Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak

William Badecker *, Frantisek Kuminiak

Department of Cognitive Science, Johns Hopkins University, Baltimore, MD 21218-2685, USA

Abstract

The experimental studies presented in this paper exploit agreement attraction in order to examine the mechanisms underlying the production of subject–verb agreement in Slovak. Our experiments verify that the processes which specify the gender feature on past tense verbs are subject to interference from local nouns, and that the likelihood of interference from the local noun depends on the relative markedness of the subject phrase's head and local noun gender. The results suggest that, unlike the asymmetries observed with the two-term opposition between singular and plural, the markedness asymmetries observed with the three Slovak genders are not explicable solely in terms of the under-specification (or default status) of the unmarked value. The results also indicate that the likelihood of attraction depends on the case form of the head and local noun. The effects relating to gender markedness and to grammatical case provide evidence that producing subject–verb agreement requires the retrieval of an agreement source from content-addressable working memory.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Working memory; Content-based retrieval; Agreement; Case ambiguity

Introduction

Most theories of sentence production recognize distinct conceptual, syntactic and phonological levels of encoding (Bock, 1982; Bock & Levelt, 1994; Dell, 1986; Garrett, 1980, 1982; Levelt, 1989), but both the complexity of the process and the potential for interaction among levels make it difficult to sort out how planning is carried out at each level. Planning agreement is a good case in point. At a first approximation, agreement is constructed at a syntactic level of planning. Yet subject–predicate agreement in English and other languages

is also responsive to the conceptual/referential number and gender of a subject phrase even when these diverge from morphological values (Pollard & Sag, 1994; Wechsler & Zlatić, 2003). In the domain of language processing, the probability of producing plural agreement with a morphologically singular subject noun phrase can be modulated by the conceptual number of the referent denoted by that phrase (Eberhard, 1999; Vigliocco, Butterworth, & Semenza, 1995, 1996b), and the accuracy of gender agreement for speakers of languages like French and Italian is also affected by the connection between the morphological gender of a phrasal head and the 'natural gender' of the phrase's referent (Vigliocco & Franck, 1999). Several recent studies of agreement have used the differences between syntactic and conceptual number to gain leverage on

the distinction between syntactic and pre-syntactic planning stages (Bock, Nicol, & Cutting, 1999, 2001; Bock et al., 2004). In this paper we will explore the role that morpho-syntactic markedness and morpho-syntactic ambiguity play in producing gender agreement in Slovak to further isolate details of syntactic planning stages.

The hierarchic structure of language often creates contexts that require non-contiguous constituents to agree. This poses an interesting challenge to the processing system that can be freely observed in the agreement errors that speakers (and writers) produce. Both naturally occurring agreement errors (Francis, 1986; Strang, 1966; Zandvoort, 1961) and experimentally elicited errors (e.g, Bock & Cutting, 1992; Bock & Miller, 1991) reveal that these performance lapses are predominantly instances of "attraction" to an erroneous agreement source: A target verb appears to enter into an agreement relation with another noun or noun phrase in the local context whose feature specification differs from that of the actual subject. For example, in the error "*the response from all the actors **were** terrific,*" the local noun "*actors*" (or the local noun phrase "*all the actors*") is said to attract agreement away from the singular head of the subject "*response*". Attraction errors are observed in cases of pronominal agreement with antecedents (Bock, Eberhard, Cutting, Meyer, & Schriefers, 2001; Meyer & Bock, 1999), as well as in cases of subject predicate agreement.

Attraction errors have other properties that inform us about the types of representations and processes that are engaged in sentence formulation. For example, attraction errors are influenced by the specific inflectional values of the agreement source. Planning is more vulnerable to interference from a local noun that is plural when the head of the subject noun phrase is singular (e.g., *the key to the cabinets* . . .), than to interference from a singular local noun when the head is plural (e.g., *the keys to the cabinet* . . .; Bock & Cutting, 1992; Bock & Eberhard, 1993; Bock & Miller, 1991; Eberhard, 1997). This attraction asymmetry for number agreement has been replicated in studies across a wide range of languages, including Spanish (Vigliocco, Butterworth, & Garrett, 1996a), Italian (Vigliocco et al., 1995), German and Dutch (Hartsuiker, Schriefers, Bock, & Kikstra, 2003), French (Fayol, Largy, & Lemaire, 1994) and Slovene (Harrison, 2004); and it is not unreasonable to suggest that the asymmetry reflects a difference between the grammatically unmarked status of singular and the marked status of plural. In comparison to these results for number, though, studies of gender agreement have not documented a corresponding markedness asymmetry (Vigliocco & Franck, 1999).

By and large, the ability of a plural local noun phrase to attract agreement depends on the morpho-syntactic character of the local noun phrase and its head. Pseudo-plurals like *rose* are no more likely to condition plural agreement as local nouns than are singular forms like *row* (whereas true plurals like *rows* are effective attractors; Bock & Eberhard, 1993); a local noun is no more able to attract agreement if it is the plural form of a collective noun like *teams* than the plural form of a non-collective noun like *players*, nor is a morphologically singular collective noun like *team* able to attract agreement (Bock et al., 2001); and the manner in which the morpho-syntactic feature for number is expressed on the local noun has little or no effect on the likelihood of error: irregular plurals like *men* are just as effective in attracting agreement as are regular plurals like *boys* (Bock & Eberhard, 1993; though see Haskell & MacDonald, 2003). One noteworthy exception to the generalization that number attraction is influenced by the morpho-syntactic character of the local noun phrase alone is the effect of paradigm status of a plural form. For example, invariant plurals like *suds* attract agreement less than plurals like *bubbles* that have competing singular forms (Bock et al., 2001).

Virtually all accounts of agreement production posit a mechanism that either copies the features of a source on to a target, or checks the features of one term against another when grammar demands an agreement relation. Accounts differ, though, on whether attraction arises because the wrong phrase is checked (e.g., Meyer & Bock, 1999), or whether checking is relatively impervious to error and local noun phrases influence the calculation by the effects they have on the feature specification of the phrase that is the target of the checking process (e.g., Eberhard, Cutting, & Bock, 2005).

Attraction and markedness

Models of agreement planning offer a varied set of explanations for the singular/plural asymmetry in agreement production and attraction. In some instances, the notion of morpho-syntactic markedness plays no direct role in the account of either. For example, in their account of agreement planning based on constraint satisfaction mechanisms, Haskell and MacDonald (2003) and Thornton and MacDonald (2003) attribute the singular/plural asymmetry to the influence of syntactically similar patterns such as the pseudo-partitive "*a flock of birds*" (Selkirk, 1977). According to this account, the frequency of the distributional pattern combining verb forms with this noun phrase structure modulates the strength of the local noun's number feature as a processing cue; and it does so asymmetrically in that a plural noun in this [*singular noun – [preposition–plural noun]*] structure can license plural agreement, whereas there is no parallel [*plural noun–[preposition–singular noun]*] structure that regularly conditions singular agreement.

Markedness based accounts take a very different approach to the uneven effect of number. For example, according to the "privative feature" hypothesis, the

representation of a plural noun includes an explicit morpho-syntactic feature [plural] as a constituent, but there is no corresponding feature [singular] marking singular nouns or noun phrases (Bock & Eberhard, 1993; Eberhard, 1997; Franck, Vigliocco, & Nicol, 2002; Vigliocco et al., 1995, 1996a, Vigliocco, Hartsuiker, Jarema, & Kolk, 1996b). Subject–verb agreement, on this account, starts with a number feature that originates on the lexical head of a noun phrase and transmits it to the phrasal root node of the subject noun phrase. Once an inflectional feature percolates to this syntactic vantage point, the feature can then be linked to (or copied onto) the verb. Errors arise when a feature that marks the subject noun phrase node moves to that syntactic position from the wrong lexical source (Franck et al., 2002; Vigliocco & Nicol, 1998; Vigliocco et al., 1995, 1996a, 1996b). The markedness asymmetry in attraction is a direct consequence of the privative status of [plural]. There is no feature [singular] available to percolate (by hypothesis, singular agreement is default agreement), therefore there is no feature [singular] that can percolate beyond the limits of the local noun phrase to cause attraction.

A more recent proposal dubbed the ''Marking and Morphing Model'' encodes morpho-syntactic number using a continuously valued Singular-and-Plural feature (Bock, 2004; Eberhard et al., 2005). On this proposal, noun phrases begin with two semi-independent sources of number information: A Singular-and-Plural value that is part of the morpho-syntactic specification of the immediate lexical constituents (most notably the head noun, but also specifiers like ''one'' or ''these''), and the Singular-and-Plural value that the phrase itself receives from ''number marking'' mechanisms (to register the conceptual number of an intended referent). The ultimate Singular-and-Plural value for the noun phrase is calculated in working memory by a spreading activation process that sums the Singular-and-Plural values of the lexical constituents (weighted to reflect the strength of the link that binds each lexical item to the noun phrase node) with the initial Singular-and-Plural value of the noun phrase (Eberhard et al., 2005). The singular/plural asymmetry follows on this account from the algebraic nature of spreading activation and from some assumptions about the Singular-and-Plural feature's activation values that encode singular (default) and plural number. Plural noun phrases bear a positive value for the Singular-and-Plural feature, whereas singular nouns and noun phrases have activation values for this feature that range from negative values to zero. An activation value at or near zero coming from a singular local noun phrase will not diminish the non-zero value that a plural head noun contributes to the calculation when activations from the two sources are summed. However, a plural local noun phrase will have some positive Singular-and-Plural value to transmit to the root node of the subject noun phrase, so there is some chance that adding

this activation value to the (typically zero) activation value of a singular lexical head will push the summed activation level for this feature to the threshold for inducing plural agreement.

Although the Marking and Morphing Model's spreading activation account of agreement attraction differs substantially from the percolation account in many respects, in other respects it bears some important similarities to percolation. As in the privative feature plus percolation account, producing subject–verb agreement requires transmitting the number specification of the subject noun phrase to the verb, and as in that account, attraction errors arise when the number specification (here, the Singular-and-Plural value) of a local noun phrase spreads from a local noun phrase to the root of the subject noun phrase. In addition, the two theories explain the singular/plural attraction asymmetry in terms of the absence of a representational property for singular noun phrases (i.e., no number features on the percolation account; an activation level of zero for the number feature on the spreading activation account) that contrasts with the presence of some property that distinguishes plural noun phrases (an overt plural feature on the percolation story; a positive, non-zero value for a Singular-and-Plural feature on the activation based account). This raises the question of what types of attraction asymmetries these accounts predict when an agreement system differentiates more than two values. For example, if there are multiple grammatically significant values (as in three-way gender systems), then the feature percolation theory would predict that no asymmetry should be observed for the two non-zero (or marked) values, since they must both be explicitly marked. A similar prediction is made if, by analogy to Eberhard et al.'s (2005) treatment of number, gender is encoded in the settings for a single, continuously valued feature (though this is not the only conceivable encoding solution). For example, a single gender-linked feature could distinguish the three grammatical genders by associating each with a distinct range of activation values. Suppose, for the sake of argument, that Feminine and Masculine are respectively associated with negative and positive activation values, and that Neuter is associated with the null value. The encoding solution would also predict that asymmetries should exist when Masculine and Feminine are contrasted with Neuter, but that there should be no such asymmetry between Masculine and Feminine (the two genders associated with opposing, non-zero values).

There are, however, theories of markedness which make different types of predictions about processing asymmetries. For example, in current theories of phonological, morpho-syntactic and syntactic structure based on Optimality Theory (OT; Bresnan, 2000, 2001; McCarthy, 2002; Prince & Smolensky, 2004), the notion of marked versus unmarked is defined in a purely

relational manner. Optimality Theory is based on the idea that the structure of every language can be captured by a fixed set of violable constraints that are ranked with respect to one another. A potential linguistic form that violates a given constraint will always lose out to other potential forms that only violate lower-ranked constraints. For example, markedness relations between a language's three genders, masculine, feminine and neuter, are expressed by ranking three markedness constraints *Masculine, *Feminine and *Neuter, which respectively prohibit the expression of each of these genders. The ranking "*Feminine » *Masculine » *Neuter" entails that feminine is more marked than masculine (i.e., that expressing feminine is disfavored in comparison to expressing masculine) because feminine violates a more highly ranked markedness constraint than masculine does. Crucially, though, there is no notion of absolute markedness. Neuter is not intrinsically unmarked, it is only unmarked in comparison to another specific gender; nor does the ranking entail that the markedness asymmetry between Feminine and Neuter is greater than the markedness asymmetry between Feminine and Masculine. For any pair of candidate gender forms, the gender value corresponding to the higher ranked markedness constraint is grammatically dis-preferred in comparison to the others. To explain why relatively marked elements occur in languages at all, OT-based theories posit faithfulness constraints. These constraints require a linguistic structure to be faithful, on some relevant dimension, to the input being expressed (e.g., one faithfulness constraint might demand that the grammatical gender of a noun phrase be faithful to the gender of its lexical head). Faithfulness constraints are violable as well; and they are ranked among markedness constraints in a single partial order. In this regard, constraint ranking relativizes markedness constraints not just with respect to other markedness constraints, but with respect to the full range of constraints as well.

To account for the singular/plural performance asymmetry, for example, one might propose that the markedness bias follows from ranked grammatical faithfulness constraints and how these constraints would apply under conditions of partial information about structure. When there is sufficient information about the local syntactic structure, agreement mechanisms should only have to deal with one possible agreement controller (hence, one number value). Even the most marked feature value in the input will be chosen over alternative, less marked values, as long as a faithfulness constraint for that feature value (e.g., FaithAgr(Plural): express a plural feature in the input to agreement) is ranked above the corresponding markedness constraint (e.g., *Plural). However, some forms of uncertainty (partial information) about the local structure may give rise to inputs that include more than one potential controller (in particular, one singular and one plural candidate). If a constraint requiring faithfulness to the relatively marked plural features in the input, FaithAgr(Plural), outranks a constraint that requires faithfulness to the relatively unmarked singular features in the input, then the asymmetry will follow directly from grammatical constraint ranking and how the constraints must apply to the hypothesized partial specification of structure.

A markedness bias of this character predicts that asymmetries may arise whenever two candidates both appear to be viable agreement controllers. For example, the faithfulness ranking FaithAgr(Feminine) » FaithAgr(Masculine) » FaithAgr(Neuter) predicts an asymmetry for any gender pairing. However, if the hypothesized bias is truly a selection bias, then it should only come into play in competition situations (i.e., where two noun phrases compete with one another directly for source status). On the constraint-based account there ought to be no difference between Masculine and Feminine *with respect to their competition with Neuter.* One goal of the studies presented here is to clarify the relation between grammatically motivated markedness asymmetries and the processing asymmetries observed in the production of agreement.

Working memory retrieval and producing agreement

The general framework we propose for modeling the planning process relies on subject retrieval from a content addressable working memory system to calculate subject–verb agreement. For expository purposes, we will refer to it as the *Working Memory Retrieval Model.* Following Lewis and Vasishth (2005), we assume that the elementary units in working memory are morphosyntactically tagged lexical representations, and that phrases are maintained in this work space as hierarchically bundled arrangements of these elementary units. Each lexical entity retained in working memory is assumed to include information regarding its combinatorial properties—as in computationally oriented linguistic frameworks such as Head Driven Phrase Structure Grammar (Pollard & Sag, 1994) and Tree-Adjoining Grammar (Frank, 2002), and related processing models (e.g., Lewis & Vasishht, 2005; Ferreira, 2000, respectively). The basic premise of this working memory model is the uncontroversial claim that syntactic planning involves dynamically binding lexical identities to various functional and structural roles (such as those implicated in the creation of functional and positional level representations in Garrett (1980, 1982) and Bock & Levelt (1994)). Incremental sentence production entails that some lexical items will be bound to these grammatical roles before others are (Clark & Wasow, 1998; Smith & Wheeldon, 2001), and that in working out the plan for later portions of a sentence, the syntactic formulator and associated processing mechanisms

may need to consult binding relations that specific words or phrases enter into with respect to their grammatical role or syntactic position. Retrieving information regarding details of earlier portions of the overall sentence plan implicates working memory in general. Evidence from attentional interference effects supports the view that agreement production requires access to working memory resources (Fayol et al., 1994; Hartsuiker & Barkhuysen, 2006).

The role of working memory retrieval in agreement planning can be understood in much the same way that it is in working-memory based accounts of parsing (Gordon, Hendrick, & Johnson, 2001; Lewis & Vasishth, 2005). For example, if the inflected form of a verb **B** in a sentence plan depends on the morpho-syntactic features of noun phrase **A** that occurred earlier in that plan, then **A** must be inspected; and in order to inspect **A** for these features, **A** must be retrieved from working memory. In particular, it must be isolated from other elements—some of which may have morpho-syntactic features of the relevant sort—in the current working memory representation. We assume, for example, that the lexical subject of a sentence will be bound to representational elements that encode subjecthood: nominative case, occupying a specifier position in the extended projection of the verb phrase, occurring pre-verbally, etc. (Woolford, 2006; Zaenan & Maling, 1983). When the lexical subject must be recovered from working memory, these properties can be exploited as retrieval cues. Under the best conditions, only one part of the working memory representation of the preceding plan—the actual subject—will resonate to these retrieval cues.[1]

However, the coarseness of retrieval cues, the ambiguity of certain lexical representations, and the fact that representations in working memory decay over time can cause problems for memory retrieval operations (Gordon et al., 2001; Lewis & Vasishth, 2005), and may result either in the retrieval of multiple candidate controllers, or in no candidate retrieval at all. For example, isolating the subject may be complicated by the fact that other elements of the local plan are encoded in a manner that allows them to resonate to the retrieval cues for subjects. If a local noun resonates to case- or position-based subject retrieval cues, then this gives rise to conditions under which mis-selection may occur. The more subject-like a local noun is in terms of case marking, linear or structural position, etc., the more susceptible a cue-based retrieval mechanism will be to wrongly accessing that local noun: the working memory model attributes agreement attraction to failures of the cue-based retrieval process in sentence working memory.

How this might work can be seen in the context of retrieval mechanisms based on grammatical case. In many case-inflecting languages, the distinction between nominative case and accusative or other non-nominative cases may be preserved for some categories, but not for others. For example, German has distinct nominative and accusative determiners for masculine singular noun phrases, but not for different number or gender values; Slovak distinguishes nominative and accusative case endings on nouns in some declensions, but not in others. Some evidence indicates that case ambiguity can influence performance on agreement (Hartsuiker et al., 2003; Schriefers & van Kampen, 1993). For example, Hartsuiker et al. (2003, experiment 2) found that a plural local noun phrase in German was able to attract number agreement from a (Feminine) head noun if the local noun phrase was accusative (as in *Die Stellungnahme gegen die Demonstrationen* "the position against the demonstrations"), but not if it was dative (as in *Die Stellungnahme zu den Demonstrationen* "the position on the demonstrations"). The working memory retrieval model would account for this difference by noting that the determiner in the plural, accusative local noun phrase "*die Demonstrationen*" is case-ambiguous: *die* is the definite determiner for both nominative and accusative plural noun phrases. Hence, a case-based retrieval cue used to recover the (nominative) subject may erroneously reactivate this plural noun phrase (either instead of, or in addition to, the true subject). Because *den* is not a nominative determiner for either plural or singular nouns, the working-memory representation for the dative (unambiguously non-nominative) noun phrase "*den Demonstrationen*" should not be accessed by this case-based retrieval cue for subjects, and therefore should not induce attraction errors.

As Hartsuiker et al. (2003) note, their data do not differentiate accounts of case-ambiguity effects based directly on morpho-syntactic features from accounts based on morpho-phonological ambiguity. For example, others have proposed that morphological ambiguities can influence agreement processing not because the planning system starts out with a lexical representation that is intrinsically ambiguous with respect to case, but because feed-back from phonological forms to morpho-syntactic values can activate unintended representational elements that can interfere with the number specification of the subject phrase (Vigliocco et al., 1995). In their paper, Hartsuiker et al. suggest that feed-back from phonological representations to earlier

---

[1] With this resource, agreement may be calculated "on demand" as part of the incremental planning process: Cue-based lexical retrieval provides a mechanism for inflecting verbs and/or predicate adjectives with the flexibility to identify the properties of an agreement controller not just when the structure for that controller is first formulated, but even later when the properties of an agreement target must be worked out as part of a later processing increment (and independently of whether the agreement target was anticipated at the time that the agreement source was originally formulated).

(pre-phonological) processes may result in the activation of "both correct and incorrect feature codes" (p. 1325), which in turn is what gives rise to production of agreement errors.

Such a feed-back account of case-ambiguity effects would appear to depend on the phonological properties of the case-ambiguous form being highly correlated with the morpho-syntactic features that interfere with identifying agreement features for the subject phrase. This correspondence was present in Hartsuiker et al.'s German experiments: The case-ambiguous determiner for the local noun was *die* in all number mismatch conditions. In addition, the phonological forms in error-inducing conditions shared another property that may have added to the confusion of correct and incorrect feature codes: Agreement attraction was observed almost entirely with subject noun phrases that were headed by feminine nouns (i.e., only when the head and local noun phrases were **both** marked with the case- and number-ambiguous determiner *die*). To what extent does the case-ambiguity effect depend on this degree of correspondence between surface phonological properties of head and local noun determiners and the morpho-syntactic features that they putatively activate via feed-back? Unlike the feed-back account of case ambiguity effects, which can rely on a strong correspondence between specific phonological patterns and specific morpho-syntactic features, the working memory retrieval model attributes the case ambiguity effect directly to the different morpho-syntactic specification for case-ambiguous and case-unambiguous lexical representations. The retrieval account predicts case effects even if a phonological property that sets an unambiguously nominative form apart from its accusative counterpart (e.g., a specific inflectional suffix, or the absence of a suffix) is also a phonological property that can be seen in other nouns that are themselves case-ambiguous. Our experiments with gender agreement in Slovak should also allow us to evaluate these differing expectations regarding the source of case-ambiguity effects. A brief summary of the grammatical principles relevant to our study of gender agreement between subjects and past tense verbs in Slovak, the phenomenon which we will exploit to gauge contributions of markedness and case-ambiguity, is provided in Appendix A to this paper.

Experiment 1

Studies of gender agreement have observed effects for the biological gender of the head noun when that gender differs from the morphological gender of the head (Vigliocco & Franck, 1999; Vigliocco & Franck, 2001), but these effects had to do with competition between the conceptual gender of the referent and the morphological gender of the phrasal head. In general,

the gender of a local noun should have no bearing on the semantic gender assigned to the phrase as a whole. Accordingly, gender is well suited to study effects of morpho-syntactic markedness with little potential interference from semantically based encoding effects. The goal of Experiment 1 was to establish whether the sentence completion task will elicit gender agreement errors from Slovak speakers at all, and if so, whether there are asymmetries in the error patterns that reflect the grammatical markedness relationships between Slovak genders. It examined two specific hypotheses regarding markedness effects. Grammatical considerations lead to treating Neuter as the least marked gender in Slovak, and Feminine as the most marked gender (see discussion in Appendix A). Based on these markedness distinctions, both the privative feature approach to markedness and the ranked constraint approach predict an asymmetry between Neuter and Masculine. However, only the latter approach predicts that one will see a performance asymmetry between Masculine and Feminine. The first hypothesis to be tested in the experiment is that we will observe performance asymmetries that reflect the grammatical markedness asymmetry between Masculine and Neuter gender. If this hypothesis is correct, one should observe more errors with *Neuter–Masculine* subject phrases (i.e., subject noun phrases that contain neuter head nouns and masculine local nouns) than with *Masculine–Neuter* subject phrases. The second hypothesis to be tested is that Feminine and Masculine gender will exhibit different rates of attraction (for inanimate head and local nouns) that likewise reflects the markedness difference between Feminine and Masculine. If this hypothesis is correct, then one should observe more agreement attraction errors with *Masculine–Feminine* Subject phrases than with *Feminine–Masculine* subject phrases.

Method

Participants

Participants in this and subsequent experiments were recruited from the undergraduate and graduate community of the Mlynská Dolina university housing complex, in Bratislava, Slovakia. Thirty-two native Slovak speakers aged 18–35 (average 23) with at least 12 years of formal education participated in this experiment in exchange for a modest financial reward.

Materials

The stimulus items consisted of noun phrases with prepositional phrase modifiers. Four gender sets of 12 items each were constructed. Each item had a gender-match and a gender-mismatch variant. In the mismatch condition, the head and local nouns in the four gender sets were combined as follows:

Head-Local noun mismatch sets:

Neuter–Masculine
Masculine–Neuter

Masculine–Feminine
Feminine–Masculine

In the match condition, the head nouns remained the same, and the local nouns were exchanged for nouns of the same gender as the head. The head and the local noun were singular in all gender sets. Examples of experimental items are in Table 1.

In addition to the experimental preambles, 90 filler preambles were constructed. These included coordinate noun phrases, as well as noun phrases with adjectival modifiers, with prepositional phrase modifiers and with relative clause modifiers.

From these materials, two lists were constructed. Each list contained a total of 138 preambles: 48 experimental and 90 fillers. Out of each set of 12 experimental preambles, six of them occurred in the match condition on the first list and in the mismatch condition on the second list; and vice versa for the remaining six preambles. The items were arranged in pseudo-random order, with the constraint that no more than two experimental items occur consecutively. Each list was broken up into four approximately equal blocks, each of which began with five fillers. The order of presentation of these blocks was varied, so that material presented in the first half of one presentation order corresponded to material in the second half of the other order.

Because Slovak verbs are specified for gender only in the past tense singular, we presented speakers with temporal expressions (such as *today, yesterday*) on index cards and asked them to produce sentences referring to the time period indicated on the card. All experimental items were paired with expressions referring to the past (*yesterday, last year, the other day*), thus requiring a verb in the past tense. In total, expressions referring to the past were used 61 times; expressions referring to the present 28 times; expressions referring to the future 35 times; and general temporal expressions 14 times (these counts include filler items).

Due to concerns about possibly low error rates, we employed a secondary task to increase the likelihood of agreement errors. Specifically, each trial began with the presentation of an arbitrary three-letter 'acronym', which participants were to recall after the sentence generation phase of the trial. Fayol et al. (1994) found that the use of an attentionally demanding secondary task increased the proportion of agreement errors in a sentence generation task. In many respects the inclusion of the secondary task mirrors the effect of the attentional demands that occur in the context of normal extemporaneous speech. Furthermore, Fayol et al. observed the same markedness-related asymmetries in their experiments as others had seen when no secondary task was employed, so it appears that the introduction of such a secondary task does not substantially alter the nature of the sentence production component of the experiment.

The stimuli were tape-recorded, read by a female speaker, slightly faster than the normal speaking rate. A practice list was also prepared, consisting of 17 items, with accompanying cards and acronyms.

Procedure

Participants were run individually. They were told that they would hear a taped "fake acronym" and then the beginning of a sentence. At the same time they would be presented with a card indicating some period of time. The participants' task was to produce a complete sentence, using the phrase as the sentence's subject. The sentence should refer to the time period indicated on the card. After finishing the sentence, they were required to say aloud the acronym that they heard before the phrase.

Table 1
Examples of the gender materials from Experiment 1

Masculine–Neuter		
Match (*Masculine–Masculine*)	Pohár na čaj	The glass for tea
Mismatch (*Masculine–Neuter*)	Pohár na mlieko	The glass for milk
Neuter–Masculine		
Match (*Neuter–Neuter*)	Okno na pole	The window to the field
Mismatch (*Neuter–Masculine*)	Okno na dvor	The window to the yard
Feminine–Masculine		
Match (*Feminine–Feminine*)	Baňa na med'	The mine for copper
Mismatch (*Feminine–Masculine*)	Baňa na nikel	The mine for nickel
Masculine–Feminine		
Match (*Masculine–Masculine*)	Spor o klenot	The quarrel over the jewel
Mismatch (*Masculine–Feminine*)	Spor o korist'	The quarrel over the loot

Participants were asked to speak as quickly as possible, while trying to observe the constraints above. As for the secondary task, they were instructed to make a fast guess at the acronym if they were unable to recall it with certainty.

Since Slovak allows pre-verbal objects, there was a possibility that, for a preamble like *lístok na vlak* 'the ticket for the train,' participants might use the case-ambiguous preambles as topicalized direct objects, rather than as subjects, to produce sentences like *Lístok na vlak som kúpila včera* 'The ticket for the train, I bought yesterday'. Therefore, the instructions explicitly requested that the preambles should be used as subjects. Since the distinction between subjects and topicalized objects is not necessarily obvious to naïve speakers, examples of "correct" and "incorrect" responses were given to illustrate the difference.

Acronyms and preambles were presented through the built-in speaker on a portable tape recorder. The experimenter [FK] paused the tape after each preamble, and simultaneously turned over the index card with the temporal expression associated with the preamble. At this moment the participants were expected to repeat the preamble along with a completion. If the participant failed to understand a preamble, the experimenter repeated the preamble aloud. Seven practice trials were devoted to familiarization with the task and discussion of the constraints on the responses. On the remaining practice trials, the participants practiced responding rapidly without violating the response constraints. After the practice trials, one of the two experimental lists was presented, in four blocks, with short breaks between the blocks. Throughout the experiment, the experimenter prompted participants to in increase their performance rate whenever their responses slowed down appreciably, and to adhere to the basic task requirements regarding sentence tense, the role of the preamble in sentence responses, and accuracy on the secondary memory task when required by decrements in performance (e.g., if the participant responded with the same acronym in three successive trials). Participants were also asked to provide more variable responses if they used the same completion several times in succession, as such stock responses may not fully engage the normal sentence production process.

The experimental sessions, excluding the practice trials, were audio-taped for later analysis, using a second portable recorder and a clip-on condenser microphone.

Scoring

The participants' responses were fully transcribed and each of them was assigned into one of the following categories:

Correct. Complete, grammatical sentences that began with a correct reproduction of the preamble, used it as the sentence's subject, had correct agreement on the verb, and did not qualify as "no gender agreement," below.

Agreement error. Responses that met all the requirements for a "correct" except that they were ungrammatical because the verb did not agree with the head. Agreement errors where the verb agreed with the local noun – attraction errors – were counted separately from other agreement errors, not attributable to attraction.

Topicalization error. Responses that used the preamble as a topicalized object ('The ticket for the train, I bought yesterday'), instead of a subject. Note that when the head is case-ambiguous, such sentences are grammatical (although contrary to the instructions given to participants). Participants occasionally used even unambiguously nominative heads as topicalized objects; the resulting responses were ungrammatical (e.g., they used two unambiguously nominative phrases with a transitive verb).

A variety of other response types were scored together under the category of *Miscellaneous error.* These included:

Other grammatical errors. Responses that are ungrammatical for other reasons than a subject–verb agreement violation or an incorrectly used topicalized-object construction. A typical error of this type would be omitting a verb's obligatory argument.

No gender agreement. Responses to gender-agreement items that were correct, except that the verb was not specified for gender. (This means that these responses ignored the index card that asked for a past-tense response.)

Repetition errors. All instances in which the participants did not repeat the preamble back verbatim; also, cases when the acronym intended for memorization was repeated back by the participant before the preamble.

Completion error. Cases where the participant did not produce a complete sentence, or the completion was unintelligible.

In cases of self-corrections and multiple completions, scoring was based on the first verb uttered.

Results

For the gender-attraction stimuli (sets *Masculine–Feminine, Feminine–Masculine, Masculine–Neuter* and *Neuter–Masculine*), the application of the scoring criteria yielded 1,241 correct responses (80.8% of all responses), 52 agreement errors (3.4%), 165 (10.7%) topicalized objects, and 78 (5.1%) miscellaneous responses. Of the 52 agreement errors, 47 (90%) were errors of attraction.

The numbers of responses obtained in the individual scoring categories are shown in Table 2, broken down by stimulus sets and the match/mismatch condition (non-attraction errors appear in parentheses). Responses to the sentence completion task were examined using

Table 2
Number of sentence completions in the different scoring categories for each stimulus type in the gender materials in Experiment 1

Noun phrase type (head-local noun)	Correct	Agreement error	Topicalization error	Miscellaneous error
Masculine–Masculine	164	0	21	7
Masculine–Neuter	155	5	19	13
Neuter–Neuter	158	0	27	7
Neuter–Masculine	141	20 (3)	28	3
Feminine–Feminine	172	2 (2)	5	13
Feminine–Masculine	164	7	8	13
Masculine–Masculine	144	0	34	14
Masculine–Feminine	143	18	23	8

The number of non-attraction agreement errors (i.e., where the verb form does not conform to the gender of the local noun) is enclosed in parentheses.

repeated measures analysis of variance with participants as random factors and items as random and nested factors as appropriate. The proportion of agreement errors by condition was analyzed using correct and agreement error responses. In reporting analyses for this and subsequent experiments, we provide $\min F'$ statistics (Clark, 1973) when these are at or near significance, and by-subject (F_1) and by-item (F_2) analyses when the $\min F'$ value does not reach significance. For direct means comparisons we also report 95% confidence intervals (CIs) around differences between means computed on by-subject proportion means (Masson & Loftus, 2003).

Analysis of the combined data for *Masculine–Feminine*, *Feminine–Masculine*, *Masculine–Neuter* and *Neuter–Masculine* Gender Set conditions confirmed that agreement errors were more likely to occur in the gender mismatch than in gender match conditions (mean proportion was less than 0.01 in match conditions vs. 0.077 in mismatch conditions: CI $= \pm 0.016$ for a means difference > 0.067; $\min F'(1,73) = 15.6$, $p < .001$). Analysis also revealed a main effect for Gender Set ($F_1(3,93) = 4.02$, $p < .01$; $F_2(3,44) = 3.03$, $p < .05$ [$\min F'(3,107) = 1.72$]) and a reliable interaction between Set and Match conditions ($F_1(3,93) = 5.94$, $p < .001$; $F_2(3,44) = 4.06$, $p < .05$ [$\min F'(3,102) = 2.40$]). Planned comparisons were also carried out to further probe the effects of Gender Set. For the *Masculine–Feminine* and *Feminine–Masculine* materials, analysis revealed a main effect for Match, with more agreement errors in gender mismatch conditions (0.076) than in the match conditions (less than 0.01): CI $= \pm 0.016$ for a means difference > 0.066; $\min F'(1,49) = 5.9$, $p < .05$. There was no reliable effect for Gender Set ($F_1(1,31) = 2.94$, $p > .09$; $F_2(1,22) = 1.55$, $p > .20$); although the interaction between Match and Gender Set was marginally significant ($F_1(1,31) = 7.16$, $p < .05$; $F_2(1,22) = 3.48$, $p < .08$). Analysis of the *Masculine–Neuter* and *Neuter–Masculine* materials revealed reliable effects for Match (mismatch = 0.078, match = 0; CI $= \pm 0.016$ for a means difference of 0.078;

$\min F'(1,50) = 14.0$, $p < .001$) and Gender Set (*Masculine–Neuter* = 0.015, *Neuter–Masculine* = 0.06; CI $= \pm 0.016$ for a means difference of 0.045; $\min F'(1,53) = 3.93$, $p = .05$), and a significant interaction between Match and Gender Set conditions ($\min F'(1,53) = 3.93$, $p = .05$). Planned comparisons for each Gender Set revealed that differences in the number of gender agreement errors in match vs. mismatch conditions were not significant in the *Feminine–Masculine* or *Masculine–Neuter* Gender Sets (all Fs < 1). However, the proportion of agreement errors on the gender-mismatched items in the *Masculine–Feminine* set (mean = 0.11) differed reliably from the proportion of errors (0) in the gender match condition (CI $= \pm .03$ for a means difference of 0.11; $\min F'(1,31) = 5.57$, $p < .05$); as did the errors for gender-mismatched items in the *Neuter–Masculine* set (mean = 0.12 in mismatch vs. 0 errors in the match condition; CI $= \pm 0.035$ for a means difference of 0.12; $\min F'(1,34) = 12.9$, $p < .001$).

Analyses were also carried out on the rates of Topicalization errors in the experimental conditions. For these analyses, the rate of topicalization was computed based on the total number of Correct, Agreement Error, and Topicalized responses. A main effect for Gender Set was observed ($\min F'(3,66) = 2.90$, $p < .05$), with fewer such errors being produced for items in the *Feminine–Masculine* Gender Set (mean proportion -.036) than in any of the other Gender Set conditions (all proportions greater than 0.10; CI $= \pm 0.04$ for means difference > 0.064; $\min F'(3,66)$ 2.90, $p < 0.05$). There was no effect for Match (Fs < 1), nor was the interaction between Match and Gender Set reliable ($p > .30$).

Discussion

As anticipated, the experiment elicited gender agreement errors, and the analysis of these errors confirms that they are predominantly attraction errors: virtually all gender agreement errors occurred in the mismatch conditions (50/52), and nearly all of these (47/50) were

instances where the verb agreed in gender with the local noun. Notably, though, the proportion of agreement attraction errors was not the same across all Gender Sets. The higher rate of attraction errors in the *Neuter–Masculine* materials than in the *Masculine–Neuter* materials provides one piece of evidence that the relative markedness of the head and local noun genders contributes to the pattern of agreement errors—as does the trend toward a comparable asymmetry between the *Masculine–Feminine* and *Feminine–Masculine* materials. There are some important caveats with respect to the contrast between *Masculine–Feminine* and *Feminine–Masculine* preambles employed in this experiment. Due to the rarity of case-ambiguous feminine nouns in Slovak, some of the 12 nouns used as phrasal heads in this set were unambiguously nominative (whereas the corresponding head nouns in all of the other Gender Sets were ambiguous with respect to the nominative-accusative distinction; and all local nouns were case-ambiguous as well). Although previous reports of case ambiguity effects suggest that it is the case-ambiguity of the local noun, and not the head noun of a subject noun phrase, that is relevant, this aspect of the *Feminine–Masculine* set may well have had an influence on the pattern of performance in the experiment. The rate of Topicalization responses and agreement errors upholds this possibility, in that these response types only occurred after noun phrases containing case-ambiguous heads. So, one possibility is that the apparent difference in agreement attraction between the *Masculine–Feminine* and *Feminine–Masculine* materials actually represents an effect of unambiguously nominative phrasal heads in the *Feminine–Masculine* preambles. To address this possibility, such case-ambiguity confounds were explicitly avoided in subsequent experiments. In particular, Experiment 2 provides a direct and unambiguous contrast between

masculine and feminine local nouns regarding their ability to attract agreement.

Experiment 2

In order to clarify the source of the asymmetry in attraction rates between *Masculine–Feminine* and *Feminine–Masculine* conditions observed in Experiment 1, we constructed materials in which the gender and case ambiguity of the head and local noun were independently crossed. Case ambiguity was manipulated for both head and local noun using Masculine and Feminine nouns from different declensional paradigms. Because there was no confound with case ambiguity of the head noun for *Neuter–Masculine* and *Masculine–Neuter* items in the previous experiment, critical contrasts for evaluating effects for markedness were limited to *Masculine–Feminine* and *Feminine–Masculine* items in this experiment. To evaluate the effect of case ambiguity on gender agreement we varied whether the nouns that appear in nominative vs. accusative positions are case-unambiguous, or case-ambiguous. Examples of the stimuli we employed are provided in Table 3.

Method

Participants

Graduate and undergraduate participants were recruited from the Mlynská Dolina university housing complex, in Bratislava, Slovakia. Forty-five native Slovak speakers, ranging in age from 19 to 29 years (average 24) with a minimum of 12 years education, participated in this experiment in exchange for a modest financial reward. None of the participants in this experiment took part in Experiment 1.

Table 3
Examples of the experimental materials from Experiment 2

Head noun	Local noun condition	Preamble	
MU	Match	*Sluha*$_{M,nom}$ *pre palác*$_{M,nom \lor acc}$	The servant for the palace
	Unambiguous mismatch	*Sluha*$_{M,nom}$ *pre hostinu*$_{F,acc}$	The servant for the feast
	Ambiguous mismatch	*Sluha*$_{M,nom}$ *pre domácnosť*$_{F,nom \lor acc}$	The servant for the household
MA	Match	*Trest*$_{M,nom \lor acc}$ *za zločin*$_{M,nom \lor acc}$	The punishment for the crime
	Unambiguous mismatch	*Trest*$_{M,nom \lor acc}$ *za vraždu*$_{F,acc}$	The punishment for the murder
	Ambiguous mismatch	*Trest*$_{M,nom \lor acc}$ *za krádež*$_{F,nom \lor acc}$	The punishment for the theft
FU	Match	*Odmena*$_{F,nom}$ *pre družinu*$_{F,acc}$	The reward for the group
	Unambiguous mismatch	*Odmena*$_{F,nom}$ *pre výhercu*$_{M,acc}$	The reward for the winner
	Ambiguous mismatch	*Odmena*$_{F,nom}$ *pre projekt*$_{M,nom \lor acc}$	The reward for the project
FA	Match	*Reťaz*$_{F,nom \lor acc}$ *na mašinu*$_{F,acc}$	The chain for the machine
	Unambiguous mismatch	*Reťaz*$_{F,nom \lor acc}$ *na medveďa*$_{M,acc}$	The chain for the bear
	Ambiguous mismatch	*Reťaz*$_{F,nom \lor acc}$ *na bicykel*$_{M,nom \lor acc}$	The chain for the bicycle

(Abbreviations: M = masculine, F = Feminine, A = case ambiguous, U = case unambiguous.).

Materials

Seventy-two experimental items were constructed, each consisting of a noun phrase with a single noun head and a prepositional phrase modifier. All heads as well as local nouns in the experimental items were singular. The items were divided into four sets of 18 items each, depending on the properties of the head noun:

Masculine-Unambiguous: Masculine head with Unambiguous nominative case marking.
Masculine-Ambiguous: Masculine head with Ambiguous nominative/accusative case marking.
Feminine-Unambiguous: Feminine head with Unambiguous nominative case marking.
Feminine-Ambiguous: Feminine head with Ambiguous nominative/accusative case marking.

Each item in each set had three conditions depending on the properties of the local noun the head was paired up with. Two of these conditions were gender-mismatch conditions, in which the local noun was feminine when the head was masculine, and vice versa. In the "unambiguous mismatch" condition, the local noun was unambiguously accusative; in the "ambiguous mismatch" condition, the local noun was case-ambiguous between nominative and accusative. The third condition, "match", was a control condition, where the gender of the local noun was identical to that of the head.

Two facts ought to be mentioned about the stimuli; we will examine in the discussion section whether either one of these could possibly be a confounding factor. First, Slovak requires all masculine case-unambiguous nouns to be animate, and most of these nouns in our stimuli denoted humans. All other nouns used in both mismatch conditions denoted non-humans. Second, to keep the lists reasonably short, the case-ambiguity of the local noun was *not* independently manipulated in the "match" condition. Case-ambiguous feminine nouns are relatively scarce, as are semantically plausible constructions of the type "the *animate_1* for the *animate_2*", which would be required for a "case-unambiguous match" condition in the Masculine-Unambiguous set. So, as a concession to these distributional facts about feminine and masculine forms, the local noun was unambiguously accusative in the "match" condition for feminine head nouns and case-ambiguous in the "match" condition for masculine head nouns. Regardless of case-ambiguity, low error rates were expected after gender-matched nouns, since no true attraction errors can occur.

In addition to the experimental preambles, 52 filler preambles were used. These included coordinate noun phrases, as well as noun phrases with adjectival or prepositional phrase modifiers. The nouns in filler stimuli with prepositional phrase modifiers were gender-matched (except for one filler item). The nouns in coordinate noun phrases were always plural, but sometimes mismatched in gender. In sum, there were 8 masculine singular, 9 feminine singular, 11 neuter singular, and 24 plural filler noun phrases.

These materials were used to construct three lists. Each list consisted of 124 preambles: 72 experimental and 52 fillers. Each item was preceded by a three-letter "nonsense acronym," as in Experiment 1. Items from each condition in each of the four sets were distributed between these lists in a balanced fashion. The order of items was pseudo-random, with the constraint that no more than three experimental items could occur consecutively, and no two adjacent experimental items were from the same set. Each list was broken up into four approximately equal parts. Each of these parts began with three fillers, and the last item in each part was also a filler. The order of presentation of the four parts was kept the same.

Procedure and scoring

The experimental procedure and scoring were the same as in Experiment 1.

Results

Application of the scoring criteria yielded 2798 correct responses (86.4% of all responses), 77 agreement errors (2.4%), 155 (4.8%) topicalized objects, and 210 (6.5%) miscellaneous responses. Of the 77 agreement errors, 72 (94%) were attraction errors. The distribution of these response types over the four preamble conditions is summarized in Table 4.

Analyses of variance were performed on the data, with Gender Match, Head Gender and Head Ambiguity as independent variables and the number of agreement errors, and number of topicalized object responses as dependent variables. Note that the independent variable Match has three values: "match," "case-unambiguous mismatch" and "case-ambiguous mismatch"—where case-ambiguity relates to the form of the local noun. Separate ANOVAs were conducted with participants and with items as random factors (F_1 and F_2, respectively). To perform item analyses across stimulus sets, the variable Item was nested within the variables Head Gender and Head Ambiguity.

We observed a strong main effect for Gender Match ($\min F'(2,148) = 16.7$, $p < .001$), for Head Gender ($\min F'(1,112) = 5.06$, $p < .05$), and for Head Ambiguity ($\min F'(1,112) = 20.1$, $p < .001$). Each two-way interaction also reached significance: Gender Match by Head Gender ($\min F'(2,135) = 4.28$, $p < .05$); Gender Match by Head Ambiguity ($\min F'(2,145) = 15.6$, $p < .001$); and Head Gender by Head Ambiguity ($\min F'(1,111) = 4.38$, $p < .05$). The three-way interaction between Gender Match, Head Gender, and Head Ambiguity was also reliable ($\min F'(2,133) = 3.32$, $p < .05$). Sepa-

Table 4
Number of sentence completions in the different scoring categories for preambles Experiment 2

Head noun	Local noun condition	Correct	Agreement error	Topicalization error	Miscellaneous error
MU	Match	253	0	3	14
	Unambiguous mismatch	260	0	0	10
	Ambiguous mismatch	251	2 (1)	2	15
MA	Match	218	1 (1)	26	25
	Unambiguous mismatch	224	5	27	14
	Ambiguous mismatch	175	49 (1)	22	24
FU	Match	253	0	1	16
	Unambiguous mismatch	251	0	0	19
	Ambiguous mismatch	248	0	1	21
FA	Match	229	1 (1)	28	12
	Unambiguous mismatch	222	1 (1)	25	22
	Ambiguous mismatch	214	18	20	18

The number of non-attraction agreement errors (i.e., where the verb form does not conform to the gender of the local noun) is enclosed in parentheses. (Abreviations: M = masculine, F = Feminine, A = case ambiguous, U = case unambiguous.).

rate comparisons for the Masculine and Feminine Headed preambles affirmed the effects for Gender Match (Masculine: $\min F'(2,87) = 15.0$, $p < .001$; Feminine: $\min F'(2,34) = 5.44$, $p < .01$), for Head Ambiguity (Masculine: $\min F'(1,71) = 16.9$, $p < .001$; Feminine: $\min F'(1,74) = 8.01$, $p < .01$), and their interaction (Masculine: $\min F'(2,87) = 13.4$, $p < .001$; Feminine: $\min F'(2,120) = 5.44$, $p < .01$). That is, the overall pattern wherein errors showed up only in the Gender Mismatch condition when both head and local noun were case-ambiguous held for both Masculine- and Feminine-Headed preambles.

Comparisons were also carried out to evaluate pairwise differences between the three Gender Match conditions: Match vs. Mismatch with Case-Unambiguous local noun vs. Mismatch with Case-Ambiguous local noun. The comparison of Match vs. Mismatch with Case-Unambiguous local noun resulted in no reliable effects or interactions. When Gender Match items were contrasted with the Gender Mismatch items with case-ambiguous local nouns, though, analyses revealed main effects for Gender Match (more agreement errors on mismatch items than on gender match items: $\min F'(1,110) = 19.1$, $p < .001$), for Head Gender (more agreement errors on *Masculine–Feminine* than on *Feminine–Masculine* items: $\min F'(1,110) = 4.63$, $p < .05$), and for Head Ambiguity (more errors on items with case-ambiguous heads: $\min F'(1,112) = 18.4$, $p < .001$). All three two-way interactions reached significance (Gender Match by Head Gender: $\min F'(1,112) = 4.96$, $p < .05$; Head Match by Head Ambiguity: $\min F'(1,111) = 17.6$, $p < .001$; and Head Gender by Head Ambiguity: $F_1(1,44) = 11.7$, $p < .01$; $F_2(1,68) = 5.44$, $p < .05$ [$\min F'(1,109) = 3.71$, $p = .06$]), as did the three-way interaction between the factors ($\min F'(1,112) = 3.97$, $p < .05$).

Similar comparisons were carried out between the two Gender Mismatch conditions: Case-Ambiguous vs. Case-Unambiguous local noun. These comparisons revealed comparable effects: main effects for case-ambiguity of the local noun (more agreement errors on the case-ambiguous than on the case unambiguous mismatch items: $\min F'(1,112) = 16.4$, $p < .001$), for Head Gender (more agreement errors on *Masculine–Feminine* than on *Feminine–Masculine* items: $\min F'(1,112) = 5.55$, $p < .05$), and for head ambiguity (more errors on items with case-ambiguous heads: $\min F'(1,112) = 20.5$, $p < .001$). All three two-way interactions reached significance: Local Case ambiguity by Head Gender ($F_1(1,44) = 11.2$, $p < .01$; $F_2(1,68) = 5.77$, $p < .05$; [$\min F'(1,111) = 3.77$]); Local case ambiguity by Head Ambiguity ($\min F'(1,111) = 3.80$, $p = .05$); and Head Gender by Head Ambiguity ($\min F'(1,111) = 4.61$, $p < .05$). Likewise, the three-way interaction between the factors was reliable ($F_1(1,44) = 8.54$, $p < .01$; $F_2(1,68) = 4.40$, $p < .05$ [$\min F'(1,111) = 2.9$]).

Overall, these analyses of agreement errors reveal several strong main effects and interactions that were largely carried by the high rate of errors when head and local noun disagreed in gender and both the head and local noun were case-ambiguous. Agreement errors were virtually absent in all other conditions. Focusing, then, on the conditions in which the head noun was case ambiguous, the proportion of agreement errors for the Masculine headed noun phrases in the case-ambiguous mismatch condition (mean = 0.18) differed reliably from the match and the case unambiguous mismatch conditions (both means <0.02; CI = ±0.03 for means differences >0.16); and the proportion of agreement errors in the Feminine headed noun phrases in the case unambiguous mismatch condition (mean = 0.07) like-

wise differed reliably from the match and case unambiguous mismatch conditions (both means < 0.01; CI = ±0.026 for means differences > 0.06).

Analysis of Topicalization errors also affirmed that there was a highly reliable effect for the Case-Ambiguity of the head noun: The proportion of topicalization in conditions with case-ambiguous heads (0.098) was greater than in conditions with case-unambiguous heads (0.005; CI = ±0.05 for a means difference of 0.093; min $F'(1,106)$ = 22.6, p < .001). No other main effects or interactions were observed (Fs < 1).

Discussion

There are two major findings of the present experiment. First, the asymmetry between *Masculine–Feminine* and *Feminine–Masculine* attraction conditions was observed to be reliable: The proportion of attraction errors involving *Masculine–Feminine* preambles was reliably larger than the corresponding proportion for *Feminine–Masculine* conditions. In both gender sets these errors arose predominantly in preamble conditions where the head of the subject phrase was case-ambiguous: 21.5% of scorable responses [=correct + agreement error] for the *Masculine–Feminine* condition, and 7.8% of scorable responses for the *Feminine–Masculine* condition. This finding supports the view that Feminine behaves as though it is more marked than Masculine gender. These results are consistent with the bias account of markedness asymmetries, though they would appear to serve as counter-evidence to the account of these asymmetries provided by the privative feature account of attraction asymmetries, and at least some versions of Eberhard et al.'s (2005) activation level proposal.

The second and equally important finding that speaks to the different models of agreement processing is that the gender attraction errors in gender mismatch conditions occurred only when both the head and the local noun were case-ambiguous with respect to the nominative-accusative distinction. Sixty-six (92%) of the attraction errors arose under these conditions. Here too the facts conform to the expectations of the working memory retrieval model. If nominative case is used as a retrieval cue for the agreement source, then one would expect retrieval errors (i.e., choosing a local noun over the true head of the subject phrase) only in situations in which the local noun resonates to the retrieval cue at least as well as the head noun does. Retrieval errors may well occur when both the head and local noun are specified [NOMINATIVE ∨ ACCUSATIVE], but they certainly should not be seen when the local noun is marked [ACCUSATIVE], since a local noun that is unambiguously accusative should not resonate as well to the nominative retrieval cue as either an unambiguously nominative ([NOMINATIVE]) or a case ambiguous ([NOMINATIVE ∨ ACCUSATIVE]) head noun. The final state of affairs to

consider is when the local noun is specified [NOMINATIVE ∨ ACCUSATIVE] and the head noun is [NOMINATIVE]. Once again, the unambiguously nominative head noun should have an edge over the case ambiguous local noun: The case-based retrieval cue should not err in selecting an agreement source that *might be* nominative over one that unambiguously *is* nominative.

There are two details concerning the stimuli used in Experiment 2 which deserve some additional discussion. As we noted in the method section for this experiment, Slovak requires all masculine case-unambiguous nouns to be animate, and most of these nouns in our stimuli denoted humans. All other nouns used in both mismatch conditions denoted non-humans. There were two conditions in which these animate nouns were employed: (a) as nominative head nouns in all of the gender match and mismatch combinations for the case-unambiguous, masculine head noun conditions; and (b) as accusative local nouns in gender mismatch conditions for feminine head nouns and case-unambiguous local nouns. One might be concerned that what appeared to be an effect for case ambiguity might instead have been driven by the animacy of these case-unambiguous masculine nouns. This concern is unwarranted for several reasons, not the least of which is that previous studies that manipulated head and local noun animacy in English (where animacy is not confounded with case ambiguity) did not find any evidence that a noun phrase with an inanimate head and animate local noun conditions more attraction errors than a noun phrase with an animate head and an inanimate local noun phrase (Bock & Miller, 1991, Barker, Nicol, & Garrett, 2; Barker et al., 2001). If animacy were the actual source of the effects attributed to case ambiguity in the present study (e.g., the greater ability of an animate, case-unambiguous masculine head noun to resist attraction in comparison to an inanimate, case-ambiguous masculine head noun), then one would expect that the animate masculine nouns would also be better agreement controllers in gender mismatch conditions either when they are compared with (case-unambiguous) inanimate feminine head nouns, or when they are compared with case-ambiguous masculine nouns in *Feminine–Masculine* conditions. Both comparisons suggest that appeals to animacy are misplaced. When compared to conditions in which the head noun was case ambiguous, unambiguously nominative head nouns were resistant to attraction both in *Masculine–Feminine* mismatch conditions (where animacy and case-ambiguity of the masculine head are confounded) and *Feminine–Masculine* mismatch conditions (where the feminine head noun is unambiguously nominative, but inanimate). Likewise, when compared to conditions in which the local noun was case ambiguous, the inanimate (and case-unambiguous) feminine local nouns were no less effective at attracting agreement than the animate (case-unambiguous) masculine local nouns.

The second concession that was made to experiment length and to the rarity of case-ambiguous feminine nouns in Experiment 2 was that the local noun was unambiguously accusative in the "match" condition with feminine head nouns and case-ambiguous in the "match" condition with masculine nouns. If one grants, as we maintain, that the case-ambiguity effects for both head and local nouns are authentic, then one might have some concerns about the impact that local noun case in *Feminine–Feminine* and *Masculine–Masculine* match conditions might have on our arguments from the experiment concerning the markedness asymmetry between Masculine and Feminine genders. First, one must note that this difference across gender conditions was only for the match conditions: If the case-ambiguity of the local noun in a match conditions had any effect on agreement errors, it would be that one should observe a lower error rate in the gender-match condition with accusative local noun. But that would entail that the difference between *Feminine–Feminine* match and *Feminine–Masculine* mis-match conditions would be exaggerated if the *Feminine–Feminine* match condition has an accusative local noun and the *Feminine–Masculine* mismatch condition has a case ambiguous local noun. In other words, if the observed difference between *Feminine–Feminine* and *Feminine–Masculine* conditions is greater than what one might other-wise expect, then that would suggest that the true asymmetry between Masculine and Feminine local nouns is underestimated by our results (rather than being exaggerated by our findings). Hence, if anything the concessions that were made in order to accommodate the rarity of case-ambiguous feminine nouns must be viewed as stacking the deck against finding evidence for the markedness asymmetries we observed among genders in Slovak.

Experiment 3

The results of Experiments 1 and 2 demonstrate that the rate of agreement attraction is indeed modulated by the relative markedness of the three Slovak genders: We found more gender attraction errors with *Neuter–Masculine* subject phrases than with *Masculine–Neuter* subject noun phrases (Experiment 1), and more errors with *Masculine–Feminine* subject phrases than with *Feminine–Masculine* subject phrases—even when the confound with case-ambiguity is fully eliminated (Experiments 1 and 2). These findings are what one would expect if markedness must be understood in purely relational terms, e.g., from a bias based on ranked faithfulness constraints. The purpose of Experiment 3 is to test the final prediction of this account. Even though Masculine and Feminine exhibit an asymmetry when pitted against one another directly, the faithfulness account predicts that one should not observe a difference between the effect that Masculine and Feminine local nouns have when they occur with a Neuter Head noun. We examined this prediction by comparing the rate of agreement errors in a *Neuter–Neuter* match condition with two mismatch conditions: *Neuter–Masculine* and *Neuter–Feminine*.

Method

Participants

Forty-eight native Slovak speakers, recruited from the Mlynská Dolina university housing complex, in Bratislava, Slovakia, participated in this experiment in exchange for a modest financial reward. Participants' ages ranged 19–27, with an average of 21, and all had completed at least 12 years of formal education. None of the participants in this experiment took part in Experiments 1 or 2.

Materials

Twenty-one experimental items were constructed. Each item consisted of a noun phrase with a single nominal head and a prepositional phrase modifier. All preamble heads were grammatically Neuter nouns. The three experimental conditions were based on the Gender of the local noun, and consisted of one matching (Neuter) and two mismatching (Masculine or Feminine) Gender conditions. All head and local nouns were singular and case-ambiguous (Nominative or Accusative). Example stimuli are given in Table 5.

Eighty-one filler preambles were also constructed. These were noun phrases with prepositional phrase or adjectival modifiers or coordinate phrases. In sum, there were 3 filler items with neuter singular heads, 21 with masculine singular heads, 21 with feminine singular heads (both case-ambiguous and case-unambiguous), and 36 with plural heads.

Three lists were constructed out of these materials. Each list consisted of 102 preambles, 21 experimental and 81 filler items. The experimental items were balanced across lists and conditions. The order of items was pseudo-random; no two experimental items could occur consecutively. Each list was broken up into three parts of equal length. Each of these parts began with

Table 5
Examples of the experimental materials from Experiment 3

Neuter–Neuter (match)	*Múzeum*$_N$ *pre okolie*$_N$	The museum for the locale
Neuter–Masculine (mismatch)	*Múzeum*$_N$ *pre okres*$_M$	The museum for the county
Neuter-Feminine (mismatch)	*Múzeum*$_N$ *pre obec*$_F$	The museum for the village

at least three fillers, and the last item in each part was also a filler. The order of presentation of the three parts was kept the same. The method of cuing past-tense responses for the target items was the same as the previous experiments, as was the secondary memory task.

Procedure

The procedures for this experiment were the same as those in Experiments 1 and 2.

Results

Application of the scoring criteria yielded 788 correct responses (77.4% of all responses), 30 agreement errors (2.9%), 143 (14%) topicalized objects, and 57 (5.6%) miscellaneous responses. Of the 30 agreement errors, 22 (73%) were errors of attraction.

The numbers of responses obtained in the individual scoring categories are shown in Table 6, broken down by stimulus sets and the match/mismatch condition. If any of the agreement errors were non-attraction errors, the number of non-attraction errors appears in parentheses after the total number of agreement errors.

Responses to the sentence completion task were examined using repeated measures analysis of variance with participants and items as random factors; the analyses of agreement errors by condition for this and subsequent experiments was based on the proportions of scorable responses (i.e., the proportions of agreement errors among correct + agreement error responses for each participant and for each item). Analysis verified that the three conditions differed with respect to the rate of agreement errors: $F_1(2,47) = 4.04$, $p < .05$; $F_2(2,20) = 6.15$, $p < .01$ [min $F'(2,64) = 2.45$]. Planned comparisons revealed that there were significantly fewer gender errors in the *Neuter–Neuter* match condition (mean = 0.01) than in the *Neuter–Masculine* mismatch condition (mean = 0.05; CI = ±0.03 for a means difference of .04; min $F'(1,66) = 4.41$, $p < .05$) and in the *Neuter–Feminine* condition (mean = 0.05; CI = ±0.027 for a means difference of 0.04; $F_1(1,47) = 6.39$, $p < .05$; $F_2(1,20) = 6.21$, $p < .05$ [min $F'(1,55) = 3.14$]). However, there was no reliable difference in agreement error rates between the *Neuter–Masculine* and *Neuter–Feminine* preambles ($Fs < 1$). Further analysis revealed that there was no effect of preamble type on the rate of Topicalization errors.

Discussion

The combined results from Experiments 1 to 3 pose an interesting puzzle regarding the status of markedness in the processing system. The findings from Experiments 1 and 2 indicate that Feminine behaves as if it were more marked than Masculine, but *Neuter–Feminine* preambles in Experiment 3 did not elicit more agreement errors than the *Neuter–Masculine* preambles elicited. These facts together place an important constraint on the production system, in that it appears that the markedness status of a particular gender is defined in terms of particular oppositions (i.e., Feminine vs. Masculine; Masculine vs. Neuter; and Feminine vs. Neuter) in the local context, rather than in terms of a static ranking of the three genders (e.g., as expressed by differences in activation level). This conclusion is problematic for accounts of agreement errors based, for example, on the behavior of privative features, and for accounts based on the activation level of features themselves. On either account, if the markedness of a gender modulates a feature's "visibility" or "activation strength", then we would expect to see either (a) equal numbers of attraction errors with *Masculine–Feminine* and *Feminine–Masculine* preambles (on the assumption that they do not differ with respect to markedness); or (b) more attraction errors with *Neuter–Feminine* preambles than with *Neuter–Masculine* preambles (on the assumption that Feminine and Masculine differ with regard to their markedness). The results from Experiments 1 and 2 are at odds with the first of these predictions, and those from Experiment 3 conflict with the second. However, if markedness asymmetries follow, as they do in optimality based accounts, from a competition governed by ranked constraints (i.e., rather than from properties of representations that render them intrinsically marked vs. unmarked on some grammatical dimension), then the observed pattern of attraction asymmetries falls neatly in step with theory.

General discussion

The asymmetric interference effects that we observed in the gender agreement studies described here, along with the striking effect of case ambiguity, impose a number of interesting constraints on accounts of agreement

Table 6
Number of sentence completions in the different scoring categories for each stimulus type in Experiment 3

Noun phrase type	Correct	Agreement error	Topicalization error	Miscellaneous error
Neuter–Neuter	273	2 (2)	41	20
Neuter–Masculine	262	15 (2)	50	9
Neuter–Feminine	253	13 (4)	52	18

The number of non-attraction agreement errors (i.e., where the verb form does not conform to the gender of the head or local noun) is enclosed in parentheses.

attraction, and ultimately on theories of agreement planning that derive support from those accounts. For example, one such account of attraction asymmetries in the domain of number agreement is based on the contribution of distributional factors to the calculation of agreement (Haskell & MacDonald, 2003). On this account, the origin of the singular/plural asymmetry lies in the fact that experience with syntactic patterns such as pseudo-partitive constructions ("a group of lawyers") biases the planning system to produce plural agreement when a subject noun phrase exhibits the structural arrangement *singular noun*-[*preposition-plural noun*] (whereas no *plural noun*-[*preposition-singular noun*] construction grammatically licenses singular agreement). However, there are no parallel structures in Slovak (or any other language that we are aware of) where the gender associated with a phrase of the form *masculine noun*-[*preposition-feminine noun*] grammatically conditions feminine agreement. This and the fact that we observed asymmetries in the production of gender agreement with masculine and feminine noun phrases indicates, at a minimum, that distributional factors cannot serve as the general means for deriving markedness asymmetries.

The implications for other markedness based accounts, such as the privative feature proposal, and for the feature-activation based proposal of Eberhard et al. (2005) are also of some importance. On both proposals, it is fairly unproblematic to accommodate an attraction asymmetry within a two-valued agreement system. For example, on the feature-activation account of the number asymmetry observed in many languages (see Bock, 2004, for a review), the asymmetry derives from the fact that singular is taken as a default value (e.g., singular agreement is the norm for both singular subjects like "*the man with the plan*" and for infinitival or gerundive subjects that are fully unspecified for number), whereas plural agreement can arise only if the subject carries a significant level of non-zero activation to pass on to the agreeing predicate. However, if a third agreement value is added into the mix, then it may be difficult to attribute the asymmetry to the distinction between default and non-default—just as it is in the case of the data from Slovak gender agreement. In particular, the fact that we observed attraction asymmetries between Neuter and Masculine and between Masculine and Feminine indicates that the "weaker" partner in the gender pair cannot get this status in every case by equating "weaker" with the default interpretation of a zero activation level for a single, continuously valued agreement feature. It is also paradoxical, on such an account, that Masculine gender would be a weaker attractor than Feminine when the two are compared directly in *Feminine–Masculine* and *Masculine–Feminine* mismatch conditions (as they were in Experiments 1 and 2), but not when the two genders are compared indirectly in *Neuter–Masculine* and *Neuter-Feminine* mismatch

conditions (as in Experiment 3). The observed asymmetries cause similar problems for accounts which exploit privative features in percolation based models of agreement (Franck et al., 2002; Vigliocco & Nicol, 1998; Vigliocco et al., 1995, 1996a, 1996b). In contrast, this complex pattern of asymmetries (and symmetries) is just what one would predict if the relative ability of different genders to attract agreement reflects a relational approach to markedness and faithfulness constraints (e.g., as proposed in optimality based grammatical frameworks).

A reviewer suggested that the masculine/feminine asymmetry we observed in Experiments 1 and 2 might stem from a mechanism set up to prevent gender retrieval errors caused by form-based family resemblance. Case-ambiguous masculine nouns end in a consonant (taking a zero suffix) in the nominative and accusative, following the statistically most common pattern for masculine nouns. Case-ambiguous feminine nouns also end in a consonant, but they are in a minority among feminine nouns; most feminine nouns end in -*a*. According to this reviewer, the gender of the non-prototypical (consonant final) feminine nouns must be especially well learned during acquisition, to avoid gender retrieval errors that might otherwise be caused by their family resemblance to many masculine nouns. If these feminine nouns suppress retrieval of masculine gender and preferentially retrieve feminine gender instead, this might explain why masculine-to-feminine attraction was more frequent than feminine-to-masculine attraction.

This interesting hypothesis would deserve further testing; unfortunately, markedness, case-ambiguity, animacy and form-based resemblance are too interdependent in Slovak to be able to clearly tease them apart. In the absence of corroborating evidence, however, it is worth noting that the proposal is not without problems. If the asymmetry between masculine and feminine that we attribute to markedness were actually due to the exceptional morphology of case-ambiguous feminine nouns, one might also predict that performance asymmetries would show up in other agreement situations in which prototypicality and morpho-syntactic features collide. If in Slovak a case-ambiguous feminine noun that ends in a consonant (an atypical ending for feminine nouns, though typical for masculine nouns) is more likely to condition attraction than a local noun with the prototypical form for its gender because the atypical form provokes over-learning, then by parity of reasoning one would expect that in English local irregular plurals like *feet* and *geese* should be more likely to condition attraction than local regular plurals like *meats* and *leases*: their similarity to singular forms (e.g., *meat*, *treat*, *wheat*; *lease*, *niece*, *grease*) would, on this proposal, lead one to expect that local irregular plural nouns would be more effective than regular plurals in propagating plural agreement features. The available evidence

runs strongly against this prediction (Bock & Eberhard, 1993; Haskell & MacDonald, 2003).

The reviewer also noted that learning asymmetries might arise from simple frequency differences. Evidence for such an account might be seen in the interactions between the frequency of phonemic segments and the default status of features in the production of phonological errors. Stemberger (1991) noted that it is more common for an unmarked phonological segment like /d/ (which on some proposals is unspecified for the default place feature [coronal]) to be replaced by a marked segment such as the labial stop /b/ or the velar stop /g/ (which must be specified, respectively, as [labial] or [velar]) than for the opposite substitution pattern to occur. Stemberger (1991) attributes this anti-frequency effect to a bias for adding features (to an unmarked, or underspecified representation) over removing features that explicitly mark a non-default form. However, when this "addition bias" does not favor either of two competitors (e.g., when two explicitly marked options are placed in competition), then substitution asymmetries in phonemic errors should favor the more frequent of the two candidates (Stemberger, 1991).

The problem for an account of the masculine/feminine asymmetry based on relative frequency, as the reviewer noted himself, is that the predicted frequency-based bias (when two non-default genders compete) is not supported by our results. Feminine nouns are not more frequent than Masculine nouns; and in most frequency ranges (including that represented in our materials) they are slightly less frequent than Masculine nouns on both type and token counts of noun frequency (Mistrík, 1969). So, if any agreement asymmetry between the two non-default genders were to be observed at all on this proposal, then one might expect it to be one that tilts in favor of Masculine as the more effective attractor—not, as observed here, in favor of Feminine.

Approaches to explaining attraction asymmetries must also be examined with respect to their accounts of the more general phenomenon of agreement attraction. How does the gender (or number) of a local noun phrase (or its lexical head) come to influence the process of subject–predicate agreement? Our data speak to this issue as well. First, if the relational approach to markedness provides the correct account of attraction asymmetries, then this would suggest that attraction itself must be understood, in part at least, as involving a competition between candidate genders (since the relative markedness of a particular gender value can only be evaluated in the context of a specific competitor). On the working memory retrieval model we have outlined, the calculation of agreement will be confronted with such competitions when content-based retrieval mechanisms fail to uniquely identify the correct source of agreement (whether we construe that source in lexical, phrasal, or jointly lexical and phrasal terms). Crucially, though, this competition is set up under conditions where retrieval cues are overly inclusive for the purpose of isolating and identifying an agreement source in a particular context. Setting aside the matter of how decay affects the representational elements that are held in working memory, the hypothesized retrieval cues should include subject-oriented cues such as those based on case and syntactic position. Accordingly, the retrieval cues will fail to find a unique candidate when more than one representational element resonates strongly to the retrieval cue, such as when both a local noun phrase and the actual subject noun phrase resonate on a par with one another to the nominative retrieval cue. Since this can only happen if nominative case is a strong retrieval cue and both the head and local noun phrases are case ambiguous, this account predicts the effects of case ambiguity observed here and in Hartsuiker et al. (2003).

Whether feature percolation and spreading activation accounts of agreement and attraction can be made to accommodate the effect of case ambiguity remains an open question at this point. For example, one might attempt to account for the effect of case ambiguity observed in this study (and in Hartsuiker et al., 2003) by hypothesizing that specific case features, when present in a structure, could inhibit the transmission of gender or number feature values (Kay Bock, personal communication). One question that arises is what specific case features must carry this account: The feature [ACCUSATIVE] could do the work, but it is not needed (because attraction from a case ambiguous local noun is blocked by an unambiguously nominative head); [NOMINATIVE] could do the work, but it also is not required (since an unambiguously accusative local noun will not attract agreement from a case-ambiguous head). These observations suggest that either feature should be strong enough to inhibit the transmission of the local noun's feature to the root of the subject phrase. But it will not be enough to imbue these abstract features with the inhibitory power, if only because it matters so much how the features are introduced into a specific syntactic context. The studies here show that the lexical head nouns of the subject and local noun phrases are strong disambiguating sources, and Hartsuiker et al.'s (2003) results also suggest that determiners that carry case information can also block attraction if they are unambiguous for case. But many familiar grammatical frameworks (e.g., Pollard & Sag, 1994) would also lead one to expect case disambiguation coming from outside a noun phrase, which one might expect would also stop the spread of errant features. For example, the preposition "na" (for) in a structure like "Ret'az na bicykel" (the chain for the bicycle) assigns accusative case to its complement. Why does this unambiguous case feature not inhibit the spread of features (or feature values) that originate in the complement noun phrase?

Though we will not try to argue that a mechanism using case to inhibit the transfer of number or gender features to the subject's root node cannot be made to work, we also will refrain from attempting to evaluate the proposal in its underspecified form. What we can do, though, is suggest that there should be a place to look in order to compare the memory retrieval proposal with an account based on spreading activation of agreement feature values (plus inhibition based on case form). If a specific case configuration blocks attraction on the spreading activation proposal, it should not matter what type of agreement relation the noun phrase enters into: The case forms that block the spread of activation from one number feature to another should exhibit this blocking effect an any instance of attraction (e.g., pronoun agreement). This is because the spreading-activation account locates the effect of case in the calculation of the subject noun phrase's agreement features (not in the mechanism that gets some other syntactic element to agree with the subject). The memory retrieval account, on the other hand, attributes the effect of case to the fact that (nominative) case is a retrieval cue for subjects. If identifying the antecedent of a pronoun does not exploit subject-oriented retrieval cues, then one should observe a dissociation between the effect of case ambiguity in subject–verb agreement and in pronoun-antecedent agreement.

The working memory retrieval account of agreement processing locates attraction errors in competition among elements in working memory that retrieval mechanisms have identified as agreement sources. Contests between head and local nouns arise because the mechanism for retrieving the agreement source (cue-based retrieval) makes the agreement process susceptible to mis-identifying a source. The susceptibility is due at least in part to the fact that some local nouns will resonate to certain retrieval cues as well as the intended retrieval target (the head of the subject phrase). Models of working memory also identify representational similarity among elements in memory as a factor that can contribute to the loss of information about their relative order (Lewandowsky & Murdock, 1989), and it is not difficult to see how this too could undermine processes for encoding and retrieving agreement sources during on-line sentence planning (Lewis & Vasishth, 2005). So, in addition to encoding similarities between head and local nouns based on morpho-syntactic features for case, there may be other sources of similarity that also contribute to the probability of retrieval errors (and therefore agreement errors). Gordon et al. (2001) propose, for example, that in addition to case, similarity in gender, number, animacy and/or person are syntactic features that also may diminish the ability to encode or maintain a distinct representation of grammatical role for two or more referring elements in working memory. Evidence supporting effects of similarity on agreement may be observed, for example, in Barker et al.'s (2001)

discovery that referent animacy raised rates of attraction only when the head and local nouns were either both animate or both inanimate.

Finally, the retrieval model maintains that as the representation of syntactic relations among sentence elements decays over time, keeping head and local noun apart will depend on the number of structural distinctions that are maintained between the two. One simple way in which syntactic structure can be encoded is in terms of dominance and precedence relations (Gorrell, 1995; Marcus, Hindle, & Fleck, 1983). For example, the hierarchic structure indicated by the bracketing [NP$_i$ *the picture* [PP *of* [NP$_j$ *the boys*]]] can be encoded as a set of dominance relations {Dom(NP$_i$, *the*), Dom(NP$_i$, *picture*), ..., Dom(NP$_i$, *boys*), Dom(NP$_i$, PP), ... Dom(NP$_i$, NP$_j$), Dom(PP, NP$_j$), Dom(NP$_j$, *boys*), ...}. As more of these dominance relations are lost to the effects of decay, the probability increases that working memory will fail to maintain any structural asymmetry between the head noun '*picture*' and the local noun '*boys*'. Conversely, the more of these relations that initially distinguish a head and local noun, the more likely it is that the syntactic role of a head and local noun can be kept distinct. The probability that subject-oriented retrieval mechanisms will wrongly recover a local noun (rather than the lexical head of the subject) is inversely proportional to the local noun's syntactic depth in the subject noun phrase (i.e., to the number of dominance relations that it enters into as the dominated element). Alternatively (or in addition), working memory may be aided in keeping head and local noun apart if it has access to information that identifies the planning increment that contains the subject head. Numerous studies of sentence production have shown that sentence planning is carried out in increments (Clark & Wasow, 1998; Ford & Holmes, 1978; Garrett, 1980, 1982; Holmes, 1988), and the probability that a local noun will be included in the same planning chunk as the head noun of a subject also diminishes as the syntactic depth of the local noun increases. On either approach (or on some combination of the two), the working memory retrieval model of agreement production can account for the effect of syntactic embedding (Bock & Cutting, 1992; Franck et al., 2002) without requiring features to literally climb through a data-structure as the feature percolation hypothesis would require them to do during the normal calculation of agreement relations.

Appendix A

The Slovak gender system and markedness

Each Slovak noun bears one of three grammatical genders (Masculine, Feminine or Neuter), and these nouns condition agreement processes with determiners, attributive and predicate adjectives, participial verbs, and—in the past tense—with finite

verbs. Subject–Verb agreement requires that verbs distinguish the three nominal genders in the past tense singular: *muž*$_{M,sg}$ *odišiel*$_{M,sg}$ 'the man left', *žena*$_{F,sg}$ *odišla*$_{F,sg}$ 'the woman left', *dieťa*$_{N,sg}$ *odišlo*$_{N,sg}$ 'the child left'. These gender distinctions are neutralized in plural verbal morphology: *muži/ženy/deti odišli* 'the men/women/children left'. To establish markedness relations between the three Slovak genders, one can exploit the criteria summarized and evaluated by Battistella (1990). As Battistella notes himself, though, one must bear in mind that these criteria may at times provide incomplete or contradictory evidence regarding markedness relations.

Evidence that Feminine gender is marked with respect to Masculine is abundant for nouns with human referents: The word *student*$_M$ can be used to refer to a male student or a student of unspecified sex; the related feminine form *študentka*$_F$ can only refer to a female student. Similarly, *študenti*$_{M,pl.}$ can be used to refer to a mixed group of males and females, while *študentky*$_{F,pl.}$ can only refer to a group consisting entirely of females. The equivalents of 'who', 'nobody', 'anyone', etc., are grammatically always masculine (*Kto otehotnel*$_M$? 'Who became pregnant?'), making Masculine the unmarked option for human referents. Outside the domain of human referents, relatively impoverished Feminine adjectival morphology (cf. Masculine singular genitive suffix-*ého*, dative-*ému*, locative-*om* and Feminine singular genitive/dative/locative-*ej*) and overt suffixes on Feminine pronouns and past tense verbs argue for the markedness of Feminine gender.

There is also evidence that Neuter is the maximally unmarked, or default, gender of the three. Non-nominal subjects, such as empty, infinitival or clausal subjects, condition Neuter agreement: *pršalo*$_N$ 'it rained$_N$'; *fajčiť zakázané*$_N$ 'it is prohibited$_N$ to smoke'; *prekvapilo*$_N$ *ma, že ste prišli* 'it surprised$_N$ me that you came.' Acronyms, names of letters, or quoted words typically condition Neuter agreement: *x* /iks/ *je menšie*$_N$ *ako 5* 'x is smaller$_N$ than 5', *hlasné*$_N$ *"ach"* 'a loud$_N$ 'ah''. 'What', 'nothing' and 'anything' are also Neuter. However, the relative markedness of Neuter and Masculine is not fully evident. For example, for specifically human referents Masculine, not Neuter, appears to be the default, and in terms of overt marking, Neuter gender is overtly marked with a suffix on pronouns and past tense verbs while Masculine is not.

In summary, Feminine appears marked in comparison to both Masculine and Neuter, and there is at least some evidence which indicates that Masculine is marked in comparison to Neuter, at least outside the domain of human referents.

Morphophonology vs. morphosyntax

In Slovak, there are morphological sub-classes of nouns which differ from one another in their case forms. Membership in these classes is based in some instances on semantic properties of the noun (e.g., animacy for Masculine nouns), in other instances on some combination of phonological criteria (e.g., ending in a hard vs. soft consonant) and arbitrary class membership (Rubach, 1993; Short, 1993).

Across different morphological sub-classes, the same suffix can encode different case and number features. For example, the suffix -*a* in the Masculine word *sluha* 'servant' unambiguously marks the noun as nominative singular, but in the word *muža* '(of a) man', a Masculine word from another morphological sub-class, it marks the noun as genitive or accusative singu-

lar. The -*o* suffix in the Masculine word *dedo* 'grandpa' stands for nominative singular, but in the Neuter word *miesto* 'place' it is ambiguously nominative or accusative singular. The zero suffix in the Masculine word *muž* 'man' signifies unambiguous nominative singular, in the Masculine word *stroj* 'machine' it is ambiguously nominative or accusative singular, and in the Neuter word *miest* 'of places' it is genitive plural. In general, it is not possible to determine a word's case or its case ambiguity status on the basis of the suffix alone unless the word's morphological sub-class is identified first.

This many-to-many relationship between Slovak case features and suffixes has implications regarding the possible mechanisms for case-ambiguity effects. In German, for example, the article *die* always indicates a case-ambiguous, nominative or accusative, noun phrase (that it is either both Singular and Feminine on the one hand, or Plural with gender unspecified on the other). On the feed-back view, the phonological form *die* in a local noun phrase can feed activation back to both nominative and accusative case features, sometimes interfering with the computation of agreement. The ambiguous morphological element in this case is a free-standing grammatical formative. In Slovak, in comparison, the phonological form of a suffix such as -*a*, which indicates nominative for certain nouns and accusative for others, cannot simply feed back to both nominative and accusative case. Otherwise, nouns ending in -*a* would be associated with agreement errors as if they were ambiguous (since the suffix itself is ambiguous). But the unambiguously nominative masculine and feminine nouns that end in -*a* and the unambiguously non-nominative (genitive/accusative) masculine nouns that end in -*a* do not allow gender errors to occur (see Experiment 2). The absence of errors after case-unambiguous nouns with potentially ambiguous suffixes shows that Slovak case ambiguity effects are not based on ambiguity of the phonological form of the suffix *per se*; being ambiguous or unambiguous is a property of a lexical item and its declension class. This would suggest that the type of multiple specification that can give rise to case-ambiguity effects must be tied to a specific inflected lemma in order for the ambiguity to function as ambiguous with regard to the agreement mechanism.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jml.2006.08.004.

References

Barker, J., Nicol, J., & Garrett, M. (2001). Semantic factors in the production of number agreement. *Journal of Psycholinguistic Research, 30*, 91–114.

Battistella, E. (1990). *Markedness: The evaluative superstructure of language*. Albany, NY: State University of New York Press.

Bock, K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*, 1–47.

Bock, K. (2004). Psycholinguistically speaking: Some matters of meaning, marking and morphing. In B. H. Ross (Ed.). *The psychology of learning and motivation* (Vol. 44, pp. 109–144). San Diego, CA: Elsevier.

Bock, K., & Cutting, J. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language, 31*, 99–127.

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

Bock, K., Eberhard, K., Cutting, J. C., Meyer, A., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive Psychology, 43*, 83–128.

Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego: Academic Press.

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

Bock, K., Nicol, J., & Cutting, J. C. (1999). The ties that bind: Creating number agreement in speech. *Journal of Memory and Language, 40*, 330–346.

Bresnan, J. (2000). Optimal syntax. In J. Dekkers, F. van der Leeuw, & J. van de Weijer (Eds.), *Optimality theory: Phonology, syntax and acquisition* (pp. 334–385). Oxford: Oxford University Press.

Bresnan, J. (2001). Explaining morphosyntactic competition. In M. Baltin & C. Collins (Eds.), *Handbook of contemporary syntactic theory* (pp. 11–44). Oxford: Blackwell.

Clark, H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335–359.

Clark, H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 38*, 201–242.

Dell, G. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283–321.

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

Eberhard, K. (1999). The accessibility of conceptual number to the processes of Subject–Verb agreement in English. *Journal of Memory and Language, 41*, 560–578.

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

Fayol, M., Largy, P., & Lemaire, P. (1994). Cognitive overload and orthographic errors: When cognitive overload enhances subject–verb agreement errors. A study in French written language. *Quarterly Journal of Experimental Psychology A, 47*, 437–464.

Ferreira, F. (2000). Syntax in language production: An approach using tree-adjoining grammars. In L. Wheeldon (Ed.), *Aspects of language production* (pp. 291–330). Philadelphia, PA: Psychology Press.

Ford, M., & Holmes, V. (1978). Planning units and syntax in sentence production. *Cognition, 6*, 35–53.

Francis, N. W. (1986). Proximity concord in English. *Journal of English Linguistics, 19*, 309–318.

Frank, R. (2002). *Phrase structure composition and syntactic dependencies*. Cambridge, MA: MIT Press.

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.

Garrett, M. (1980). Levels of processing in speech production. In B. Butterworth (Ed.), *Language production, Vol. 1: Speech and talk* (pp. 177–220). New York, NY: Academic Press.

Garrett, M. (1982). Production of speech: Observations from normal and pathological language use. In A. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 19–76). New York, NY: Academic Press.

Gordon, P., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 1411–1423.

Gorrell, P. (1995). *Syntax and parsing*. Cambridge, UK: Cambridge University Press.

Harrison, A. J. (2004). Three way attraction effects in Slovene. In *Proceedings of the Second CamLing Postgraduate Conference on Language Research*. Cambridge: University of Cambridge.

Hartsuiker, R., & Barkhuysen, P. (2006). Language production and working memory: The case of subject–verb agreement. *Language and Cognitive Processes, 21*, 181–204.

Hartsuiker, R., Schriefers, H., Bock, K., & Kikstra, G. (2003). Morphophonological influences on the construction of subject–verb agreement. *Memory and Cognition, 31*, 1316–1326.

Haskell, T., & MacDonald, M. (2003). Conflicting cues and competition in subject–verb agreement. *Journal of Memory and Language, 48*, 760–778.

Holmes, V. (1988). Hesitations and sentence planning. *Language and Cognitive Processes, 3*, 323–361.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Lewandowsky, S., & Murdock, B. (1989). Memory for serial order. *Psychological Review, 96*, 25–57.

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

Masson, M., & Loftus, G. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57*, 203–220.

McCarthy, J. (2002). *A thematic guide to optimality theory*. Cambridge, UK: Cambridge University Press.

Marcus, M., Hindle, D., & Fleck, M. (1983). D-Theory: Talking about talking about trees. In *Proceedings of the 21st Annual Meeting of the Association of Computational Linguistics* (pp. 129–136). Menlo Park, CA: ACL.

Mistrík, J. (1969). *Frekvencia slov v slovenčine [Frequency of words in Slovak]*. Bratislava: Vydavateľstvo SAV.

Meyer, A., & Bock, K. (1999). Representations and processes in the production of pronouns: Some perspectives from Dutch. *Journal of Memory and Language, 41*, 281–301.

Pollard, C., & Sag, I. (1994). *Head-driven phrase structure grammar*. Chicago, IL: University of Chicago Press.

Prince, A., & Smolensky, P. (2004). *Optimality Theory: Constraint interaction in generative grammar*. Oxford, UK: Blackwell.

Rubach, J. (1993). *The lexical phonology of Slovak*. Oxford, UK: Oxford University Press.

Schriefers, H., & van Kampen, A. (1993). Syntaktische Prozesse bei der Sprachproduktion: Zur Numerus-Kongruenz zwischen Subjekt und Verb. *Sprache & Kognition, 12*, 205–216.

Selkirk, E. (1977). Some remarks on Noun phrase structure. In A. Akmajian, P. Culicover, & T. Wasow (Eds.), *Studies in formal syntax* (pp. 285–316). New York, NY: Academic Press.

Short, D. (1993). Slovak. In B. Comrie & G. Corbett (Eds.), *The Slavonic languages* (pp. 533–592). London: Routledge.

Smith, M., & Wheeldon, L. (2001). High level processing scope in spoken sentence production. *Cognition, 73*, 205–246.

Stemberger, J. (1991). Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language, 30*, 161–185.

Strang, B. (1966). Some features of S-V concord in present day English. In I. Cellini & G. Meclchiori (Eds.), *English Studies Today: Forth Series* (pp. 73–87). Rome: Edizioni di Storia et Letteratura.

Thornton, R., & MacDonald, M. (2003). Plausibility and grammatical agreement. *Journal of Memory and Language, 48*, 740–759.

Vigliocco, G., Butterworth, B., & Garrett, M. (1996a). Subject–verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition, 61*, 261–298.

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 and 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., Hartsuiker, R., Jarema, G., & Kolk, H. (1996b). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes, 11*, 407–442.

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

Wechsler, S., & Zlatić, L. (2003). *The many faces of agreement*. Stanford: CLSI Press.

Woolford, E. (2006). Case-agreement mismatches. In C. Boeckx (Ed.), *Agreement Systems* (pp. 317–339). Philadelphia, PA: John Benjamins.

Zaenan, A., & Maling, J. (1983). Passive and oblique case. In L. Levin, M. Rappaport, & A. Zaenen (Eds.), *Papers in lexical functional grammar*. Bloomington, IN: IULC.

Zandvoort, R. (1961). Varia syntactica. In *Language and Society: Essays Presented to Arthur M. Jensen on his 70th Birthday*. Copenhagen: Berlingske Bogtrykkeri.