426	138
NOMATCH/UNGRAM	432	135
Attractor Region		
MATCH/GRAM	485	211
MATCH/UNGRAM	497	215
NOMATCH/GRAM	503	210
NOMATCH/UNGRAM	497	196
Adverb Region		
MATCH/GRAM	565	270
MATCH/UNGRAM	542	227
NOMATCH/GRAM	552	224
NOMATCH/UNGRAM	546	231
Second Spillover Region		
MATCH/GRAM	428	108
MATCH/UNGRAM	448	119
NOMATCH/GRAM	425	95
NOMATCH/UNGRAM	443	113

Table B.5: Raw condition grand avergage reading times across participant means in milliseconds for additional regions in Experiment 3.

Coefficient	$\hat{\beta}$	SE	t
Relative Clause Verb Region			
Intercept	**435.70**	**18.42**	**23.66**
Ambiguity:AMBIG	0.94	11.73	0.08
Match:NO	-5.74	11.48	-0.50
Grammaticality:UNGRAM	-14.26	11.44	-1.25
Item Order	**-1.01**	**0.05**	**-18.31**
Length	**11.50**	**2.91**	**3.95**
Previous Region RT	**0.10**	**0.01**	**13.31**
Amb:UNAM × Match:NO	-4.77	16.60	-0.29
Amb:UNAM × Gram:UNGRAM	7.02	16.61	0.42
Match:NO × Gram:UNGRAM	2.26	16.33	0.14
Amb:UNAM × Match:NO × Gram:UNGRAM	17.43	23.65	0.74
Attractor Region			
Intercept	**579.38**	**66.19**	**8.75**
Ambiguity:AMBIG	-5.92	18.53	-0.32
Match:NO	22.78	16.87	1.35
Grammaticality:UNGRAM	4.07	15.91	0.26
Item Order	**-1.40**	**0.08**	**-18.16**
Length	-8.42	11.04	-0.76
Previous Region RT	**0.20**	**0.01**	**14.45**
Amb:UNAM × Match:NO	9.56	23.99	0.40
Amb:UNAM × Gram:UNGRAM	19.55	23.09	0.85
Match:NO × Gram:UNGRAM	-8.68	22.71	-0.38
Amb:UNAM × Match:NO × Gram:UNGRAM	-24.95	32.88	-0.76
Adverb Region			
Intercept	**564.51**	**45.75**	**12.34**
Ambiguity:AMBIG	-11.72	20.19	-0.58
Match:NO	**-37.33**	**18.17**	**-2.05**
Grammaticality:UNGRAM	-19.24	18.11	-1.06
Item Order	**-1.72**	**0.09**	**-19.42**
Length	**16.93**	**6.15**	**2.75**
Previous Region RT	**0.13**	**0.01**	**9.76**
Amb:UNAM × Match:NO	34.98	26.29	1.33
Amb:UNAM × Gram:UNGRAM	-14.30	26.30	-0.54
Match:NO × Gram:UNGRAM	24.38	25.85	0.94
Amb:UNAM × Match:NO × Gram:UNGRAM	-6.25	37.43	-0.17
Second Spillover Region			
Intercept	**467.28**	**14.29**	**32.71**
Ambiguity:AMBIG	8.94	9.94	0.90
Match:NO	-2.25	7.45	-0.30
Grammaticality:UNGRAM	**20.02**	**7.43**	**2.69**
Item Order	**-1.02**	**0.04**	**-27.72**
Length	2.28	1.80	1.27
Previous Region RT	**0.10**	**0.01**	**11.72**
Amb:UNAM × Match:NO	-7.49	10.77	-0.70
Amb:UNAM × Gram:UNGRAM	-12.52	10.77	-1.16
Match:NO × Gram:UNGRAM	-9.17	10.59	-0.87
Amb:UNAM × Match:NO × Gram:UNGRAM	15.68	15.34	1.02

Table B.6: Mixed effects regression coefficients for other regions in Experiment 3. Significant coefficients ($|t| > 2$) are in bold and marginal coefficients ($|t| > 1.65$) are in italics.

Condition	Mean	SD
Subject Region		
Sg/Match/Gram	422	146
Sg/Match/Ungram	449	175
Sg/NoMatch/Gram	452	172
Sg/NoMatch/Ungram	434	159
Pl/NoMatch/Ungram	476	212
Pl/NoMatch/Gram	455	180
Pl/Match/Ungram	455	181
Pl/Match/Gram	457	217
Complementizer Region		
Sg/Match/Gram	371	102
Sg/Match/Ungram	387	106
Sg/NoMatch/Gram	387	113
Sg/NoMatch/Ungram	399	129
Pl/NoMatch/Ungram	397	127
Pl/NoMatch/Gram	393	126
Pl/Match/Ungram	406	129
Pl/Match/Gram	420	142
Relative Clause Verb Region		
Sg/Match/Gram	374	115
Sg/Match/Ungram	370	118
Sg/NoMatch/Gram	364	117
Sg/NoMatch/Ungram	366	116
Pl/NoMatch/Ungram	389	135
Pl/NoMatch/Gram	390	129
Pl/Match/Ungram	391	135
Pl/Match/Gram	404	149
Attractor Region		
Sg/Match/Gram	400	153
Sg/Match/Ungram	423	178
Sg/NoMatch/Gram	420	176
Sg/NoMatch/Ungram	414	167
Pl/NoMatch/Ungram	419	152
Pl/NoMatch/Gram	414	161
Pl/Match/Ungram	430	155
Pl/Match/Gram	450	187
Adverb Region		
Sg/Match/Gram	428	150
Sg/Match/Ungram	452	186
Sg/NoMatch/Gram	456	188
Sg/NoMatch/Ungram	449	185
Pl/NoMatch/Ungram	442	153
Pl/NoMatch/Gram	451	175
Pl/Match/Ungram	458	177
Pl/Match/Gram	454	165
Second Spillover Region		
Sg/Match/Gram	362	94
Sg/Match/Ungram	376	95
Sg/NoMatch/Gram	361	96
Sg/NoMatch/Ungram	380	101
Pl/NoMatch/Ungram	370	99
Pl/NoMatch/Gram	363	92
Pl/Match/Ungram	367	96
Pl/Match/Gram	373	98

Table B.7: Raw condition avergage grand reading times across participant means in milliseconds for additional regions in Experiment 4.

Coefficient	$\hat{\beta}$	SE	t
Relative Clause Verb Region			
Intercept	**379.16**	**17.33**	**21.88**
SubjNum: Pl	**24.33**	**8.20**	**2.97**
Match:No	-9.12	8.17	-1.12
Grammaticality:Ungram	-6.22	8.19	-0.76
Item Order	**-0.61**	**0.04**	**-16.76**
Length	*4.81*	*2.58*	*1.86*
Previous Region RT	**0.08**	**0.01**	**12.70**
SubjNum:Pl × Match:No	2.33	11.59	0.20
SubjNum:Pl × Grammaticality:Ungram	-0.58	11.60	-0.05
Match:No × Grammaticality:Ungram	5.80	11.56	0.50
SubjNum:Pl × Match:No × Grammaticality:Ungram	-2.94	16.40	-0.18
Attractor Region			
Intercept	**333.03**	**33.49**	**9.95**
SubjNum: Pl	**32.43**	**11.31**	**2.87**
Match:No	13.42	11.28	1.19
Grammaticality: Ungram	**20.01**	**10.54**	**1.90**
Item Order	**-0.85**	**0.05**	**-17.83**
Length	**9.47**	**4.05**	**2.34**
Previous Region RT	**0.21**	**0.01**	**17.49**
SubjNum: Pl × **Match:**No	**-35.49**	**16.98**	**-2.09**
*SubjNum:*Pl *× Grammaticality:*Ungram	*-28.17*	*14.94*	*-1.89*
*Match:*No *× Grammaticality:*Ungram	*-26.99*	*14.88*	*-1.81*
*SubjNum:*Pl *× Match:*No *× Grammaticality:*Ungram	*40.35*	*21.12*	*1.91*
Adverb Region			
Intercept	**489.73**	**28.15**	**17.40**
*SubjNum:*Pl	*20.51*	*11.07*	*1.85*
Match: No	**29.06**	**11.02**	**2.64**
*Grammaticality:*Ungram	*18.31*	*11.05*	*1.66*
Item Order	**-0.81**	**0.05**	**-16.14**
Length	-5.08	3.99	-1.27
Previous Region RT	**0.11**	**0.01**	**10.69**
SubjNum:Pl × Match:No	-24.19	15.64	-1.55
SubjNum:Pl × Grammaticality:Ungram	-4.18	15.66	-0.27
Match:No × Grammaticality:Ungram	-24.07	15.59	-1.54
SubjNum:Pl × Match:No × Grammaticality:Ungram	-0.18	22.13	-0.01
Second Spillover Region			
Intercept	**303.85**	**11.14**	**27.27**
SubjNum:Pl	6.46	5.52	1.17
Match:No	-0.32	5.50	-0.06
Grammaticality:Ungram	4.41	5.53	0.80
Item Order	**-0.44**	**0.03**	**-17.48**
Length	**5.11**	**1.42**	**3.61**
Previous Region RT	**0.20**	**0.01**	**23.49**
SubjNum:Pl × Match:No	-6.26	7.80	-0.80
SubjNum:Pl × Grammaticality:Ungram	-4.31	7.82	-0.55
Match:No × Grammaticality:Ungram	7.90	7.78	1.01
SubjNum:Pl × Match:No × Grammaticality:Ungram	-3.44	11.04	-0.31

Table B.8: Mixed effects regression coefficients for other regions in Experiment 4. Significant coefficients ($|t| > 2$) are in bold and marginal coefficients ($|t| > 1.65$) are in italics.

References

Antón-Méndez, I., Nicol, J. L., & Garrett, M. F. (2002). The relation between gender and number agreement processing. *Syntax*, *5*, 1–25.

Badecker, W., & Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language*, *56*, 65–85.

Badecker, W., & Lewis, R. (2007). A new theory and computational model of working memory in sentence production: Agreement errors as failues of cue-based retrieval. Paper Presented at the 20th annual CUNY Sentence Processing Conference. San Diego, La Jolla.

Badenes-Ribera, L., Frías-Navarro, D., Monterde-i Bort, H., & Pascual-Soler, M. (2015). Interpretation of the p value: A national survey study in academic psychologists from spain. *Psicothema*, *27*, 290.

Badenes-Ribera, L., Frias-Navarro, D., Iotti, B., Bonilla-Campos, A., & Longobardi, C. (2016). Misconceptions of the p-value among chilean and italian academic psychologists. *Frontiers in Psychology*, *7*.

Baker, M. C. (2008). *The Syntax of Agreement and Concord*. Cambridge, UK: Cambridge University Press.

Béjar, S. (2003). *Phi-Syntax: A Theory of Agreement*. Ph.D. thesis University of Toronto Toronto.

Béjar, S., & Rezac, M. (2009). Cyclic Agree. *Linguistic Inquiry*, *40*, 35–73.

Bock, K., & Eberhard, K. M. (1993). Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes*, *8*, 57–99.

Bock, K., & Miller, C. (1991). Broken agreement. *Cognitive Psychology*, *23*, 45–93.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 365–376.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. (2000). Minimalist inquiries. In D. M. Martin, Roger, & J. Uriagereka (Eds.), *Step by Step: Essays on Minimalist Sytnax in Honor of Howard Lasnik* chapter 3. (pp. 89–156). Cambridge, MA: MIT Press.

Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: A Life in Language* (pp. 1–52). Cambridge, Massachusetts: MIT Press.

Collaboration, O. S. et al. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, aac4716.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.

Cumming, G. (2008). Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science*, *3*, 286–300.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*, 7–29.

Cumming, G., & Maillardet, R. (2006). Confidence intervals and replication: where will the next mean fall? *Psychological methods*, *11*, 217.

Dank, M., & Deutsch, A. (2010). The role of morpho-phonological factors in subject–predicate gender agreement in Hebrew. *Language and Cognitive Processes*, *25*, 1380–1410.

Deutsch, A., & Dank, M. (2009). Conflicting cues and competition between notional and grammatical factors in producing number and gender agreement: Evidence from Hebrew. *Journal of Memory and Language*, *60*, 112–143.

Deutsch, A., & Dank, M. (2011). Symmetric and asymmetric patterns of attraction errors in producing subject–predicate agreement in Hebrew: An issue of morphological structure. *Language and Cognitive Processes*, *26*, 24–46.

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*, 85–103.

Eberhard, K., Cutting, J., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review*, *112*, 531–559.

Eberhard, K. M. (1997). The marked effect of number on subject–verb agreement. *Journal of Memory and Language*, *36*, 147–164.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American statistical Association*, *82*, 171–185.

Engelmann, F., Jäger, L. A., & Vasishth, S. (2015). The determinants of retrieval interference in dependency resolution: Review and computational modeling. Ms. (submitted), Universität Potsdam.

Francis, G. (2012). Publication bias and the failure of replication in experimental psychology. *Psychonomic Bulletin & Review*, *19*, 975–991.

Franck, J. (2011). Reaching agreement as a core syntactic process. *Natural Language & Linguistic Theory*, *29*, 1071–1086.

Franck, J., Colonna, S., & Rizzi, L. (2015). Task-dependency and structure-dependency in number interference effects in sentence comprehension. *Frontiers in Psychology*, *6*, 349.

Franck, J., Vigliocco, G., Antón-Méndez, I., Collina, S., & Frauenfelder, U. (2008). The interplay of syntax and form in sentence production: A cross-linguistic study of form effects on agreement. *Language and Cognitive Processes*, *23*, 329–374.

Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes*, *17*, 371–404.

Fuchs, Z., Polinsky, M., & Scontras, G. (2015). The differential representation of number and gender in Spanish. *The Linguistic Review*, *32*, 703–737.

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, *33*, 587–606.

Gillespie, M., & Pearlmutter, N. J. (2013). Against structural constraints in subject–verb agreement production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 515.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, *31*, 337.

Haller, H., & Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers. *Methods of Psychological Research*, *7*, 1–20.

Hartsuiker, R. J., Schriefers, H., Bock, K., & Kikstra, G. (2003). Morphophonological influences on the construction of subject-verb agreement. *Memory & Cognition*, *31*, 1316.

Häussler, J. (2009). *The Emergence of Agreement Attraction Errors During Sentence Comprehension*. Ph.D. thesis Universität Konstanz Konstanz, Germany.

Häussler, J., & Bader, M. (2009). Agreement checking and number attraction in sentence comprehension: Insights from German relative clauses. *Travaux du Cercle Linguistique de Prague*, *7*.

Hedges, L. V., & Olkin, I. (1980). Vote-counting methods in research synthesis. *Psychological Bulletin*, *88*, 359.

Hoekstra, R., Finch, S., Kiers, H. A., & Johnson, A. (2006). Probability as certainty: Dichotomous thinking and the misuse of p values. *Psychonomic Bulletin & Review*, *13*, 1033–1037.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. (2nd ed.). Sage publications.

Jäger, L., Chen, Z., Li, Q., Lin, C.-J. C., & Vasishth, S. (2015). The subject-relative advantage in chinese: Evidence for expectation-based processing. *Journal of Memory and Language*, *79*, 97–120.

Jäger, L. A., Engelmann, F., & Vasishth, S. (In Press). Similarity-based interference in sentence comprehension: Literature review and bayesian meta-analysis. *Journal of Memory and Language*, .

Just, M., Carpenter, P., & Woolley, J. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*, 228–238.

Kirby, K. N., & Gerlanc, D. (2013). Bootes: An r package for bootstrap confidence intervals on effect sizes. *Behavior research methods*, *45*, 905–927.

Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PLOS ONE*, *9*, 1–8.

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement in spanish comprehension. *Journal of Memory and Language*, *82*, 133–49.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*, 375–419.

Lorimor, H., Bock, K., Zalkind, E., Sheyman, A., & Beard, R. (2008). Agreement and attraction in Russian. *Language and Cognitive Processes*, *23*, 769–799.

Mahowald, K., James, A., Futrell, R., & Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *Journal of Memory and Language*, *91*, 5–27.

Malko, A., & Slioussar, N. (To Appear). Gender features in processing: Evidence from Russian. *Frontiers in Psychology: Encoding and Navigating Linguistic Representations in Memory*, *6*.

Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological methods*, *5*, 241.

Nicol, J. L., Forster, K. I., & Veres, C. (1997). Subject–verb agreement processes in comprehension. *Journal of Memory and Language*, *36*, 569–87.

Norris, M. (2014). *A Theory of Norminal Concord*. Ph.D. thesis University of California, Santa Cruz.

Oaks, M. (1986). Statistical inference: A commentary for the social and behavioral

2264   sciences.

2265   Pearlmutter, N. J. (2000). Linear versus hierarchical agreement feature processing in
2266   comprehension. *Journal of Psycholinguistic Research*, *29*, 89–98.

2267   Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence
2268   comprehension. *Journal of Memory and Language*, *41*, 427–456.

2269   Pollock, J.-Y. (1989). Verb movement, universal grammar, and the structure of IP.
2270   *Linguistic Inquiry*, *20*, 365–424.

2271   Potvin, P. J., & Schutz, R. W. (2000). Statistical power for the two-factor repeated
2272   measures anova. *Behavior Research Methods, Instruments, & Computers*, *32*, 347–
2273   356.

2274   Preminger, O. (2011). *Agreement as a Fallible Operation*. Ph.D. thesis MIT Cambridge,
2275   MA.

2276   R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R
2277   Foundation for Statistical Computing Vienna, Austria.

2278   Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological
2279   Bulletin*, *114*, 510–532.

2280   Rhode, D. (2003). *Linger: A Flexible Platform for Language Processing Experiments*.
2281   MIT.

2282   Rosenthal, R., & Dimatteo, M. (2001). Meta-analysis. *Annual Review of Psychology*,
2283   *52*, 59–82.

2284   Ryding, K. C. (2005). *A Reference Grammar of Modern Standard Arabic*. Cambridge,
2285   UK: Cambridge UP.

2286   Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in
2287   psychology: Implications for training of researchers. *Psychological Methods*, *1*,
2288   115–129.

2289   Sicuro Corrêa, L. M., Almeida, D. A. d. A., & Porto, R. S. (2004). On the representation
2290   of Portuguese gender-inflected words in the mental lexicon. *Brain and Language*,
2291   *90*, 63–73.

2292   Sprouse, J., & Almeida, D. (In Press). Design sensitivity and statistical power in ac-
2293   ceptability judgment experiments. *Glossa: a journal of general linguistics*, .

2294   Stanley, D. J., & Spence, J. R. (2014). Expectations for replications are yours realistic?
2295   *Perspectives on Psychological Science*, *9*, 305–318.

2296   Staub, A. (2009). On the interpretation of the number attraction effect: Response time
2297   evidence. *Journal of Memory and Language*, *60*, 308–27.

2298   Staub, A. (2010). Response time distributional evidence for distinct varieties of number
2299   attraction. *Cognition*, *114*, 447–54.

2300   Tanner, D. (2011). *Agreement Mechanisms in Native and Nonnative Language Process-
2301   ing: Electrophysiological Correlates of Complexity and Interference*. Ph.D. thesis
2302   University of Washington Seattle, WA.

2303   Tanner, D., Nicol, J. L., & Brehm, L. (2014). The time-course of feature interference in
2304   agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Jour-
2305   nal of Memory and Language*, *76*, 195–215.

2306   Trubetskoy, N. (1939/1958). *Grundzüge der Phonologie*. Göttingen: Vandenhoeck &
2307   Ruprecht.

2308   Tucker, M. A. (2011). The morphosyntax of the Arabic verb. In N. LaCara, A. Thomp-
2309   son, & M. A. Tucker (Eds.), *Morphology at Santa Cruz: Papers in Honor of Jorge*

Hankamer (pp. 177–211). Santa Cruz, CA: LRC Publications.

Tucker, M. A., Idrissi, A., & Almeida, D. (2015). Representing number in the real-time processing of agreement: Self-paced reading evidence from Arabic. *Frontiers in Psychology*, *6*.

Tucker, M. A., & Wagers, M. W. (2010). Agreement attraction in infinitival complements. In G. McGuire (Ed.), *UCSC LRC Laboratory Report 2010* (pp. 23–43). Santa Cruz, CA: LRC Publications.

Vasishth, S., & Nicenboim, B. (2016). Statistical methods for linguistic research: Foundational ideas–part I. *Language and Linguistics Compass*, *10*, 349–369.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48.

Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language*, *34*, 186–215.

Vigliocco, G., & Franck, J. (1999). When sex and syntax go hand in hand: Gender agreement in language production. *Journal of Memory and Language*, *40*, 455–478.

Vigliocco, G., & Franck, J. (2001). When sex affects syntax: Contextual influences in sentence production. *Journal of Memory and Language*, *45*, 368–390.

Vigliocco, G., & Nicol, J. (1998). Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? *Cognition*, *68*, B13–B29.

Vigliocco, G., & Zilli, T. (1999). Syntactic accuracy in sentence production: The case of gender disagreement in Italian language-impaired and unimpaired speakers. *Journal of Psycholinguistic Research*, *28*, 623–648.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*, 206–237.

Wehr, H. J. C. e. (1976). *A Dictionary of Modern Written Arabic*. Spoken Language Services Ithaca, NY.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American psychologist*, *54*, 594.