Journal of Memory and Language 97 (2017) 208–244

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Sources of relative clause processing difficulty: Evidence from Russian

Iya K. Price *, Jeffrey Witzel

Department of Linguistics and TESOL, University of Texas at Arlington, 701 Planetarium Place, (Mailbox 19559), Arlington, TX 76019, USA

ARTICLE INFO

Article history:
Received 12 September 2016
Revision received 27 July 2017
Available online 29 August 2017

Keywords:
Sentence processing
Relative clauses
Russian
Self-paced reading
Eye tracking
Corpus

ABSTRACT

This study investigates the sources of processing difficulty in complex sentences involving relative clauses (RCs). Self-paced reading and eye tracking were used to test the comprehension of Russian subject- and object-extracted RCs (SRCs and ORCs) that had the same word-order configuration, but different noun phrase (NP) types (full NPs vs. pronouns) in the embedded clause. In both SRCs and ORCs, this NP intervened between the modified noun and the RC verb. A corpus analysis and acceptability rating experiment indicated different frequency/preference profiles for this word order depending on RC type and embedded NP type. In line with these profiles, processing difficulty was revealed early in the embedded clause for less frequent/dispreferred constructions. Later in the embedded clause, the processing of the RC verb was comparable for both SRCs and ORCs when the same number of NP arguments was available for integration. While there were no indications of an ORC penalty at or after this verb, late-stage comprehension difficulty was found for full-NP ORCs, but not for their pronominal counterparts, suggesting that similarity-based interference in combination with ORC structure influences the overall comprehension of these sentences. Taken together, these findings support a hybrid model under which independent sources of processing difficulty affect different stages of RC comprehension.

Introduction

During language comprehension, different types of information contribute to the interpretation of sentences. These include syntactic and semantic information as well as information about the frequency of occurrence of certain structures. One way to determine how these information types are used during real-time sentence comprehension is to investigate sentences that cause processing disruptions. Examples of such sentences are those that involve relative clauses (RCs) (1a-b). An RC is a subordinate clause that typically modifies a noun phrase (NP). Under standard syntactic analyses, RCs contain an extracted constituent that is linked with the modified NP (Heim & Kratzer, 1998). Example 1a shows a case in which the extracted constituent (__) is the subject, or a subject-extracted RC (SRC), while 1b provides an example of an object-extracted RC (ORC).

1. a. The reporter [that __ attacked the senator] admitted the error.
   b. The reporter [that the senator attacked __] admitted the error.

Generally, research has shown that ORCs are more difficult to comprehend than SRCs (e.g., Gibson, 1998, 2000; King & Just, 1991; Staub, 2010; Staub, Dillon, & Clifton, 2017; Traxler, Morris, & Seely, 2002). This is the case not only in English, but also in other languages, including Chinese (Lin & Bever, 2006; Vasishth, Chen, Li, & Guo, 2013, but see Hsiao & Gibson, 2003), Dutch (Mak, Vonk, & Schriefers, 2002), Hungarian (MacWhinney & Pleh, 1998), and Japanese (Miyamoto & Nakamura, 2003; Nakamura & Miyamoto, 2013). Understanding the nature of this disparity – and more generally, what makes some RCs more difficult to process than others – has the potential to shed light on fundamental properties of the language processing system, including how different sources of information contribute to online sentence comprehension.

A number of models of RC processing have been proposed. These accounts attribute processing costs in these sentences to different sources, including subject-object structural asymmetries (Clifton & Frazier, 1989; Hawkins, 1999; Lin & Bever, 2006; MacWhinney & Pleh, 1998; O'Grady, 1997; Townsend & Bever, 2001; Traxler, Williams, Blozis, & Morris, 2005; Traxler et al., 2002), syntactic expectations (Hale, 2001; Levy, 2008; MacDonald & Christiansen, 2002; Reali & Christiansen, 2007), memory costs (Gibson, 1998, 2000; Gordon, Hendrick, & Johnson, 2001, 2004; Gordon, Hendrick, Johnson, & Lee, 2006; Gordon, Hendrick, & Levine, 2002; Johnson, Lowder, & Gordon, 2011; King & Just, 1991; Lewis & Vasishth, 2005; Van Dyke & McElree,

* Corresponding author.
E-mail address: iya.khelm@uta.edu (I.K. Price).

http://dx.doi.org/10.1016/j.jml.2017.07.013
0749-596X/

2006), or a combination of these factors (Levy, Fedorenko, & Gibson, 2013; Staub, 2010; Staub et al., 2017). One way to test among these models is to examine where difficulty occurs during the incremental processing of RC sentences (Gibson, 1998, 2000; Gordon et al., 2001; Grodner & Gibson, 2005; Levy et al., 2013; Staub, 2010; Staub et al., 2017). Indeed, these models often predict processing costs at different points in the clause. For instance, expectation-based models predict processing difficulty at the first indication that the RC involves a less frequent construction, with ORCs generally occurring less frequently than SRCs. Memory-based accounts, on the other hand, predict additional processing time when arguments are integrated at the RC verb, which often occurs across greater distance and over more potentially interfering material in ORCs compared to SRCs. However, in many languages, such as English, it is difficult to test among these competing accounts because of word-order differences between SRCs and ORCs (as in the examples above). These disparities make it difficult to compare the relevant regions of these clauses and to tease out the nature of observed processing differences. Russian offers a potential solution to this problem because it has a relatively flexible word order. This makes it possible for Russian SRCs and ORCs to have the same lexical material in the RC in the same linear order (with case-marking distinguishing between the RC types), allowing for clearer comparisons of the processing of these clauses.

The present study took advantage of this word-order flexibility to examine potential sources of processing difficulty in RC sentences. Specifically, using self-paced reading (SPR) and eye tracking, this study investigated the online processing of Russian SRC and ORC sentences in which an NP argument intervened between the modified noun and the RC verb in both sentence types. This created a configuration in which the same number of NP arguments was available for integration at the RC verb in both SRCs and ORCs. This design thus allowed for an examination into whether RC processing difficulty relates to integration – in which case, there should be comparable memory costs for SRCs and ORCs when the number of integrated NPs is held constant – or to structural asymmetries – in which case, there should be particular processing difficulty for ORCs even under these conditions. Furthermore, the influence of syntactic expectations was investigated by using full NPs and pronouns in the embedded clause. As indicated by a corpus analysis and an offline acceptability rating experiment, these NP types are associated with very different word-order frequencies/preferences. This made it possible to investigate the role of expectation-based processing in these sentence types while again holding word order constant. In these ways, the present study attempted to assess different potential sources of processing difficulty in RC sentences and thus to test among competing models of the processing of these sentence types.

The first models of interest are those that attribute asymmetries in RC processing to structural differences. For example, the incremental minimalist parser theory (Lin & Bever, 2006) posits that processing difficulty for ORCs is due to differences in extracting from subject and object positions. This account holds that SRCs are easier to comprehend due to the shorter structural distance between the extracted constituent and its extraction site (see also Hawkins, 1999; O'Grady, 1997). Other accounts attribute this ORC penalty to a preference for analyzing the modified noun as the subject of the RC (Clifton & Frazier, 1989; Traxler et al., 2002, 2005), which results in the correct analysis for SRCs, but not for ORCs. Under such models, processing difficulty for ORCs relates to structural reanalysis. Another structure-based model is the perspective maintenance account (MacWhinney & Pleh, 1998), which explains processing difficulty for RCs in terms of shifts in the perspective of the subject. For example, in sentences with subject-modifying SRCs as in 1a, the main-clause subject (*the reporter*) is also the subject of the RC. In sentences with subject-modifying ORCs as in 1b, however, the perspective of the subject shifts because the main-clause subject is the object of the RC. Other models explain RC processing asymmetries in terms of word-order heuristics (Holmes & O'Regan, 1981; Townsend & Bever, 2001). Under these accounts, an initial parse provisionally assigns thematic roles by mapping the input onto canonical word-order templates. In the case of English, the ease of processing SRCs is attributed to the fact that they conform to the canonical SVO/agent-action-patient word order for this language.

Similar to models that suggest an important role for word-order-based heuristics are those that attribute processing costs in RCs to the frequencies of constructions in the language and the experience of language users. For instance, expectation-based theories (Hale, 2001; Levy, 2008) predict processing difficulty when unexpected constructions are encountered. Generally, in full-NP RCs – i.e., RCs in which the embedded NP is a full descriptive noun (e.g., *the senator* in 1a-b) – ORCs are less frequent than SRCs (Gordon & Hendrick, 2005; Reali & Christiansen, 2007). Under expectation-based models, processing difficulty for these ORCs is attributed to this frequency disparity. Support for such models comes from studies showing a switch in the SRC-ORC processing asymmetry when ORCs occur more frequently than SRCs. In English, this is the case with pronominal RCs – i.e., when the NP in the RC is a pronoun as in 2a-b (Reali & Christiansen, 2007; Roland, Dick, & Elman, 2007).

2. a. The reporter [that __ attacked you] admitted the error.
   b. The reporter [that you attacked __] admitted the error.

In line with this frequency difference, Reali and Christiansen (2007) showed that pronominal ORCs (2b) were easier to process than pronominal SRCs (2a). This was taken to indicate that expectation-based accounts make more accurate predictions for RC processing difficulty than structural asymmetry models, which cannot easily explain these findings.

Another class of accounts attributes RC processing differences to working memory effects. Specifically, these accounts explain RC processing costs in terms of the encoding, storage, and structural integration of the NPs involved in the clause. For example, according to the dependency locality theory (DLT) (Gibson, 1998, 2000; Warren & Gibson, 2002), integration costs are generally higher in ORCs (1b) because an additional discourse referent (*the senator*) is introduced before the dependency between the modified noun (*the reporter*) and RC verb (*attacked*) can be resolved. Similarity-based interference accounts (Gordon et al., 2001, 2002, 2004, 2006), on the other hand, posit that RC comprehension is hindered when the sentence requires similar NPs to be encoded and held in working memory before they can be integrated with the verb, as is often the case in ORCs (Gordon et al., 2006). Comparably, cue-based retrieval accounts explain this interference in terms of the argument requirements of the RC verb (Lewis & Vasishth, 2005; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006). Under these accounts, the integration of an RC verb like *attacked* with its arguments involves setting retrieval cues for NPs that can act as its agent and patient. RC processing difficulty arises when there is more than one candidate NP for these roles.

It is important to note that the absence of processing disruptions for pronominal ORCs, which expectation-based models attribute to their relatively high frequency, receives different explanations under these memory-based accounts. Under the DLT, processing difficulty for pronominal ORCs like (2b) is reduced because the referent of the RC pronoun is easily accessible, making