Detecting Agreement Errors in Georgian: 
Implications for Predictive Parsing

Steven Foley and Matt Wagers
University of California Santa Cruz, Santa Cruz
Draft, March 2019

Abstract: While agreement is a well-studied phenomenon in psycholinguistics, little is known about the processing of typologically rarer types of agreement dependencies, like object–verb agreement. The present study investigates Georgian, a language with both subject–verb and object–verb agreement. In a speeded acceptability judgement experiment, participants were tasked with detecting errors in local-person (first- and second-person) subject and object agreement morphology. A Signal Detection Theory analysis reveals that participants are (i) more sensitive to agreement errors on verbs with canonical morphosyntactic properties than they are to errors on verbs with noncanonical properties, and (ii) more sensitive to errors in object agreement than they are to errors in subject agreement. Both results, we propose, follow from differing levels of certainty about the form of the sentence-final verb. As a consequence of Georgian’s split-ergative case system, the syntactic role of an argument is often ambiguous until the verb is encountered. Previous studies on Georgian sentence processing have identified a few simple biases that parsers use to navigate these temporary ambiguities. Applied to our stimuli, these parsing strategies conspire against verb forms with noncanonical properties and/or local-person subject agreement. The parser is less likely to predict these features, and thus less sensitive to errors involving them.

Keywords: agreement, case, Georgian, Signal Detection Theory, predictive parsing

1. Introduction

Argument–verb agreement dependencies are well studied in psycholinguistics, but most of the literature has focused on subject agreement (Bock & Miller 1991, Berg 1998, Badecker & Kuminia 2007, Nevins et al. 2007, Wagers et al. 2009, Molinaro et al. 2011, Dillon et al. 2013, Mancini et al. 2013, and many others). In contrast, there have been only a handful of studies investigating agreement dependencies involving objects, and to our knowledge these have nearly all focused on Basque (Zawiszewski & Friederici 2009, Díaz et al. 2011, Santesteban et al. 2013, Chow et al. 2018). Therefore, it’s an open question to what extent subject–verb agreement (S.AGR) and object–verb agreement (O.AGR) are processed with the same cognitive mechanisms, and to what extent morphosyntactic variation across languages can influence those mechanisms.

The present study broadens the typology by comparing S.AGR and O.AGR dependencies in Georgian. This language’s morphosyntactic properties make it an especially good testing ground. Georgian verbs can register both subjects and objects, but the mapping between agreement morphemes and the syntactic role of their controllers is not a straightforward one. In certain tenses, role–affix associations ‘invert’: affixes which usually register subjects will register objects, and vice versa. Similar dissociations between form and role are found in nominal case morphology, so the syntactic role of certain nouns will be ambiguous until encountering tense morphology on the verb. The confluence of these factors means that predicting what kind of agreement morphology will appear on a verb is far from trivial in this language.
Detecting Agreement Errors in Georgian: Implications for Predictive Parsing

Applying a Signal Detection Theory analysis to data from a speeded acceptability experiment, we find that Georgian speakers are more sensitive to errors involving the default morpheme–role mapping than ones involving the rarer ‘inverse’ mapping, and more sensitive errors involving O.AGR than S.AGR. The former finding is unsurprising, assuming that non-canonical morphosyntactic properties are more difficult to process and less likely to be predicted by the parser. The latter finding, though, is somewhat unexpected; O.AGR is intuitively more marked than S.AGR, and facts from typology (Moravcsik 1974, Baker 2008, Bobaljik 2008, Siewierska 2013), acquisition (Meisel & Ezeizabarrena 1996), and language impairment (Laka & Erriondo-Korostola 2001) are consonant with this intuition. So, if there is a processing asymmetry between the two dependencies, one might expect O.AGR to be the more difficult one. However, we are hesitant to conclude that the processing asymmetry we observe is rooted primarily in the syntactic role of the agreement controller. Taking into account parsing strategies previously observed during Georgian sentence processing (Skopeteas et al. 2012; cf. Foley & Wagers 2017; Lau et al. 2018, submitted), we instead propose that the sensitivity to agreement errors correlates with the parser’s degree of certainty about the form of the upcoming verb. Under this view, there need not be any inherent difficulty associated with S.AGR.

2. Background
2.1 Previous research

Differences between the processing of S.AGR and O.AGR have been investigated in several studies on Basque. Basque is a scrambling SOV language whose (auxiliary) verbs agree with all arguments. Its morphology is robustly ergative: intransitive subjects and direct objects receive one case value (absolutive), and transitive subjects another (ergative); verbal agreement affixes are likewise ergative-aligned (Laka 1996, Hualde & Ortiz de Urbina 2003, de Rijk 2008).

Across a series of ERP studies and one production experiment, a fairly clear picture emerges: S.AGR and O.AGR seem to be processed qualitatively differently, but it’s not obvious that one or the other is more difficult to process. For example, Zawiszewski & Friederici (2009), interested in how the controller or length of an agreement dependency might affect its processing profile, compared the ERPs elicited by S.AGR and O.AGR errors in SOV sentences (Experiment 1), and by O.AGR errors in either SOV or OSV sentences (Experiment 2). Ungrammatical stimuli had verbs with person and number values that did not appear on either argument. Their behavioral data (acceptability judgement accuracy and latency) indicate that participants are better at correctly rejecting agreement errors than at endorsing grammatical forms, and that their performance worsens in sentences with non-canonical word order (OSV). However, they observe no difference in performance between S.AGR and O.AGR errors. As for the electrophysiological results, the authors find that both S.AGR and O.AGR errors trigger an early negativity (which they label an N400) and a late positivity (a P600); the N400 was stronger for S.AGR errors, and the P600 stronger for O.AGR errors. The P600 is an ERP frequently occasioned by morphosyntactic violations (see Molinaro et al. 2011 for an overview), but the N400 — which is typically associated with difficulties in lexical–semantic integration or thematic role association — is a less canonical result of errorful agreement. This particular asymmetry, they suggest, implies that comprehenders are more sensitive to S.AGR in early stages of processing, and more sensitive to O.AGR in later stages. As for the effects of word order, the authors observe no significant differences between the ERPs elicited by SOV O.AGR violations and OSV ones; they do observe processing difficulty in grammatical OSV sentences, though, both at the sentence-initial absolutive argument (which,
Detecting Agreement Errors in Georgian: Implications for Predictive Parsing

notably, is not unambiguously a scrambled object; absolutive case can appear on intransitive subjects) and at the sentence-final verb.

Zawiszewski & Friederici (2009)’s N400–P600 pattern is replicated by Zawiszewski et al. (2016) (who look just at S.AGR), but not by Díaz et al. (2011). These latter authors also investigated ERPs elicited by S.AGR and O.AGR violations. They find that both kinds of errors elicit P600s, but they observe no evidence that the two kinds of agreement dependencies trigger ERPs of different magnitudes. Errors in O.AGR also trigger an early posterior negativity. Due to the time course of this potential (which surfaces between 200 and 300 ms after the onset of the ungrammatical verb), and to the fact that auditory rather than written stimuli were used in their experiment, Díaz et al. speculate that this early negativity is more likely to be an N200 than an N400. The N200 indexes violations of phonological expectations, and in O.AGR conditions, grammatical and ungrammatical agreeing auxiliaries are phonologically distinct by the second segment of the word (*dute vs. dituzte). If their interpretation is correct, then Basque parsers must make very specific predictions about the form of an upcoming auxiliary.

Besides the modality of their presentation, Díaz et al. (2011)’s stimuli depart from Zawiszewski & Friederici (2009)’s in another important way. Ungrammatical verb forms only ever had errors in number agreement, not person agreement. It may be that different phi-features are processed differently, such that only person agreement violations will elicit an N400. However, Nevins et al. (2007), who compare agreement violations across a number of features in Hindi, do not observe an N400 associated with ungrammatical person agreement.

A third ERP study on Basque agreement is Chow et al. (2018). They compare the processing profiles of transitive S.AGR (which cooccurs with overt O.AGR) and intransitive S.AGR (which does not), manipulating the grammaticality of number agreement with these two argument classes. Across both behavioral and electrophysiological measures, they find that agreement dependencies on verbs with different valencies are processed asymmetrically. Looking at acceptability judgement accuracy, their participants were more sensitive to errors in intransitive than transitive S.AGR. ERP findings are similar. Both kinds of agreement errors occasion a classic P600, but only ungrammatical agreement with intransitive (i.e., absolutive) subjects caused an early posterior negativity. This recalls Díaz et al. (2011)’s results, where agreement with direct objects (i.e., absolutive arguments) triggers early negativity. Chow et al. speculate that the differences in electrophysiological responses to absolutive and ergative agreement is related to the formal syntactic properties of the dependencies. According to Arregi & Nevins (2012), absolutive agreement in Basque is the result of feature transfer between an argument and the verb; ergative agreement, on the other hand, is actually a species of pronominal cliticization. Perhaps this syntactic difference manifests in processing as differing ERP patterns.

Finally, Santesteban et al. (2013) conduct a production experiment on Basque. Participants were presented with SO_ and OS_ preambles which they had to complete with an inflected auxiliary verb; different conditions manipulated the number features of the preamble arguments. Echoing Zawiszewski & Friederici (2009)’s behavioral results, Santesteban et al. find that non-canonical word orders (OS_ preambles) lead to more production errors, but they find no more mistakes in S.AGR conditions than in O.AGR conditions. However, they do observe that, when arguments have different number values, more ungrammatical S.AGR forms are produced in SO_ conditions, and more ungrammatical O.AGR forms are in OS_ conditions. This demonstrates that agreement
Detecting Agreement Errors in Georgian: Implications for Predictive Parsing

attraction (Bock & Miller 1991, et seq.) is alive and well in Basque, and that both S.AGR and O.AGR may be attracted (cf. Bhatia 2017 on Hindi).

2.2 Morphosyntactic properties of Georgian

Georgian is a South Caucasian language with a split-ergative case system and rich verbal agreement (Aronson 1990, Hewitt 1995). The dominant word order is SOV, but other orders are possible, especially when information structure is appropriately manipulated (Skopeteas & Fanselow 2010). Verbs agree with local-person (first- or second-person) arguments of all stripes, and also certain third-person arguments. Agreement affixes can be divided into two classes, which we refer to as ‘Set A’ and ‘Set B’. How these classes of morphemes are mapped to syntactic roles depends on factors including verb argument structure and tense. In the Direct Mapping, which obtains for most verbs in most tenses, subjects control Set A affixes, and objects control Set B affixes. Examples in (1) are representative of this pattern: local-person subjects trigger the Set A prefixes v− ‘1.A’ and Ø− ‘2.A’, and local-person objects trigger the Set B prefixes m− ‘1.B’ and g− ‘2.B’. These prefixes can compete with each other for realization in a single morphological slot (Béjar & Rezac 2009, Nevins 2011), but since there are no third-person Set A or Set B prefixes, this competition will not arise in the kinds of clauses we consider here, which only ever have one local-person argument.

(1) Direct Mapping: Set A morphemes (v−…) = S.AGR, Set B morphemes (m−…) = O.AGR
  a. me datv-s v-naxav. 1SG.(NOM) bear-DAT 1.A-see.FUT
     ‘I will see the bear.’
  b. šen datv-s Ø-naxav. 2SG.(NOM) bear-DAT 2.A-see.FUT
     ‘You will see the bear.’
  c. datv-i me m-naxavs. bear-NOM 1SG.(DAT) 1.B-see.FUT
     ‘The bear will see me.’
  d. datv-i šen g-naxavs. bear-NOM 2SG.(DAT) 2.B-see.FUT
     ‘The bear will see you.’

Contrasting with the Direct pattern is the Inverse Mapping, which is found in only dative subject constructions. These include clauses with normal transitive verbs in a few tenses (like the present perfect, typically used as a past evidential in Georgian), and also clauses with experiencer-subject verbs in any tense. Now, subjects control Set B morphemes, and objects control Set A morphemes. Examples in (2) illustrate.

(2) Inverse Mapping: Set B morphemes (m−…) = S.AGR, Set A morphemes (v−…) = O.AGR
  a. me datv-i m-naxavs. 1SG.(DAT) bear-NOM 1.B-see.PERF
     ‘I have [apparently] seen the bear.’
  b. šen datv-i g-naxavs. 2SG.(DAT) bear-NOM 2.B-see.PERF
     ‘You have [apparently] seen the bear.’
Detecting Agreement Errors in Georgian: Implications for Predictive Parsing

c. `datv-s  me  v-unaxivar.
bear-DAT  1SG(NOM)  1.A-see.PERF
‘The bear has [apparently] seen me.’

d. `datv-s  šen  Ø-unaxixar.
bear-DAT  2SG(NOM)  2.A-see.PERF
‘The bear has [apparently] seen you.’

Table 1 summarizes the Direct and Inverse Mappings.

<table>
<thead>
<tr>
<th></th>
<th>S.AGR</th>
<th>O.AGR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Direct Mapping</strong></td>
<td>Set A affixes</td>
<td>Set B affixes</td>
</tr>
<tr>
<td></td>
<td>(v–…)</td>
<td>(m–…)</td>
</tr>
<tr>
<td><strong>Inverse Mapping</strong></td>
<td>Set B affixes</td>
<td>Set A affixes</td>
</tr>
<tr>
<td></td>
<td>(m–…)</td>
<td>(v–…)</td>
</tr>
</tbody>
</table>

**Table 1: Summary of Georgian agreement mappings**

Georgian’s case system is as mercurial as its agreement system. Due to the language’s split ergativity (Harris 1981, Harris 1990, Nash 2017), the clause’s tense determines what case suffixes appear on the arguments of a normal transitive verb. For example, we find a nominative subject and a dative object in the future (3a); an ergative subject and a nominative object in the aorist (or perfective past; 3b); and a dative subject and a nominative object in the present perfect (3c). While this split-ergative pattern is expressed transparently by third-person arguments, it’s obscured by local-person arguments, since first- and second-person pronouns have syncretic nominative, ergative, and dative forms (4).

(3)  Tense-based split ergativity
   a. `datv-i  vepxv-s  naxavs.
bear-NOM  tiger-DAT  see:FUT.3SG
   ‘The bear will see the tiger.’

   b. `datv-ma  vepxv-i  naxa.
bear-ERG  tiger-NOM  see:AOR.3SG
   ‘The bear saw the tiger.’

   c. `datv-s  vepxv-i  unaxavs.
bear-DAT  tiger-NOM  see:PERF.3SG
   ‘The bear has [apparently] seen the tiger.’

(4)  Case syncretisms in local-person pronouns
   a. `me  vepxv-s  vnaaxay.
1SG(NOM)  tiger-DAT  see:FUT.1SG
   ‘I will see the tiger.’

   b. `me  vepxv-i  vnaaxe.
1SG(ERG)  tiger-NOM  see:AOR.1SG
   ‘I saw the tiger.’

   b. `me  vepxv-i  minaxaxs.
1SG(DAT)  tiger-NOM  see:PERF.1SG
   ‘I have [apparently] seen the tiger.’
Detecting Agreement Errors in Georgian: Implications for Predictive Parsing

These facts mean that an argument’s form is not necessarily a reliable cue to its syntactic role, at least not before encountering reliable tense cues. Ergative case on a noun entails that it is a subject, but nominative and dative case could both signal a subject or object. The situation is yet more vexing for local-person pronouns, which do not give any morphological cues to their syntactic role.

The effect of pronominal syncretisms on sentence processing has not yet been investigated in Georgian, but the effect of case ambiguities has. How Georgian parsers navigate the temporary ambiguity of nominative and dative is the question that drives Skopeteas et al. (2012). Their two acceptability judgement studies manipulate the word order (SOV or OSV) and case frame ($S_{NOM}>O_{DAT}$ or $S_{DAT}>O_{NOM}$) of simple declarative clauses. The arguments of the stimuli were displayed by together on a screen for 5000 ms, then the verb was added. Participants were prompted to judge the acceptability of the sentence. (While all experimental items were grammatical, there were both grammatical and ungrammatical fillers.) In both experiments, Skopeteas et al. find a main effect of case alignment on judgement response latencies, with $S_{DAT}>O_{NOM}$ clauses prompting slower responses. In the second experiment, they observe a main effect of word order, and also an interaction of word order and case frame: OSV clauses were judged more slowly than SOV ones, and $O_{NOM}S_{DAT}V$ clauses were the worst.

The authors infer from these results that Georgian parsers have two strategies for dealing with ambiguous case morphology. First, parsers are biased towards predicting that nominative DPs will be subjects, and dative DPs to be objects — even though these mappings aren’t necessary in the language. Second, parsers are inclined to predict that the linearly first argument in a clause is its subject — even though the language has flexible word order. The first bias is generally stronger than the second, explaining why $S_{DAT}O_{NOM}V$ stimuli were judged so slowly, but the two factors compound in $O_{NOM}S_{DAT}V$ clauses, explaining why they were the slowest in Experiment 2.

Suggestive evidence in favor of this parsing picture, or at least the dative–objecthood link, comes from Foley & Wagers (2017). In a reading time experiment on Georgian relative clauses, they find that sentence-initial human head nouns in the dative case are read more slowly than either nominative or ergative ones. Animal and inanimate nouns, on the other hand, did not exhibit a dative cost. Assuming (i) high-animacy nouns are more likely to be associated with subject position, (ii) low-animacy nouns are more likely to be associated with object position, and (iii) dative morphology is associated with object position, then the observed case–animacy interaction is entirely expected.

2.3 Current study

A cohesive set of results has emerged from the literature on Basque agreement processing: qualitatively different responses are elicited by $S_{AGR}$ and $O_{AGR}$ errors. However, it remains to be seen whether the patterns in Basque can be replicated, either in other languages or with different methodologies. That question is the impetus for the present study, a speeded acceptability task on Georgian. While Zawiszewski & Friederici (2009) and Diaz et al. (2011) did not find any significant differences in their acceptability judgement data attributable to an $S_{AGR}$–$O_{AGR}$ manipulation, we employ a methodology which should be more sensitive than a simple judgement task. Our participants also rate the confidence of their acceptability decision from one to three, resulting in a richer 1–6 judgement × confidence scale (Table 2). A Signal Detection analysis using
Detecting Agreement Errors in Georgian: Implications for Predictive Parsing

this scale will aim to tease apart differences in sensitivity to various kinds of agreement errors in Georgian (Macmillan & Creelman 2004).

<table>
<thead>
<tr>
<th>Unacceptable judgements</th>
<th>Acceptable judgements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very confident</td>
<td></td>
</tr>
<tr>
<td>Somewhat confident</td>
<td></td>
</tr>
<tr>
<td>Not confident</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Not confident</th>
<th>Somewhat confident</th>
<th>Very confident</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 2: Combined acceptability judgement–confidence rating scale

There are two dimensions of an agreement dependency in Georgian which may affect how it is processed. First is the controller. ERPs elicited by S.AGR and O.AGR errors in Basque are qualitatively different. Electrophysiological evidence does not clearly indicate that one or the other of S.AGR and O.AGR agreement is more difficult to process, but it has been observed that O.AGR is somewhat rarer cross-linguistically than S.AGR (Baker 2008:196; Siewierska 2013), and when it does occur, O.AGR is claimed to entail the existence of S.AGR elsewhere in the language (Moravcsik 1974, Bobaljik 2008). Moreover, O.AGR is acquired later than S.AGR (Meisel & Ezeizabarrena 1996), and it is more susceptible to aphasic impairment (Laka & Erriondo-Korostola 2001), at least in Basque. These facts hint at a processing disadvantage for O.AGR — a disadvantage that might manifest in our study as longer acceptability judgement latencies, lower confidence ratings, or less sensitivity to grammaticality errors in conditions involving O.AGR.

The second important factor is the morphosyntactic mapping of an agreement dependency. The inverse mapping is marked in many ways: it’s less frequent than the direct mapping; it involves a unique syntactic structure (Marantz 1989, et seq.); and the tenses that trigger it are semantically marked, conveying evidentiality. For these reasons, it’s reasonable to hypothesize that forms involving the inverse mapping will be more difficult to process than ones with the direct mapping.

What complicates the picture is the constellation of morphological ambiguities discussed in the previous section. In the typical out-of-the-blue clause encountered in an experimental setting, there will be at best suggestive evidence for either the syntactic role of a potential agreement controller or the morphosyntactic mapping employed by the verb. It’s not clear, then, what kinds of predictions the parser will make about an upcoming agreement dependency.

3. Methods

3.1 Participants

Fifty-five native speakers of Georgian (mean age = 25.6 years, standard deviation = 9.8) living in Georgia participated in this study. They were recruited online, and were paid 15 Georgian lari (GEL) for their participation.

3.2 Materials

Sixty-four item sets were constructed in a 2×2×2 design crossing local-person AGREEMENT dependency (S.AGR or O.AGR), morphosyntactic MAPPING (direct or inverse), and GRAMMATICALITY of agreement morphology (grammatical or ungrammatical). A sample itemset is given
Detecting Agreement Errors in Georgian: Implications for Predictive Parsing

in (6). The sentences consist of two DPs (the subject and object), a one-word adjunct (a locative or temporal PP, or a manner adverbial), and the clause-final verb. Itemsets were counterbalanced for word order (SOXV or OSXV), the type of local-person argument (1SG or 2SG), and tense (inverse verbs were always present perfect, but the direct verbs were either present, future, imperfective past, or conditional). All nouns in argument position referred to humans.

(5) a. **S.AGR, Direct Mapping**

\[\text{me} \quad bič-s \quad ťge-ši \quad \{\checkmark \text{vnaxav}, \quad \ast \text{mnaxavs}\}.\]

1SG(.NOM) boy-DAT forest-in see:FUT.1SG>3SG see:FUT.3SG>1SG

‘I will see the boy in the forest.’

b. **S.AGR, Inverse Mapping**

\[\text{me} \quad bič-i \quad ťge-ši \quad \{\checkmark \text{minaxavs}, \quad \ast \text{vunaxivar}\}.\]

1SG(.DAT) boy-NOM forest-in see:PERF.1SG>3SG see:PERF.3SG>1SG

‘I have [apparently] seen the boy in the forest.’

c. **O.AGR, Direct Mapping**

\[\text{bič-i} \quad \text{me} \quad ťge-ši \quad \{\checkmark \text{mnaxavs}, \quad \ast \text{vnaxav}\}.\]

boy-NOM 1SG(.DAT) forest-in see:FUT.3SG>1SG see:FUT.1SG>3SG

‘The boy will see me in the forest.’

d. **O.AGR, Inverse Mapping**

\[\text{bič-s} \quad \text{me} \quad ťge-ši \quad \{\checkmark \text{vunaxivar}, \quad \ast \text{mnaxavs}\}.\]

boy-DAT 1SG(.NOM) forest-in see:PERF.3SG>1SG see:PERF.1SG>3SG

‘The boy has [apparently] seen me in the forest.’

Note that the ungrammatical verb forms express all the features from both arguments, but they do so in the wrong way. For example, (5a) has a first-person subject, so the grammatical verb form has first-person S.AGR while the ungrammatical form has first-person O.AGR. Contrast this with the grammaticality manipulations of Zawiszewski & Friederici (2009), Diaz et al. (2011), and Chow et al. (2018), whose ungrammatical forms always introduced a feature that wasn’t present among the arguments, or lacked one that was. We believe our manipulation is a stronger test. Intuitively, ensuring that the observed features are associated with the right morphemes is a more difficult task that either recognizing that a verb form bears a novel phi-feature, or recognizing that it lacks an expected one.

These experimental materials were distributed to participants among eight lists using a Latin Square, such that each participant saw exactly one version of each item. They were distributed among 64 filler items of comparable complexity. The order of materials was randomized across participants. All materials were constructed by the first author in consultation with two native Georgian speakers.

3.3 Procedure

Subjects participated in the study online, via Ibex Farm (Drummond 2013). Upon accessing the experiment link, they input demographic information and gave informed consent. All materials and instructions were written in Georgian, using the Georgian script. The experiment began with four practice items to familiarize participants with the task. These were followed by the 128 experimental and filler sentences of the experiment proper. All stimuli were presented one word at a time, each word appearing in the center of the screen for 450 ms, with 100 ms between each
Detecting Agreement Errors in Georgian: Implications for Predictive Parsing

At the offset of the final word, participants had 3,500 ms to judge the acceptability of the sentence they just read. They pressed the F key to answer სცორი ‘correct’, and J for არასცორი ‘incorrect’. After that, they had unlimited time to rate how confident they were in that judgement, using a three-point scale: the 1 key indicated არა თავდაჯერებული ‘not confident’, and the 3 key indicated ჯალიან თავდაჯერებული ‘very confident’. After reading and judging all 128 sentences, the experiment concluded with an optional debriefing question.

3.4 Analysis

Response latencies were analyzed using linear mixed-effects regression; acceptability judgement accuracy was analyzed using logistic mixed-effects regression. The AGREEMENT conditions were coded by two coefficients using centered sum contrasts (S.AGR = +½ and O.AGR = −½). Likewise for the MAPPING conditions (direct = +½ and inverse = −½) and GRAMMATICALITY conditions (grammatical = +½ and ungrammatical = −½). Unless otherwise stated, maximal random effects structure was included (Barr et al. 2013). Models were estimated using the lme4 package in R (Bates et al. 2014). T-tests were calculated using Satterthwaite's method via the lmerTest package in R (Kuznetsova et al. 2017).

The Signal Detection Theory analysis derives empirical receiver operating characteristics (ROCs) from the judgement-confidence ratings (Macmillan & Creelman 2004). Sensitivity measures, including $d_a$, were calculated using the pROC package in R (Robin et al. 2011). We also used this package’s implementation of DeLong, DeLong & Clarke-Pearson (1988)’s non-parametric test for comparing the area under correlated ROC curves.

4. Results

There are three clear results. AGREEMENT, MAPPING, and GRAMMATICALITY all had significant effects on various behavioral metrics. Unsurprisingly, ungrammatical sentences were easier to reject than grammatical ones were to endorse, and inverse forms were more difficult to judge than direct ones. Somewhat surprisingly, though, O.AGR errors were easier to detect than S.AGR errors.

Descriptive statistics are given in Table 3. The accuracy of acceptability judgements was fairly high (similar to the accuracy in Lau et al. submitted’s study on Georgian), but markedly higher in ungrammatical conditions. For the most part, ungrammatical conditions were also judged more quickly. Confidence ratings were very high across the board, and experimental conditions had little impact on them. Finally, casual inspection of $d_a$, a robust measure of sensitivity related to the area under a ROC curve, suggests that participants were more sensitive to O.AGR and direct mapping conditions.

Table 4 reports the results of a logistic mixed-effects model on acceptability judgement accuracy. There are significant main effects of AGREEMENT, MAPPING, and GRAMMATICALITY, with conditions involving S.AGR, inverse mapping, and grammatical agreement being judged with significantly less accuracy than their counterparts.

Turning to response times, Table 5 gives the results of mixed effects modeling on latencies for correctly-judged trials. There is a main effect of GRAMMATICALITY, with grammatical conditions being judged slower. There was also a significant interaction of AGREEMENT and GRAMMATICALITY: the difference between latencies for S.AGR and O.AGR conditions is greater in
Detecting Agreement Errors in Georgian: Implications for Predictive Parsing

the ungrammatical conditions (where O.AGR conditions elicit much faster reaction times) than in the grammatical conditions. The easiest conditions to judge, in other words, were the ones with ungrammatical object agreement.

<table>
<thead>
<tr>
<th>Agr.</th>
<th>Mapping</th>
<th>Gram.</th>
<th>Accuracy</th>
<th>RT</th>
<th>Confidence</th>
<th>$d_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S.AGR</td>
<td>Direct (5a)</td>
<td>✓</td>
<td>66%</td>
<td>1072 (37)</td>
<td>2.73 (0.026)</td>
<td>1.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*</td>
<td>82%</td>
<td>1025 (34)</td>
<td>2.75 (0.028)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inverse (5b)</td>
<td>✓</td>
<td>60%</td>
<td>1023 (25)</td>
<td>2.72 (0.028)</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*</td>
<td>79%</td>
<td>1069 (35)</td>
<td>2.71 (0.029)</td>
<td></td>
</tr>
<tr>
<td>O.AGR</td>
<td>Direct (5c)</td>
<td>✓</td>
<td>65%</td>
<td>1054 (35)</td>
<td>2.74 (0.027)</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*</td>
<td>86%</td>
<td>910 (32)</td>
<td>2.76 (0.027)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inverse (5d)</td>
<td>✓</td>
<td>65%</td>
<td>1078 (34)</td>
<td>2.72 (0.027)</td>
<td>1.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*</td>
<td>82%</td>
<td>963 (33)</td>
<td>2.72 (0.028)</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3:** Summary of descriptive statistics. Given are acceptability judgement accuracy, mean judgement latency (in ms, with SE), mean confidence rating (1 = not confident, 3 = very confident; with SE), and $d_a$ (a measure of sensitivity).

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>SE</th>
<th>$z$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.2</td>
<td>0.12</td>
<td>10</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>AGREEMENT</td>
<td>−0.17</td>
<td>0.085</td>
<td>−2.0</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>MAPPING</td>
<td>0.20</td>
<td>0.085</td>
<td>2.3</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>GRAMMATICALITY</td>
<td>−1.1</td>
<td>0.086</td>
<td>−12</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>AGR:MAP</td>
<td>0.14</td>
<td>0.17</td>
<td>0.85</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>AGR:GRAM</td>
<td>0.15</td>
<td>0.17</td>
<td>0.91</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>MAP:GRAM</td>
<td>−0.10</td>
<td>0.17</td>
<td>−0.60</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>AGR:MAP:GRAM</td>
<td>0.34</td>
<td>0.34</td>
<td>1.0</td>
<td>&lt; 1</td>
</tr>
</tbody>
</table>

**Table 4:** Logistic mixed-effects model for response accuracy (this model only includes by-participant and by-item slopes; other models didn’t converge).
Detecting Agreement Errors in Georgian: Implications for Predictive Parsing

<table>
<thead>
<tr>
<th></th>
<th>β</th>
<th>SE</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1058</td>
<td>50</td>
<td>21</td>
<td>58</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>AGREEMENT</td>
<td>15</td>
<td>24</td>
<td>0.63</td>
<td>72</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>MAPPING</td>
<td>-8.8</td>
<td>23</td>
<td>-0.38</td>
<td>2287</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>GRAMMATICALITY</td>
<td>166</td>
<td>51</td>
<td>3.2</td>
<td>48</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>AGR:MAP</td>
<td>31</td>
<td>45</td>
<td>0.67</td>
<td>2280</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>AGR:GRAM</td>
<td>-169</td>
<td>54</td>
<td>-3.1</td>
<td>49</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>MAP:GRAM</td>
<td>69</td>
<td>45</td>
<td>1.5</td>
<td>2285</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>AGR:MAP:GRAM</td>
<td>28</td>
<td>91</td>
<td>0.30</td>
<td>2285</td>
<td>&lt; 1</td>
</tr>
</tbody>
</table>

Table 5: Linear mixed-effects model for response latency (in ms; model includes slopes only for agreement, grammaticality, and their interaction). Degrees of freedom, calculated by Satterthwaite’s method, are rounded to whole numbers.

Finally, we plot the results of a Signal Detection analysis in Figure 1. Pairwise comparison reveals a significant difference between the inverse-mapping curves (p < 0.05), but no significant difference between the direct-mapping curves (p = 0.44).

Figure 1: Fitted ROC curves, plotted in z-space. On the x-axis is the z-score of the miss rate (incorrectly rejected grammaticals); on the y-axis is the z-score of the correct rejection rate (correctly rejected ungrammaticals).
5. Discussion

There are two important findings from this experiment. First, we observed that inverse morphosyntactic mappings were harder to judge accurately. This confirms our prediction that, as inverse verb forms are so morphosyntactically and semantically marked, they should be harder to process. The semantic properties of the inverse forms used in this experiment — the present perfect, whose primary use is a past evidential — may be especially relevant. We hypothesize that, out of the blue, evidential clauses require processing effort to accommodate pragmatically; this cost is likely even greater for evidentials with local-person arguments. The intuitions of native speaker consultants support this suspicion. So, in order for inverse forms to put their best foot forward in a processing experiment like this, follow-up experiments may benefit from context sentences to facilitate the evidential semantics contributed by the present perfect.

The second important takeaway is that errorful O.AGR is easier to detect than errorful S.AGR. Taken at face value, this is surprising. Typological, acquisition, and language impairment facts would suggest that O.AGR is the harder type of dependency to process. However, it’s instructive, we think, to explore the data further. Recall that the itemsets in this experiment were counterbalanced for word order, half being SOV and half OSV. An exploratory Signal Detection Theory analysis partitioning data by word order, as presented in Figure 2, is illuminating. It shows that participants were more sensitive to S.AGR errors in OSV stimuli, and more sensitive to O.AGR errors in SOV stimuli.

![Figure 2: Fitted ROC curves, plotted in z-space, partitioned by word order.](image)

On the x-axis is the z-score of the misses rate (incorrectly rejected grammaticals); on the y-axis is the z-score of the correct rejections (correctly rejected ungrammaticals).

To understand why word order might affect processing in this way, we considered how morphological cues accrue incrementally in the stimuli. Examples in (7) schematize the SOV sentences. The string with the highest sensitivity, (7a), begins with a nominative-marked DP. According to Skopeteas et al. (2012), Georgian parsers initially posit that nominative morphology indicates subjecthood. If parsers assume that this nominative DP is the subject (eliminating the
Detecting Agreement Errors in Georgian: Implications for Predictive Parsing

possibility of inverse mapping, which is only found in dative subject constructions), and that the following subject local-person pronoun is the object, then they will have a very strong expectation for a verb with direct-mapping O.AGR. In support for this interpretation is how poorly participants perform in response to OSV stimuli with the same argument preamble (i.e., inverse S.AGR conditions, 9d).

(7) Schematized SOV stimuli
a. DP-NOM me XP {m-Verb.FUT, *v-Verb.FUT} [O.AGR, Direct; $d_a = 1.47$]
b. DP-DAT me XP {v-Verb.PERF, *m-Verb.PERF} [O.AGR, Inverse; $d_a = 1.26$]
c. me DP-DAT XP {v-Verb.FUT, *m-Verb.FUT} [S.AGR, Direct; $d_a = 1.02$]
d. me DP-NOM XP {m-Verb.PERF, *v-Verb.PERF} [S.AGR, Inverse; $d_a = 0.73$]

Now consider the about the worst SOV string (7d). We suspect that parsers are least sensitive to errors here because they form a very strong foiled expectation. According to Skopeteas et al. (2012)’s second parsing bias — link the first DP in a clause to subject position — parsers will be inclined to treat the local-person object as the subject. Encountering a nominative DP afterwards might lead parsers to infer that the syncretic pronoun is actually in the dative case, but this seems unlikely, given the general dispreference for positing dative subject constructions. An alternative is the infer that the pronoun is actually ergative, as in (8). Since verbs with ergative subjects always display the direct agreement mapping, this would lead to a strong expectation for direct-mapping S.AGR. Encountering the inverse verb in (7d) requires revising this parse, a process which may expend processing resources or simply fail.

(8) me datv-i v-naxe.
1SG(ERG) bear-NOM 1.A-see.AOR
‘I saw the bear.’

Turning to the OSV stimuli, participants were most sensitive to direct S.AGR violations (9a). If parsers have a strong expectation for dative DPs to be objects, even ones which are linearly first in the clause (as data from Skopeteas et al. 2012 and Foley & Wagers 2017 suggest), then it follows that they will anticipate direct-mapping S.AGR in this string. The other possible preambles (9c–d) all violate parsers’ expectations in one way or another: (9b) has the more appealing continuation (8) as a competitor; (9c) and (9d) both involve dative subjects, which the parser is in general loath to posit. This accounts for the lower sensitivity metrics associated with these strings.

(9) Schematized OSV stimuli
a. DP-DAT me XP {v-Verb.FUT, *m-Verb.FUT} [S.AGR, Direct; $d_a = 1.44$]
b. me DP-NOM XP {m-Verb.FUT, *v-Verb.FUT} [O.AGR, Direct; $d_a = 1.16$]
c. me DP-DAT XP {v-Verb.PERF, *m-Verb.PERF} [O.AGR, Inverse; $d_a = 1.12$]
b. DP-NOM me XP {m-Verb.PERF, *v-Verb.PERF} [S.AGR, Inverse; $d_a = 1.08$]

While the interpretation outlined here bears little on a central impetus of this experiment — namely, whether S.AGR or O.AGR is more difficult to process — it does present a principled and coherent picture of incremental sentence processing in Georgian that aligns well with the patterns found by Skopeteas et al. (2012). Four principles at work in Georgian seem to be the following.
(10) Parsing biases in Georgian
   a. Predict that DP-NOM will be a subject.
   b. Predict that DP-DAT will be an object.
   c. Avoid predicting a dative subject construction.
   d. Predict that a clause-initial DP will be a subject.

Future research, though, will be necessary to understand why these descriptive generalizations hold in this language. In particular, why might nominative and dative be associated with subject and object position, respectively? One possibility is that these are the most frequent mappings in the language. Impressionistically, though, Georgian’s split-ergative case system means neither nominative objects nor dative subjects are especially rare. Corpus research will be necessary to determine what kind of statistical discrepancy there really is.

6. Conclusion

This experiment has explored a poorly explored empirical terrain — agreement dependencies controlled by either subjects or objects — by applying Signal Detection Theory (SDT) to speeded acceptability judgement data. While behavioral measures in previous studies could not consistently reveal a processing difference between various agreement types (Zawiszewski & Friederici 2009, Díaz et al. 2011, Santesteban et al. 2013; cf. Chow et al. 2018), SDT has proven itself a more sensitive way to compare different agreement dependencies in Georgian. Our results show that Georgian speakers are most sensitive to (i) agreement dependencies involving ‘direct’ morphosyntactic mapping, and (ii) object–verb agreement dependencies. We attribute the first finding to the fact that the direct mapping is less marked and more frequent than the alternative ‘inverse’ mapping. As for the second finding, we conclude it stems from parsing principles specific to Georgian. Incremental morphosyntactic ambiguities are thick on the ground in the language, and previous psycholinguistic research suggests that comprehenders navigate these ambiguities with a small number of parsing biases (Skopeteas et al. 2012; Foley & Wagers 2017; Lau et al. 2018, submitted). Splitting the data up by word order, the patterns strongly suggest that parsers make specific predictions about the form of upcoming verbal agreement (cf. Díaz et al. 2011), and that when these expectations are dashed, it is more difficult to judge the acceptability of agreement morphology.

Acknowledgements

We thank audiences at UC Santa Cruz, UMass Amherst, the 2018 California Meeting on Psycholinguistics at the University of Southern California, and CUNY 32 (Boulder) for helpful discussion. Special thanks to Amanda Rysling for her feedback and suggestions, and to Irma Miminoshvili and Mariam Navadze for their assistance constructing stimuli and navigating logistics in Georgia. All errors are our own.

References

Detecting Agreement Errors in Georgian: Implications for Predictive Parsing


Detecting Agreement Errors in Georgian: Implications for Predictive Parsing


