Attraction Errors in Gender Agreement: Evidence from Russian*  

Anton Malko  
St. Petersburg State University  
Natalia Slioussar  
Utrecht institute of Linguistics OTS and St. Petersburg State University

1 Introduction

When we want to understand the inner workings of complex systems, we often turn to errors. The study of language is no exception: many methods are based on analyzing errors in production and our reaction to them in perception. In this paper, we look at one well-studied type of errors: attraction errors in subject–predicate agreement (cf. (1)-(2)).

(1) The key to the cabinets are rusty.  
(2) The hypotheses one entertain influence the outcome.

The term attraction is used to describe the following phenomenon: in (1) the verb are erroneously agrees not with the head of the subject DP, but with an intervening DP the cabinets, which is said to attract agreement. In (2), DP the hypotheses acts as an attractor. Importantly, this shows that attraction depends on the syntactic structure rather than on linear proximity, as it was initially assumed (cf. Jespersen 1924, a.o.).

* The experiment described in this paper was run in the Laboratory for Cognitive Studies, St.Petersburg State University. The grant 12-06-00382 from the Russian Foundation for Basic Research is gratefully acknowledged. We have benefitted from comments of many linguists and are particularly grateful to Colin Phillips.
Attraction errors frequently occur naturally and are produced in high numbers in experimental conditions. In comparison, agreement errors without attraction (cf. (3)) are very rare.

(3) The key (to the cabinet) are rusty.

In the last two decades, many experimental studies focused on attraction errors, primarily in English (cf. Bock & Miller 1991, Clifton et al. 1999, Eberhard et al. 2005, Solomon & Pearlmutter 2004, Vigliocco & Nicol 1998; Wagers et al. 2009 a.o.), but also in other languages: French, Italian, Spanish, Dutch, German etc. (cf. Fayol et al. 1994, Franck et al. 2002, 2006, Hartsuiker et al. 2003, Vigliocco et al. 1995, 1996, a.o.). Analyzing syntactic, semantic and morphological factors affecting their frequency, we gain insights into the inner working of the grammar. For example, how does agreement, one of the basic syntactic operations, work in real time? How are various features that take part in it represented in the grammar?

Some important findings that are most relevant for our research are presented in section 2, along with several remaining questions. One of them motivated our interest to attraction errors in gender agreement. All previous studies focused on number agreement, with the exception of Badecker and Kuminiai’s (2007) work on Slovak, which is discussed in section 3. Section 4 presents our own experiment on Russian. Section 5 contains the general discussion and conclusions.

2 Previous studies: some findings and remaining questions

First of all, let us say a couple of words about experimental methods. The principle method of studying agreement errors is eliciting them in experimental conditions and identifying the factors that influence their frequency. Usually participants of an experiment are provided with DPs (termed preambles) and are asked to compose sentences using them as subjects (cf. Bock and Miller 1991, a.o.). Alternatively, participants are first given predicates in the indefinite form, then subject DPs and are asked to combine them in one sentence (cf. Vigliocco et al. 1995, a.o.). Agreement errors are also studied in perception. It was demonstrated that people tend to overlook the same errors that they produce more often (cf.
Pearlmutter et al. 1999, a.o.). This tendency can be traced in reading times, in grammaticality judgment accuracy and in ERP data.

The first approaches to agreement attraction explained it by linear proximity, arguing that the verb agrees with the closest DP rather than with the subject (cf. Francis 1986, Jespersen 1924, Quirk et al. 1972, a.o.). Then many studies proved that the linear order plays a very minor role (if any at all), and the phenomenon depends on the syntactic structure of the sentence. For example, in (Franck et al. 2002), preambles like (4a) provoked significantly more errors than preambles like (4b), where the potential attractor is linearly closer to the verb, but hierarchically further away from the head of the subject DP. Examples like (2) above also illustrate this point very convincingly.

(4) a. the inscription on the doors to the toilet
   b. the inscription on the door to the toilets

In all examined languages, only plural attractors provoked a significant number of agreement errors, while the incidence of errors with singular attractors (cf. (5)) was comparable to the incidence of errors without attraction (cf. (2)). This pattern is referred to as the plural markedness effect. Most proposed accounts claim that number is a privative feature, i.e. only plural nouns are marked for number, while singular ones are not, or that singular is the default value of this feature. Consequently, only plural nouns can interfere with number agreement (cf. Eberhard et al. 2005, Franck et al. 2002, Vigliocco et al. 1995, a.o.).

(5) The keys to the cabinet is rusty.

As it seems to us, to test these explanations and to find out whether attraction indeed depends on default or zero vs. non-default feature values, it is crucial to look at other features – for example, at three-member gender systems. We address this question in our study, so we will come back to it shortly.

Data from languages with morphologically marked cases show that the formal resemblance of the attractor to a subject is an important factor. For example, in an experiment on German (Hartsuiker et al. 2003), preambles like (6a), where the form of the attractor is ambiguous
between accusative and nominative, provoked significantly more errors than preambles like (6b), where the attractor is unambiguously dative.

(6) a. die Stellungnahme gegen die Demonstrationen
    the\textsubscript{F,NOM,SG} position against the\textsubscript{ACC,PL} demonstrations

b. die Stellungnahme zu den Demonstrationen
    the\textsubscript{F,NOM,SG} position on the\textsubscript{DAT,PL} demonstrations

Finally, there are different views on where exactly attraction errors arise. Most authors argue that this happens when subject DPs are constructed (both in production and in perception): they erroneously take the number feature not from their head, but from a dependent element (cf. Eberhard et al. 2005, Franck et al. 2002). Arguments supporting a different point of view can be found in (Wagers et al. 2009). According to it, errors arise when correctly constructed subject DPs are accessed to determine the number on the agreeing verb (constructing a sentence in production or checking its correctness in perception). Several DPs are active in our working memory at this point, and if an attractor formally resembles a subject, we are prone to confusion.

3 Badecker and Kuminiak’s (2007) study

Now let us turn to attraction errors in gender agreement, which we focus on in our study. In the first experiments where the authors looked both at number and at gender, virtually no such errors were found (cf. Vigliocco et al. 1995). However, a recent study on Slovak (Badecker and Kuminiak 2007) showed that gender agreement is also vulnerable to attraction and exhibits markedness effects.

Badecker and Kuminiak used the sentence completion method. Slovak has three genders, and verbs agree with their subjects in gender only in past tense singular forms. So participants were given preamble DPs (that they were asked to use as subjects) and tense adverbials to make sure that generated target sentences contain past tense verbs. Conditions and sample preambles from Badecker and Kuminiak’s experiment 1 are presented in Table 1. Target preambles contained only singular nouns, and the forms of intervening DPs were ambiguous between accusative and nominative.
Participants made very few errors without attraction in conditions 1-4 and significantly more attraction errors in conditions 5-8. In condition 6 (M–N) there were significantly more errors than in condition 5 (N–M). In condition 7 (M–F) there were significantly more errors than in condition 8 (F–M). In another experiment, the authors showed that the number of errors in conditions ‘N–M’ and ‘N–F’ is comparable.

Badecker and Kuminiak concluded that feminine is more marked than masculine, and masculine is more marked than neuter. Importantly, they did not suggest a markedness hierarchy (F>>M>>N), but rather markedness relationships in gender pairs. However, this conclusion is based only on one study on Slovak. In our opinion, it is important to see whether this pattern will be replicated in other languages. So we planned a similar study on Russian.

4 Our experiment

4.1 Subjects, materials and procedure
30 native speakers of Russian, 8 male and 22 female, participated in our experiment. Ages range from 18 to 50.

Russian is similar to Slovak with respect to the grammar features relevant for our study. Russian also has three genders, and verbs agree with their subjects in gender only in past tense singular forms. In some nominal paradigms, nominative and accusative forms are different, in the

<table>
<thead>
<tr>
<th>Head: gender</th>
<th>Attractor: gender match conditions</th>
<th>Attractor: gender mismatch conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>M (condition 1), e.g. pohár na čaj ‘glassM,NOM,SG for teaM,ACC,SG’</td>
<td>N (condition 5), e.g. pohár na mlieko ‘glassM,NOM,SG for milkN,ACC,SG’</td>
</tr>
<tr>
<td>N</td>
<td>N (condition 2)</td>
<td>M (condition 6)</td>
</tr>
<tr>
<td>M</td>
<td>M (condition 3)</td>
<td>F (condition 7)</td>
</tr>
<tr>
<td>F</td>
<td>F (condition 4)</td>
<td>M (condition 8)</td>
</tr>
</tbody>
</table>

Table 1. Conditions in Badecker and Kuminiak’s (2007) experiment

1 Both condition 1 and condition 3 contain ‘M–M’ preambles. But masculine dependent nouns from condition 1 alternate with neuter nouns in condition 2 (examples are given in the table), while masculine dependent nouns from condition 3 alternate with feminine nouns in condition 4 (e.g. spor o klenot/korist ‘quarrelM over jewelF/lootF’).
others they coincide in plural or in both numbers. Thus, many features of our study replicated those from Badecker and Kuminiak’s (2007) first experiment.

In our pilot experiment, we also used the same method as Badecker and Kuminiak did, but did not see any gender agreement errors after running six subjects. This can be explained by the fact that in morphologically rich languages such errors are in general very rare (cf. Lorimor et al. 2008). In Badecker and Kuminiak’s study, they occurred in 3% cases on average. Since the number of errors varies from subject to subject, the probability to elicit no errors from several people in a row is considerably high. Nevertheless, we decided to switch to a different method in our main experiment in hope to elicit more errors.

We used a modified version of the method developed by Vigliocco et al. (1995). In every trial, participants first saw a predicate, then a subject and were asked to combine them into one sentence and to say it aloud. In half of the trials, the predicates did not agree with the subjects in gender. In Table 2, experimental conditions are listed and several examples are provided (we had twice as many conditions as Badecker and Kuminiak did because of the predicate match/mismatch factor).

<table>
<thead>
<tr>
<th>Subject DP: the gender of the head and the attractor</th>
<th>Predicate: gender match and mismatch conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>M–M, e.g. recept na porosok ‘prescriptionM,NOM,SG for powderM,ACC,SG’</td>
<td>M (condition 1) / F (condition 2), e.g. byl prosročennym (wasM,SG expiredM,INSTR,SG) or byla prosročennoj (wasF,SG expiredF,INSTR,SG)</td>
</tr>
<tr>
<td>M–F, e.g. recept na maz’ ‘prescriptionM,NOM,SG for ointmentF,ACC,SG’</td>
<td>M (condition 3) / F (condition 4), e.g. byl prosročennym (wasM,SG expiredM,INSTR,SG) or byla prosročennoj (wasF,SG expiredF,INSTR,SG)</td>
</tr>
<tr>
<td>F–F</td>
<td>F (condition 5) / M (condition 6)</td>
</tr>
<tr>
<td>F–M</td>
<td>F (condition 7) / M (condition 8)</td>
</tr>
<tr>
<td>M–M</td>
<td>M (condition 9) / N (condition 10)</td>
</tr>
<tr>
<td>M–N</td>
<td>M (condition 11) / N (condition 12)</td>
</tr>
<tr>
<td>N–N</td>
<td>N (condition 13) / M (condition 14)</td>
</tr>
<tr>
<td>N–M</td>
<td>N (condition 15) / M (condition 16)</td>
</tr>
</tbody>
</table>

Table 2. Experimental conditions and sample stimuli from our study
Sample stimuli in conditions 1-4 given in Table 2 represent one set: two variants of the subject DP (one head and two different dependent DPs) and two variants of the predicate (matched or mismatched in gender with the subject). We constructed 48 such sets, 12 for each of the four conditions. All nouns in the subject DPs were singular. All attractors were in accusative, and their forms were homonymous with the nominative forms. All predicates consisted of the verb *byt’* ‘to be’ in the Past tense and a participle or an adjective in instrumental singular.\(^2\)

In addition to that, we constructed 100 fillers, also consisting of a predicate and a subject. Subjects contained singular or plural heads and adjectives or DP modifiers (not in accusative). Predicates were similar to the ones in target stimuli and did not agree with subjects in gender in one third of the cases.

Each participant saw only one stimulus from each set. Consequently, we had four experimental lists. The number of conditions was balanced for every list. All lists began with ten fillers, and then fillers and experimental items were presented in pseudo-random order, with the constraint that no more than two experimental items occur consecutively.

The experiment was run on a Macintosh computer using PsyScope software. In every trial, a fixation point appeared on the screen for 300 ms, then a predicate and a subject were presented for 800 ms each. After that participants saw a screen with a second counter, prompting them to respond faster. After they responded, the experimenter pressed a key, and the next trial started. All participants’ responses were tape-recorded. An experimental session lasted around seven minutes.

4.2 Results

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

1. Correct (C): the sentence is grammatical, the subject and the predicate provided as stimuli are repeated faithfully.

\(^2\) In Russian, participles, adjectives and nouns in predicates can appear either in nominative or in instrumental. Often only one variant is grammatical, but sometimes both are, or one is marginally acceptable, while the other is fine. Meaning nuances associated with them can be very subtle. It will suffice to say that we chose instrumental forms because, unlike nominative ones, they suited all our stimuli. But, if the subjects occasionally responded with nominative forms, we did not count this as a mistake.
2. Agreement error (A): the sentence is correct except for a gender agreement error.
3. Repetition error (R): the sentence is grammatical, but the subject or the predicate are repeated incorrectly.
4. A combination of repetition error and agreement error (R+A).
5. Incomplete response (I): the participant utters only a part of the sentence or says nothing at all.
6. A combination of incomplete response and agreement error (I+A): the sentence is incomplete, but a verb, a participle or an adjective was uttered and did not agree with the subject (cf. (7a-b)).

(7) a. Recept na maz’ byla...
   prescription_m.nom.sg for ointment_f.acc.sg was_f.sg

b. Recept na maz’ prosročennaja...
   prescription_m.nom.sg for ointment_f.acc.sg expired_f.nom.sg

The number of errors in each category is given in Table 3. In case of self-corrections, only the first variant was counted.

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct (C)</td>
<td>1018</td>
<td>71.2%</td>
</tr>
<tr>
<td>Agreement error (A)</td>
<td>61</td>
<td>4.3%</td>
</tr>
<tr>
<td>Repetition error (R)</td>
<td>111</td>
<td>7.8%</td>
</tr>
<tr>
<td>Repetition and agreement error (R+A)</td>
<td>9</td>
<td>0.6%</td>
</tr>
<tr>
<td>Incomplete response (I)</td>
<td>224</td>
<td>15.7%</td>
</tr>
<tr>
<td>Incomplete response and agreement error (I+A)</td>
<td>7</td>
<td>0.5%</td>
</tr>
</tbody>
</table>

Table 3. The distribution of responses in our study

The results of our experiment prove that gender agreement in Russian is subject to attraction. There were 77 agreement errors in categories ‘A’, ‘R+A’ and ‘I+A’, and only 13 out of them were not due to attraction (the difference is significant according to the chi-square test, p<0.001). Attraction errors were most frequent in predicate mismatch conditions (i.e. when the gender of the predicate provided as stimulus coincided with the gender of the attractor and not with the gender of the subject). However, 11 errors occurred in predicate match conditions, i.e. participants changed the correct gender of the predicate they were

*1* i.e. only one component of the predicate contains an agreement error.
provided with to an incorrect one due to attraction. It is also interesting that in three cases participants first uttered a correct sentence and then gave a second response, agreeing the verb with the attractor (since we always counted the first response, these cases were included in the category ‘C’).

Table 4 shows how errors of different types listed in Table 3 were distributed across experimental conditions. If two numbers are given, separated by the ‘+’ sign, the first represents agreement errors with attraction, and the second without it.

<table>
<thead>
<tr>
<th>Conditions 1/2</th>
<th>C</th>
<th>A</th>
<th>R</th>
<th>R+A</th>
<th>I</th>
<th>I+A</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/4</td>
<td>69/69</td>
<td>0/0</td>
<td>7/4</td>
<td>0/0</td>
<td>9/12</td>
<td>0/0</td>
</tr>
<tr>
<td>5/6</td>
<td>65/57</td>
<td>2+1/15+1</td>
<td>4/3</td>
<td>0/1</td>
<td>16/14</td>
<td>1/3</td>
</tr>
<tr>
<td>7/8</td>
<td>65/50</td>
<td>0/0+3</td>
<td>8/10</td>
<td>0/0</td>
<td>17/20</td>
<td>0/0</td>
</tr>
<tr>
<td>9/10</td>
<td>74/64</td>
<td>0/0+3</td>
<td>6/8</td>
<td>0/0</td>
<td>10/15</td>
<td>0/0</td>
</tr>
<tr>
<td>11/12</td>
<td>69/59</td>
<td>1/9</td>
<td>8/7</td>
<td>0/1</td>
<td>12/13</td>
<td>0/1</td>
</tr>
<tr>
<td>13/14</td>
<td>64/68</td>
<td>0+1/0+2</td>
<td>10/9</td>
<td>0/0</td>
<td>15/11</td>
<td>0/0</td>
</tr>
<tr>
<td>15/16</td>
<td>62/63</td>
<td>6+1/11</td>
<td>9/5</td>
<td>0/2</td>
<td>12/9</td>
<td>0/0</td>
</tr>
</tbody>
</table>

Table 4. The distribution of responses for each experimental condition

Now let us look at the distribution of attraction errors of all types (A, R+A, I+A) depending on the gender of the subject and the attractor.

<table>
<thead>
<tr>
<th>Conditions 3 and 4 (M–F)</th>
<th>Correct responses</th>
<th>Attraction errors</th>
<th>Other errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/4</td>
<td>119 (66+53)</td>
<td>22 (3+19)</td>
<td>39 (21+18)</td>
</tr>
<tr>
<td>7/8</td>
<td>116 (66+50)</td>
<td>11 (1+10)</td>
<td>53 (23+30)</td>
</tr>
<tr>
<td>11/12</td>
<td>136 (64+68)</td>
<td>12 (1+11)</td>
<td>42 (21+21)</td>
</tr>
<tr>
<td>15/16</td>
<td>125 (62+63)</td>
<td>19 (6+13)</td>
<td>36 (22+14)</td>
</tr>
</tbody>
</table>

Table 5. Agreement attraction errors across experimental conditions

Table 5 makes clear that the incidence of attraction errors is different for different gender combinations. In the pair ‘M–F’ (where the head of the subject is masculine and the attractor is feminine) there were

---

4 Due to a minor mistake, there are 85 responses in these conditions rather than 90.
significantly more errors than in the pair ‘F–M’ (p=0.045 according to the chi-square test). Badecker and Kuminiak obtained the same result in their experiment on Slovak. As for the ‘M–N’ and ‘N–M’ combinations, we had more errors in the latter case, but this difference was not significant, unlike in the Slovak study.

In the future, we plan to replicate Badecker and Kuminiak’s experiment where they compared ‘N–M’ and ‘N–F’ combinations. We are also going to run a parallel comprehension study: numerous previous experiments demonstrated that agreement errors that tend to occur more often in production cause bigger delays in reading times. This way we will be able to see whether the difference between Slovak and Russian was accidental or would be observed again.

5 General discussion and conclusions

Further experiments are necessary to formulate the final explanation of markedness effects in gender agreement. Nevertheless, some conclusions can already be made. The plural markedness effect observed in all previous studies on number agreement was explained as follows. Either singular was claimed to be the default value of the number feature, or this feature was claimed to be privative, i.e. only plural nouns were assumed to be marked for number. Even if some of these assumptions are correct, results from Slovak and Russian experiments show that markedness effects in agreement are not limited to such configurations. Even if some gender value, say, neuter, is considered to be default or zero, markedness effects arise not only when it is compared to feminine and masculine, but also when masculine is compared to feminine.

What can cause such markedness effects? In case of number, possible explanations cannot be teased apart because all factors point in one direction: singular is more frequent, is morphologically unmarked in many languages, is considered to be less complex cognitively etc. The situation with gender is less straightforward. Only Russian examples are discussed below. In some cases, we know that Slovak exhibits the same properties, but in the others we are not sure.

The relation between masculine and feminine is more or less clear. Masculine is more frequent and is used as the default option in cases like (8a-b), where the word každyj ‘each’ appears in masculine. Interestingly, (8b) sounds worse than (8a), and the same construction is ungrammatical
with inanimate nouns, although they can be used in such sentences without gender conflict (cf. (9a-b)). Moreover, masculine nouns are used to refer to groups of people of mixed or uncertain gender, or to an arbitrary member of such groups. This generalization is discussed by Yanovich (2013) who shows that it does not hold for animals or inanimate things.

(8) a. Мужчина и женщина каžдыj с’eli po manM.NOM.SG and womanF.NOM.SG eachM.NOM.SG atePST.PL preP.DISTR jabluku
appleACC.SG

b. Еж свин’ja каžдыj с’eli
hedgehogM.NOM.SG and swineF.NOM.SG eachM.NOM.SG atePST.PL po
preP.DISTR jabluku.
appleACC.SG

(9) a. *Дивan и кроват’ каžдыj стоили целое
sofaM.NOM.SG and bedF.NOM.SG eachM.NOM.SG costPST.PL wholeACC.SG
стоjание.
fortuneACC.SG

b. Кушетка и кроват’ каžдaja стоили целое
couchF.NOM.SG and bedF.NOM.SG eachF.NOM.SG costPST.PL wholeACC.SG
стоjание.
fortuneACC.SG

The situation with masculine and neuter is more complex. Neuter is the least frequent gender, but the default agreement in impersonal constructions uses neuter endings (cf. (10)). As for sentences like (11), not all speakers of Russian accept them, but those who do use the masculine form of každyj ‘each’. Constructing a good example with non-human animates is problematic, but this might be due to independent factors. The relevant neuter words (mlekopitajuščee ‘mammal’, životnoe ‘animal’, nasekomoe ‘insect’ etc.) tend to be abstract, and sentences combining two of them with the neuter form každoe ‘each’ also sound strange despite the absence of gender conflict.

(10) Светalo.
dawnPST.N.SG
‘It dawned.’
Now let us turn to previous studies dedicated to agreement errors in Russian (Fedorova 2004; Lorimor et al. 2008; Wilson & Nicol 1999; Yanovich & Fedorova 2006). Although they looked only at number agreement, Wilson and Nicol and then Yanovich and Fedorova demonstrated that the incidence of number errors depends on the gender of the head noun of the subject DP. Errors arise most often with neuter heads and least often with masculine ones. This correlates with the frequency of the three genders in Russian: masculine is the most frequent, and neuter is the least frequent.

We find these data very interesting because they show that the phenomenon of markedness is non-homogeneous, with different forms behaving as marked in different grammatical configurations or constructions. This means that markedness does not depend on one particular property of features, for example, on the representation of these features in the grammar or on the frequency of different feature values, but may reflect various properties in different cases. We will come back to this problem to analyze it in more detail once we have the results from our further experiments. It is also interesting to check whether markedness effects are more pronounced if all properties point into one direction, as in the case of number.

References


---

5 Ditja ‘child’ is somewhat archaic, so an archaic masculine noun *voin* is used instead of *mužčina ‘man’ (cf. (8a)) so that the sentence does not sound strange.


slioussar@gmail.com, fynjy8@mail.ru