Brain & Language 108 (2009) 40–55

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Illusory licensing effects across dependency types: ERP evidence

Ming Xiang [a,*], Brian Dillon [b], Colin Phillips [b,c]

[a] Department of Linguistics, Harvard University, Boylston Hall, Third Floor, Cambridge, MA 02138, USA
[b] Department of Linguistics, University of Maryland, College Park, MD, USA
[c] Neuroscience and Cognitive Science Program, University of Maryland, College Park, MD, USA

ARTICLE INFO

Article history:
Accepted 9 October 2008
Available online 12 November 2008

Keywords:
Sentence processing
Semantic/pragmatic processing
Anaphora
Binding
Negative polarity items (NPI)
ERP

ABSTRACT

A number of recent studies have argued that grammatical illusions can arise in the process of completing linguistic dependencies, such that unlicensed material is temporarily treated as licensed due to the presence of a potential licensor that is semantically appropriate but in a syntactically inappropriate position. A frequently studied case involves illusory licensing of negative polarity items (NPIs) like *ever* and *any*, which must appear in the scope (i.e., c-command domain) of a negative element. Speakers often show intrusive licensing effects in sentences where an NPI is preceded but not c-commanded by a negative element, as in **The restaurants that no newspapers have recommended in their reviews have ever gone out of business*. Existing accounts of intrusive licensing have focused on the role of general memory retrieval processes. In contrast, we propose that intrusive licensing of NPIs reflects semantic/pragmatic processes that are more specific to NPI licensing. As a test of this claim, we present results from an ERP study that presents a structurally matched comparison of intrusive licensing in two types of linguistic dependencies, namely NPI licensing and the binding of reflexive anaphors like *himself*, and *herself*. In the absence of a potential licensor, both NPIs and reflexives elicit a P600 response, but whereas there is an immediate ERP analog of the intrusion effect for NPI licensing, no such effect is found for reflexive binding. This suggests that the NPI intrusion effect does not reflect general-purpose retrieval mechanisms.

1. Introduction

In order to parse effectively, the human sentence processor requires the ability to integrate material that may not be locally adjacent, due to the many different types of long-distance relations that are found in natural languages. Even a relatively unremarkable sentence like (1) could not be parsed without this ability:

(1) It's the old man that the boys like who watches the neighbor's cat.

Upon reaching the verb *watches* a listener must figure out what the subject of the verb is. In order to do this, he must assess the structure that he has already processed, query the structure, and decide upon an appropriate subject for the verb. In accounting for the parsing of long-distance dependencies, such as the dependency between the subject and verb in (1), several aspects of the process require explicit description. The first task is to characterize how structured representations are encoded in memory. The second is some sort of retrieval or querying operation, which is used to extract information from these representations. In recent years a number of proposals have aimed to explicitly characterize these memory-related processes in sentence processing (e.g., Gordon, Hendrick, Johnson, & Lee, 2006; Lewis & Vasishth, 2005). An important question for these approaches is how the proposed data structures are used to implement grammatical constraints on structural dependencies. For example, in (1) above there are several referring expressions in the sentence (e.g., *the old man, the boys, the neighbor*). However, the only licit subject of the verb *watches* is *the old man* by virtue of its structural configuration. Similar constraints arise in a number of domains in sentence processing, such as the interpretation of anaphora. These constraints rely on hierarchically structured representations that must be tracked during language processing in order for the parser to accurately single out grammatically licit antecedents. Accordingly, representations of structure in memory must be organized in such a way that retrieval operations can make appropriate decisions about acceptable or unacceptable targets.

However, in the parsing of long-distance dependencies it is still not clear how faithfully the human sentence processor implements grammatical constraints. A number of case studies have demonstrated faithfulness to grammatical constraints (e.g., Kazanina, Lau, Lieberman, Yoshida, & Phillips, 2007; Stowe, 1986; Sturt, 2003; Traxler & Pickering, 1996, among others), but there are also studies that suggest that the parser operates with somewhat less precision. A number of experimental results have demonstrated 'intrusion' effects in the completion of syntactic dependencies

* Corresponding author. Fax: +1 617 496 4447.
E-mail address: mxiang@fas.harvard.edu (M. Xiang).

0093-934X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2008.10.002

(Badecker & Straub, 2002; Drenhaus, Saddy, & Frisch, 2005; Gordon et al., 2006; Vasishth, Brüssow, Lewis, & Drenhaus, 2008). Intrusion occurs when speakers retrieve non-target or structurally inappropriate items that share some measure of similarity with the appropriate target. In the processing of syntactic dependencies, this may lead to the formation of spurious dependencies with grammatically illicit antecedents. This effect has been shown through decreased accuracy in judgment tasks (Drenhaus et al., 2005; Sturt, 2003), as well as decreased surprise at ungrammatical words in studies using eye-tracking (Vasishth et al., 2008) and self-paced reading (Pearlmutter, Garnsey, & Bock, 1999).

There are many factors that could underlie the seeming variation in the results in the literature. In this paper, we address two issues. First, previous studies have investigated intrusion effects in very different constructions, such as wh-questions, pronoun binding, subject-verb agreement, and licensing of negative polarity items (NPIs) such as *any, ever*, or *lift a finger*. Although discussions of intrusion effects tend to treat these dependencies as a homogenous class, careful linguistic examination of these constructions reveals many differences. Therefore, the establishment of these relations online could involve very different mechanisms that may be more or less susceptible to intrusion. This paper compares the intrusion profile of negative polarity item licensing and reflexive binding in a single experiment. As will be argued below, NPI licensing relies heavily on the semantic and pragmatic properties of the sentences in which the NPIs occur (Chierchia, 2006; Fauconnier, 1975; Giannakidou, 1998; Israel, 2004; Kadmon & Landman, 1993; Krifka, 1995). In contrast, reflexive binding reflects a process that more straightforwardly makes explicit reference to previously processed items and structure. Since these two constructions arguably involve long-distance relations that are established through different processes, a direct comparison of their intrusion profiles can provide insight into the specific nature of the errors that speakers make when building linguistic dependencies.

The second issue addressed here is methodological: we use ERP measures to obtain detailed time course information on the intrusion effect for different types of dependencies, which in turn can provide important clues to the exact source of intrusion effects.

2. Parsing long-distance dependencies

2.1. Accuracy in dependency completion

Much previous work shows that the human parser can use structural constraints to guide its operations, so that grammatically illicit structural relations are avoided. For example, a number of studies indicate that when completing a wh-dependency, comprehenders are sensitive to island constraints, such that the well-known 'active' gap construction mechanisms are inoperative in domains that span island boundaries (e.g., Phillips, 2006; Stowe, 1986; Traxler & Pickering, 1996). Similar sensitivity to linguistic constraints in the completion of syntactic dependencies has been shown for anaphora resolution as well. A particularly relevant example for this study is an eye-tracking study by Sturt (2003) on the processing of reflexives like *himself* and *herself*, which demonstrated a retrieval mechanism that appears to immediately respect grammatical constraints on antecedents for reflexives. Accurate retrieval in the resolution of anaphoric relations has also been noted in studies using several other constructions and methodologies, such as in experiments on the processing of backward anaphora (Cowart & Cairns, 1987; Kazanina et al., 2007), and studies of forwards anaphora using cross-modal lexical priming (Nicol & Swinney, 1989) and self-paced reading (Clifton, Kennison, & Albrecht, 1997).

This accuracy is not unanimously reported in all studies on anaphora resolution. In addition to an intrusion effect observed by Sturt in later eye-tracking and off-line measures, other results have suggested that inaccessible antecedents may influence the parser during the resolution of pronoun reference. For example, Badecker and Straub (2002) and Kennison (2003) have found that inaccessible antecedents exerted an influence on the resolution of reflexives and pronouns.

The results on intrusion effects in the processing of anaphoric dependencies offer a mixed picture of how accurately the retrieval system can target grammatically licit antecedents. One possible source for this discrepancy is that the conflicting findings reflect distinct stages of processing. None of the methodologies discussed above have probed the fine-grained time course of dependency completion. Self-paced reading times do not allow us to infer at what point in processing intrusion effects might arise. Although the distinction between early and late eye movement measures (e.g., Sturt, 2003) provides some measure of the time course of anaphora resolution, such measures only provide information at the granularity of individual fixations (approximately 250 ms per fixation).

2.2. Intrusion in NPI licensing

Some recent findings on negative polarity item (NPI) licensing have been presented as a strong challenge to the parser's ability to accurately retrieve structural information. At first glance NPI licensing shares important commonalities with the processing of anaphora, and with reflexives in particular. Both phenomena typically require a licensor in a structurally appropriate position. However, as we will demonstrate below, the relation between an NPI and its licensor differs in important ways from the relation between a reflexive and its antecedent, and these differences may impact how intrusion effects should be interpreted in each case.

NPIs, such as *any, ever*, or *lift a finger*, are lexical items that are licensed in specific semantic contexts (e.g., negation, conditionals, comparative constructions). For example, in (3a) the English NPI *ever* is licensed by the negatively quantified NP *no professor*, and the lack of the negative quantifier in (3b) leads to ungrammaticality.

(3) a. *No* professor will *ever* say that.
b. *A professor will *ever* say that.

Moreover, the licensing relation generally must also satisfy structural constraints. The negatively quantified NP in (3a) is in a position that c-commands *ever*. If the licensor fails to c-command the NPI this leads to unacceptability, as in (4), where the potential licensing NP is a relative clause subject that does not c-command *ever*.

(4) *The professor that *no* student likes will *ever* say that.

Drenhaus and colleagues (2005) tested native German speakers' sensitivity to the licensing conditions of the German NPI *jemals* ('ever'). Speeded acceptability judgments and ERP measures were collected for sentences like (5).

(5) *a.* *Kein* Mann, der einen Bart hatte, war *jemals* glücklich.
"No man who had a beard was ever happy."
b. *Ein Mann, der *keinen* Bart hatte, war *jemals* glücklich.
"*A man who had no beard was ever happy."
c. *Ein Mann, der einen Bart hatte, war *jemals* glücklich.
"*A man who had a beard was ever happy."

Both (5b) and (5c) are ungrammatical. Whereas (5c) has no licensor at all, (5b) contains a potential licensor *keinen Bart* ('no beard') that linearly precedes the NPI, but it is not in a structurally accessible position. Results of the speeded acceptability judgment task showed that responses in the inaccessible licensor condition were both slower and less accurate than in the other two conditions (70% accuracy in (5b) compared to 85% for (5a) and 83% for (5c)). ERP recordings showed that an N400 component was elicited in both ungrammatical conditions. However, the N400 for the inaccessible licensor condition was significantly reduced compared to the condition with no potential licensor at all. In another recent study of German NPIs, Vasishth et al. (2008) demonstrated faster reading times for sentences with an intrusive licensor than for sentences with no licensor, but only in later reading time measures such as second pass fixations. We have also found intrusive licensing effects in judgment and reading-time studies in English (Xiang, Dillon, & Phillips, 2006).

The critical finding from these previous studies is that the intrusive licensor decreased the effects of disruption associated with processing an unlicensed NPI. This suggests that in some proportion of trials participants treated these cases on a par with grammatical cases. Borrowing insights from the ACT-R cognitive architecture (Adaptive Control of Thought-Rational, Anderson, 2005; Anderson & Lebiere, 1998), Vasishth, Lewis and their colleagues (Lewis, Vasishth, & Van Dyke, 2006; Vasishth, Drenhaus, Saddy, & Lewis, 2005) have argued that the intrusion effect in NPI licensing reflects a general cue-based memory retrieval mechanism that is sensitive to partial cue overlap. For instance, in sentences like (5) the NPI *ever* is assumed to initiate a search for a lexical licensor that matches two cues: a semantic cue (i.e., a negation) and a structural cue (i.e., the licensor should appear in a position that c-commands the NPI). In (5a) the search yields a licensor that matches both cues, and in (5c) the search fails to find a match to either cue. In (5b) the search yields only a partially matched item—a semantically appropriate licensor that is in an inappropriate position. However, this partial cue overlap, coupled with stochastic fluctuation in base activation levels, may be sufficient for the structurally inaccessible licensor to be retrieved and treated as a licensor for the NPI *ever* on some percentage of trials.

This account of NPI intrusion essentially treats NPI licensing on a par with reflexive binding, as a structural dependency that is formed by retrieving an antecedent from previously processed material. It therefore makes the prediction that constructions involving similar types of structural dependencies should involve similar retrieval mechanisms, such that we should find a similar 'illusion of grammaticality' effect in both reflexive binding and NPI licensing. However, much work in linguistics suggests that NPI licensing and reflexive binding involve substantially different relations. Specifically, whereas reflexive binding involves dependencies between two specific syntactic constituents, many studies of NPI licensing have concluded that NPIs are licensed by the semantics and pragmatics of entire propositions, rather than involving dependencies with specific lexical items. This then raises the possibility that NPI intrusion is driven by a mechanism that is more specific to NPI licensing, rather than reflecting a general property of retrieval failure.

2.3. Semantic vs. syntactic dependencies

Whereas the intrusion effect has been robustly demonstrated for NPI licensing, studies of reference resolution processes show a more mixed pattern of results. The contrasting findings about NPI licensing and reference resolution may reflect the different methodologies that have been used across studies, or they may reflect fundamental differences in the two phenomena, despite the fact that psycholinguistic studies have often treated them as similar types of structural dependencies. If the two phenomena do indeed differ in fundamental ways, then the differences in previous results might reflect contrasts in how memory representation and retrieval operate in two qualitatively different domains.

In Vasishth et al. (2005), the characterization of intrusive NPI licensing is based upon the notion of a retrieval mechanism that seeks a licensor in memory. However, this view of NPI licensing may not be general enough to encompass the wide range of licensing environments that NPIs occur in. Although it is often convenient to talk about certain lexical items as specific licensors, this should not obscure the fact that the root of the licensing conditions does not lie in the lexical licensors themselves. There is a wide range of possible licensing environments, some of which do not contain an obvious lexical licensor, as shown by examples like (6).

(6) a. Has John ever cleaned his own dishes?
b. The reason one *ever* bothers to decant a wine is to leave the sediment [...] behind in the bottle. [*SouthWest Airlines Spirit* August 1994: 47] (Israel, 1998)

Vasishth et al. (2005) treat c-command as a critical licensing condition. Although c-command is often used to describe the general facts about NPI licensing, it is less straightforward to state it as a formal condition on NPI licensing. (7a) shows that when an NPI is within the semantic scope of a negation, it can be licensed, even though it is not syntactically c-commanded by the negation. (7b) shows that the same is not true of reflexive binding, which requires the antecedent to c-command the reflexive.

(7) a. *Nobody*'s mother has *ever* complained about his grades.
b. *Nobody$_j$'s mother criticizes himself$_j$ for his grades.

There is a broad consensus among current leading theories of negative polarity phenomena that NPI licensing reflects an interaction between the lexical properties of NPIs and the semantics and pragmatics of entire propositions (e.g., Chierchia, 2006; Fauconnier, 1975; Giannakidou, 1998; Israel, 2004; Kadmon & Landman, 1993; Krifka, 1995; Ladusaw, 1992), rather than a direct structural relation between the NPI and a licensing item. We discuss the specifics of such accounts in more detail in the Discussion section.

Reflexive binding, on the other hand, is more clearly identifiable as a process that links one lexical item with another in a structural configuration, regardless of whether this is characterized in syntactic (Chomsky, 1981) or semantic terms (e.g., Jackendoff, 1992; Steedman, 1997; Van Valin & LaPolla, 1997), without recourse to more intricate semantic or pragmatic properties of the proposition and the discourse (but see Zribi-Hertz, 1989, for a different view). It is important to mention that a separate class of reflexives, known as logophors, does demonstrate sensitivity to discourse relations (Clements, 1975; Pollard & Sag, 1992; Reinhart & Reuland, 1993; Runner, Sussman, & Tanenhaus, 2006; Sells, 1987). However, this discussion focuses on non-logophoric reflexive binding relations, which show no such sensitivity. Since the reflexive binding relation is more easily classified as an item-to-item dependency it may be a better candidate for testing the predictions of a retrieval-based model such as ACT-R. In contrast, such an account seems less attractive for NPIs, which display sensitivity to a wide range of other constraints and do not always require the presence of overt licensors.

2.4. The present experiment

In light of the evidence for qualitatively different representations underlying NPI licensing and reflexive binding, we

hypothesized that the NPI intrusion effect might reflect mechanisms more specific to NPI licensing, and might not extend to the domain of reflexive binding. Existing findings on intrusion effects in sentence processing neither confirm nor disconfirm this hypothesis, since no previous studies have presented a side-by-side comparison of the two types of dependencies. In addition, existing findings on intrusion in anaphor resolution provide relatively coarse-grained time course information. In this study, we compared the intrusion effect on NPI licensing and reflexive binding in the same structural configurations and in the same experiment, using ERP recordings, which offer a fine-grained and continuous measure of brain activity.

ERPs have several attractive properties. The first, and most germane to the question at hand, is their temporal precision. They can track scalp voltages associated with cognitive processes with millisecond precision, allowing us to assess the relative time course of the intrusion effect. Furthermore, studies of language processing using ERPs have described a number of temporally and topographically distinct components elicited by different types of linguistic material. For example, syntactically well-formed but semantically anomalous input elicits the N400 response (Kutas & Hillyard, 1980), a central negativity that generally peaks around 400 ms after the anomalous word. Syntactic and morphological anomalies and garden paths characteristically evoke the P600 component (Friederici, Pfeifer, & Hahne, 1993; Hagoort, Brown, & Groothusen, 1993), a posterior positivity that canonically appears around 600 ms post-stimulus, although it shows substantial latency variation. Some cases of syntactic or morphological anomaly also elicit an (early) Left Anterior Negativity ((E)LAN) response (Coulson, King, & Kutas, 1998; Friederici et al., 1993; Hagoort, Wassenaar, & Brown, 2003; Lau, Stroud, Plesch, & Phillips, 2006; Neville, Nicol, Barss, Forster, & Garrett, 1991), with latencies of 150–500 ms.

Precise time course information on the intrusion pattern may also prove critical in understanding the nature of the intrusion effect. If effects of intrusive licensing emerge in the ERP record at a longer latency than do effects of ungrammaticality, such that the response to intrusive licensing conditions is initially identical to the response to conditions with no licensor, then this would suggest that the intrusive licensing effect reflects a later stage of processing, perhaps as a result of error-driven repair processes. On the other hand, if the intrusive licensing effect appears in the ERP record as soon as any effect of ungrammaticality, then we may conclude that intrusion impedes the ability of speakers to detect grammatical errors, making a true 'illusion of grammaticality'. This study used ERP measures to test for the existence and detailed time course of intrusive licensing effects in reflexive binding and NPI licensing.

This study also goes beyond previous studies of intrusive licensing by comparing the effects of different NPI licensors (see Drenhaus, Blaszczak, & Schütte, submitted for publication, for comparison of multiple licensors in fully grammatical contexts). Previous studies of intrusive NP licensing have focused on the determiner *no/kein*. Since overt negation is a stereotypical licensor for NPIs, it is possible that intrusive licensing might reflect the high cooccurrence frequency between *no* and *ever*. An analysis of cooccurrence frequencies in the English Gigaword corpus (Graff, 2003) revealed that transparent negation (e.g., *not*, *no*, *didn't*, etc.) accounts for 19% of the licensing environments for *ever*, and of these the negative quantifier *no* alone accounts for 9% of cases. In order to address the possibility that intrusive NPI licensing might reflect frequent lexical associations, we tested both the highly frequent licensor *no* and the less common licensor *few* (2% cooccurrence frequency with *ever*).

3. Experimental materials and methods

3.1. Participants

Thirty-four members of the University of Maryland community participated in the ERP study. Six participants were excluded due to a high proportion of trials that needed to be rejected (more than 30% of experimental trials). The remaining 28 participants (17 females) included were aged 18–29 with a mean age of 20.1. All participants were healthy, native speakers of standard American English with no history of neurological disorder, and all were strongly right-handed based on the Edinburgh Handedness Inventory (Oldfield, 1971). All participants had normal or corrected-to-normal vision, gave informed consent and were paid $10/h for their participation, which lasted around 3 h, including set-up time.

3.2. Design and materials

The experimental materials consisted of two sets of items, one set for the test of NPI licensing and another for the test of reflexive binding. Both sets of materials included acceptable sentences with licensors in structurally accessible positions and two types of unacceptable sentence. The first type of unacceptable sentence contained no potential licensor for the NPI or reflexive, and the second type of unacceptable sentence contained a potential licensor in a structurally inappropriate position. All target items were sentences in which a subject NP was modified by an object relative clause. Structurally accessible licensors always appeared in the main clause subject position, and structurally inaccessible licensors always appeared as the subject of the relative clause, from which position they failed to c-command the NPI or reflexive in the main clause.

3.2.1. NPI materials

The NPI materials consisted of 150 sets of 5 items, which varied in terms of the presence, structural location and type of a potential NPI licensor. All items consisted of a subject NP modified by an object relative clause and followed by a main clause predicate that contained the NPI *ever*. The two grammatical control conditions had NPI licensors ('no' and 'very few') in main clause subject position. The two intrusive conditions included the same potential NPI licensors, but in the subject position of the relative clause. In the no licensor condition the NPI licensor was replaced with the quantifier *most*, which retains the quantificational force of the sentences used in the other conditions but fails to license NPIs. We chose to use the NPI licensor 'very few' instead of the shorter 'few' in order to increase the naturalness of the sentences and to reduce the risk of confusion with 'a few', a complex determiner that does not license NPIs. The relative clause verb was always followed by a four-word modifier phrase (such as a temporal adverb) that served to more clearly mark the right edge of the relative clause. The critical word, the NPI *ever*, appeared immediately following the main clause auxiliary verb. The NPI *ever* always appeared as the 14th or 15th word in the sentence. An example set of NPI items is given in Table 1, showing all five conditions. The potential licensing element and the NPI are italicized.

3.2.2. Reflexive materials

The reflexive materials consisted of 90 sets of 3 items, which varied in the presence and location of an NP that matched the reflexive in stereotypical gender. All items consisted of a subject NP modified by an object relative clause and followed by a main clause predicate that contained a 3rd person singular reflexive (*himself/herself*). The head of the relative clause was a noun that was strongly gender-stereotyped (e.g., *hairstylist* for female, and

Table 1
Sample set of items for the five conditions involving negative polarity items (NPIs).

Grammatical	{*No*/*Very few*} restaurants that the local newspapers have recommended in their dining reviews have *ever* gone out of business
Ungrammatical (intrusive licensor)	The restaurants that {*no*/*very few*} local newspapers have recommended in their dining reviews have *ever* gone out of business
Ungrammatical (no licensor)	Most restaurants that the local newspapers have recommended in their dining reviews have *ever* gone out of business

butcher for male). Previous studies have shown that gender-stereotyped nouns cause comprehender to make an early commitment to the gender of the referent, and that this commitment is strong enough to cause processing disruption upon reading a mismatching reflexive (Osterhout, Bersick, & McLaughlin, 1997; Sturt, 2003). The subject of the relative clause was a gender-specific common name (*Mary, Paul*). The reflexive items were counterbalanced, such that 45 items had subject nouns that were stereotypically male, and 45 items had subject nouns that were stereotypically female. In the acceptable control condition the reflexive was consistent with the gender stereotype of the main clause subject. In the two unacceptable conditions the reflexive violated the gender stereotype of the main clause subject (e.g., *The butcher burnt herself*). However, the two stereotype violation conditions differed in terms of the gender of the relative clause subject, which matched the gender of the reflexive in the intrusive condition and mismatched the gender of the reflexive in the no antecedent condition. Since the reflexive materials contained only apparent ungrammaticality, based on stereotypes, we refer to the conditions as *congruent*, *incongruent*, or *intrusive*, rather than in terms of grammaticality. A sample set of items is given in Table 2. The congruency manipulations in the reflexive conditions allows for closer comparison with previous studies of the processing of reflexives (Osterhout et al., 1997; Sturt, 2003), and one ERP study that directly compared stereotypical gender violations with simple gender violations found parallel effects in the two cases. The use of stereotype violations also made it more feasible to use comprehension questions, since the sentences were ultimately semantically coherent. We further discuss the impact of manipulating congruency rather than grammaticality below. An example set of reflexive materials is shown in Table 2.

Note that although the NPI materials and reflexive materials were structurally parallel, they were not lexically matched. This was unavoidable, given the design constraints in the two sets of conditions. In addition, there was a systematic difference between the two sets of materials in terms of the number of words between potential licensors and the dependent elements. Reflexives appeared 7 words after intrusive licensors, whereas NPIs appeared 11 words after intrusive licensors. The difference arose from the need to make each set of materials as natural as possible. Previous studies of NPI processing suggest that the intrusive licensing effect is stable across varying numbers of words between the NPI and the potential licensor, and hence this difference is unlikely to be responsible for any qualitative differences in results between the two types of materials.

3.2.3. Design

The experimental materials consisted of a total of 8 conditions (5 NPI conditions, 3 reflexive conditions), all with similar structures. Each participant saw 150 NPI sentences, 90 reflexive sentences, and 210 filler sentences, for a total of 450 sentences. The 150 sets of NPI items were distributed across 5 lists in a Latin Square design, and the 90 sets of reflexive items were distributed across 3 lists in a similar fashion. Among the 210 filler items, 150 had similar structures and used similar determiners to the NPI conditions, but without the NPI *ever*. The rationale for making the fillers similar in this way was to minimize the possibility that participants might develop strategic expectations for the NPI *ever* upon encountering one of the determiners used in the target sentences. The remaining 60 fillers consisted of sentences that included reflexives and names in structures of similar complexity to the target items, but with only one referent in each sentence. All of these fillers were grammatical, and they were included in order to mask the target items. All materials are available online at www.ling.umd.edu/colin.

Table 2
Sample set of items for the three conditions involving reflexives. The reflexive, the structurally correct antecedent and the intrusive antecedent are shown in italics.

Congruent	The tough *soldier* that Fred treated in the military hospital introduced *himself* to all the nurses
Intrusive	The tough *soldier* that *Katie* treated in the military hospital introduced *herself* to all the nurses
Incongruent	The tough *soldier* that Fred treated in the military hospital introduced *herself* to all the nurses

The 5 lists of NPI materials and the 3 lists of reflexive materials were combined to form 15 composite lists. Each of these lists contained 90 ungrammatical items from the NPI target items, a further 60 items that violated gender stereotype expectancy, and 300 acceptable sentences, for an overall 2:1 acceptable to unacceptable ratio.

After each experimental item, the participant was either presented with a comprehension question or was prompted to press a button to continue. Two-hundred and forty sentences in each list contained a comprehension question, and these were equally distributed across yes and no answers and across filler and target items. Each participant answered 60 questions for NPI items and 60 questions for reflexive items.

3.3. Procedure

Participants were comfortably seated in a dimly lit testing room around 100 cm in front of a computer monitor. Sentences were presented one word at a time in black letters on a white screen in a 30-point Times New Roman font. Each sentence was preceded by a fixation cross. Participants pressed a button to initiate presentation of the sentence, which began 1000 ms later. Each word appeared on the screen for 300 ms, followed by 200 ms of blank screen. The last word of each sentence was marked with a period, and 1000 ms later either a comprehension question appeared or an instruction appeared to prompt the participant to press a button to continue. Participants were instructed to read the sentences carefully without blinking and to indicate their answer with a button press on trials that included a question. Feedback was provided for incorrect responses. Each experimental session was preceded by a 4 trial practice session that included unrelated, grammatical sentences to introduce the methodology. Participants received feedback and were able to ask clarification questions about the task at that time. The experimental session was broken up by at least three break periods, and participants were able to request additional breaks at their discretion.

3.4. EEG recording and analysis

EEG was recorded from 28 Ag/AgCl electrodes, mounted in an electrode cap (Electrocap International): midline: Fz, FCz, Cz, CPz, Pz, Oz; lateral: F3/4, F7/8, FC3/4, FT7/8, C3/4, T7/8, CP3/4, TP7/8, P4/5, P7/8, O1/2. Recordings were referenced online to the left

mastoid. Additional electrodes were placed on the left outer canthus, and above and below the left eye to monitor eye movements. EEG and EOG recordings were amplified and sampled at 1 kHz using an analog bandpass filter of 0.1–70 Hz. Impedances were kept below 5 kΩ at all times.

Data from the NPI and reflexive conditions were analyzed separately, due to the different conditions, different critical words, and different syntactic and ordinal positions of the critical words. All analyses are based upon grand averages of 1200 ms intervals surrounding the critical NPI or reflexive, consisting of a 200 ms pre-stimulus interval and a 1000 ms post-stimulus interval. Data from both sub-experiments were analyzed based upon a 200 ms pre-stimulus baseline correction. However, the data from the NPI sub-experiment was also reanalyzed based on a 200 ms post-stimulus baseline interval. As discussed further below, this additional analysis was used in order to more clearly assess critical effects at the 600–800 ms interval. The choice of baseline interval did not alter the results qualitatively. Trials with ocular and other large artifacts were rejected based on visual screening. Among the 28 participants included in the analysis the total rejection rate in the NPI conditions was 13.0%, ranging from 12.0% to 13.8% across conditions, and the total rejection rate in the reflexive conditions was 13.1%, ranging from 12.2% to 15.1% across conditions. A 10 Hz low-pass filter was applied to the grand average ERPs for presentation purposes, however all analyses were performed on unfiltered data. ANOVAs were calculated based on mean voltages within a series of 200 ms time intervals that allowed continuous tracking of the evolution of ERP responses elicited by the target word (0–200 ms, 200–400 ms, 400–600 ms, 600–800 ms, 800–1000 ms).

For statistical analyses, six regions of interest (ROIs) were initially defined, consisting of three electrodes at each ROI: left anterior (FT7, F3, FC3), midline anterior (FZ, FCZ, CZ), right anterior (F4, FC4, FT8), left posterior (TP7, CP3, P3), midline posterior (CPZ, PZ, OZ), and right posterior (CP4, P4, TP8). An omnibus ANOVA comparing the factors *condition*, *laterality*, and *posteriority* was performed. No statistical analyses revealed significant main effects or interactions involving the laterality factor, and so this factor was removed from further analysis. Thus, all subsequent analyses instead used the following three ROIs: posterior (O2, OZ, O1, P3, PZ, P4), central (CP3, CPZ, CP4, C3, CZ, C4), and anterior (FC3, FCZ, FC4, F3, FZ, F4). The primary ANOVA reported here included the within-subjects factors condition (5 levels for NPIs, 3 levels for reflexives), and region (3 levels), with follow-up analyses based on planned comparisons between pairs of conditions. In order to assess potential differences between the two NPI licensors a second ANOVA crossed the factors *licensor* (*no* vs. *few*) and *position* (*accessible* vs. *intrusive*), excluding the ungrammatical no licensor condition. This ANOVA was performed for each ROI individually. All *p*-values reported below reflect application of the Greenhouse–Geisser correction where appropriate to control for violations of the sphericity assumption (Greenhouse & Geisser, 1959), together with the original degrees of freedom.

4. Results

4.1. Negative polarity item licensing

4.1.1. Comprehension questions

Overall average accuracy for trials that included a comprehension question was 89.4%. Within the NPI materials average accuracy was 86.4%, with condition means of 89.8% (*SD* 10%) for grammatical *no*, 92.2% (*SD* 11%) for grammatical *very few*, 77.4% (*SD* 18%) for intrusive *no*, 77.4% (*SD* 12%) for intrusive *very few*, and 94.9% (*SD* 7%) for the ungrammatical *no licensor* condition. Paired-sample *t*-tests showed that participants were significantly less accurate on the intrusive conditions than on the grammatical and the no licensor conditions. This was true both for the intrusive *no* condition (vs. grammatical *no*, $t_1(27) = 3.53$, $p < .01$; $t_2(59) = 2.8$, $p < .01$; vs. no licensor, $t_1(27) = 6.15$, $p < .001$; $t_2(59) = 2.1$, $p < .05$) and for the intrusive *very few* condition (vs. grammatical *very few*, $t_1(27) = 5.68$, $p < .001$; $t_2(59) = 4.7$, $p < .001$; vs. no licensor, $t_1(27) = 7.56$, $p < .001$; $t_2(59) = 5.4$, $p < .001$). Additionally, participants responded significantly less accurately in the grammatical *no* condition than in the no licensor condition ($t_1(27) = 2.64$, $p < .05$; $t_2(59) = 2.2$, $p < .05$). No other differences between conditions were reliable.

Although previous studies have shown that intrusive NPI licensors give rise to lower accuracy rates in acceptability judgment tasks, it is nevertheless striking that the same pattern was found here in comprehension question accuracy rates. Since the intrusive and no licensor conditions are ungrammatical, it was unclear to what degree speakers would converge on a coherent semantic representation for the entire sentence. Therefore, the comprehension questions for these items probed only the relative clause material, and in similar ways for the intrusive and no licensor conditions. The fact that participants nevertheless encountered greater difficulty in the intrusive conditions than in the no licensor condition suggests that the mere presence of a negative quantifier inside a relative clause might have led to increased processing difficulty. If this increased difficulty remained unresolved at the end of the relative clause, then it could have overlapped with ERP responses to the NPI, a consideration that becomes relevant in the next sections.

4.1.2. Event-related potentials

Figs. 1 and 2 display grand average ERPs to the NPI *ever* for the 18 electrodes that were included in the statistical analyses. For each potential licensor, *no* and *very few*, the grammatical and intrusive licensor conditions are plotted against the no licensor condition. Visual inspection revealed that the ungrammatical no licensor condition elicited a larger posterior positivity relative to the other two conditions, with an onset around 600 ms. Visual inspection also revealed that before 600 ms the ungrammatical condition did not diverge from the grammatical condition. However, the intrusive condition exhibited a larger negativity relative to the other two conditions that began very early at some electrodes and was more prominent in posterior regions.

Table 3 reports the results from the omnibus ANOVA that included all five conditions and three regions. The ANOVA showed a significant main effect of condition at the 200–400 ms interval and the 600–800 ms interval, and a marginally significant effect of condition at the 800–1000 ms interval. More interestingly, there was also a significant condition × region interaction in both the 600–800 and 800–1000 ms intervals. To assess the source of this result, two sets of planned comparisons were performed within each ROI. These comparisons separately compared the two grammatical conditions to their respective intrusive conditions, as well as each of these conditions to the no licensor condition. The results of these two analyses are shown in Tables 4 and 5. The average mean amplitude and the standard errors are presented in Table 6 for the last two time windows, where the P600 effects were observed.

Although visual inspection of Fig. 2 suggests a larger negativity in the no licensor condition relative to the grammatical condition in a latency range centered of around 250–350 ms, this difference did not reach significance at any ROI (posterior, $F(1,27) = 1.7$, $p > .2$; central, $F(1,27) = 2.3$, $p > .1$; anterior, $F(1,27) = 2.8$, $p > .1$).

The results of the pairwise comparisons showed that whereas the ungrammatical no licensor condition elicited a P600 relative to each of the two grammatical control conditions, the two

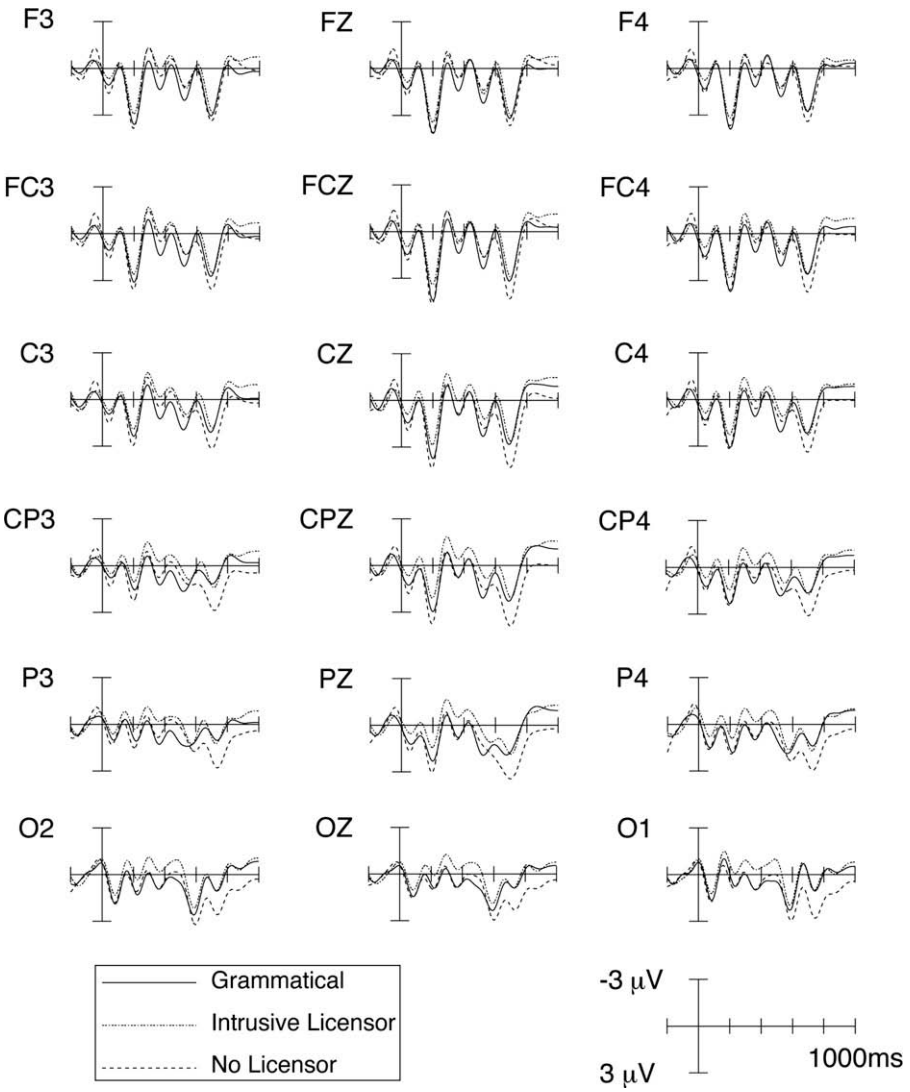

Fig. 1. Grand average waveforms at the negative polarity item *ever* for conditions with the licensor *no*. ERPs are displayed relative to a 200 ms pre-stimulus baseline interval.

ungrammatical intrusive licensor conditions did not elicit a significant P600 effect relative to the grammatical conditions. Moreover, the no licensor condition elicited a significant posterior positivity relative to the two intrusive conditions. The P600 in the no licensor conditions started around 600 ms after the onset of the critical word, and continued to the end of the epoch for both types of licensors. However, the P600 differences between the ungrammatical no licensor conditions and other conditions were not statistically reliable in the 800–1000 ms interval for the licensor *no*.

One potential concern with this result is that the intrusive licensor conditions differed from the other conditions already before the onset of the P600 effect, due to the presence of a broadly distributed negativity in the intrusive licensor conditions. However, we suggest that this effect is unlikely to directly reflect the processing of the NPI *ever*, and further analyses that eliminate the potential confound of this early effect do not impact the presence of the intrusion effect. Several observations suggest that the negativity reflects processes associated with earlier words rather than processing the NPI itself. First, the negativity was present already in the 0–200 ms interval, earlier even than reports of the ELAN component elicited by word category violations (Friederici et al., 1993; Hahne & Friederici, 1999; Neville et al., 1991). The fact that the negativity is visible almost immediately after the onset of the critical word makes it unlikely that it indexes NPI licensing processes. In addition, the fact that the negativity starts well before the appearance of a difference between the grammatical and no licensor conditions suggests that the negativity is not driven by speakers' detection of the ungrammaticality of the sentence. A possible explanation is that the negativity reflects processes associated with the preceding relative clause. This is consistent with the question comprehension data presented in the previous section. It is possible that the presence of a negative quantifier inside a relative clause may initiate extra processes to identify the relevant discourse entities that the head noun refers to and that these processes might selectively impact the comprehension of the intrusive licensor conditions. However, the design of this study was not suitable to confirm or disconfirm this suggestion.

In order to evaluate the possible impact of the early negativity on the critical P600 effect we conducted a further set of ANOVA analyses based on a 200 ms post-stimulus baseline interval. This eliminated all differences between conditions before the critical 600–800 ms interval. But, critically for our purposes, the rebaselined analysis did not change the observed effect qualitatively. In the conditions with the NPI licensor *no* the ungrammatical no licensor condition elicited a P600 relative to the grammatical control condition in the 600–800 ms interval ($F(1,27) = 4.4$, $p < .05$).

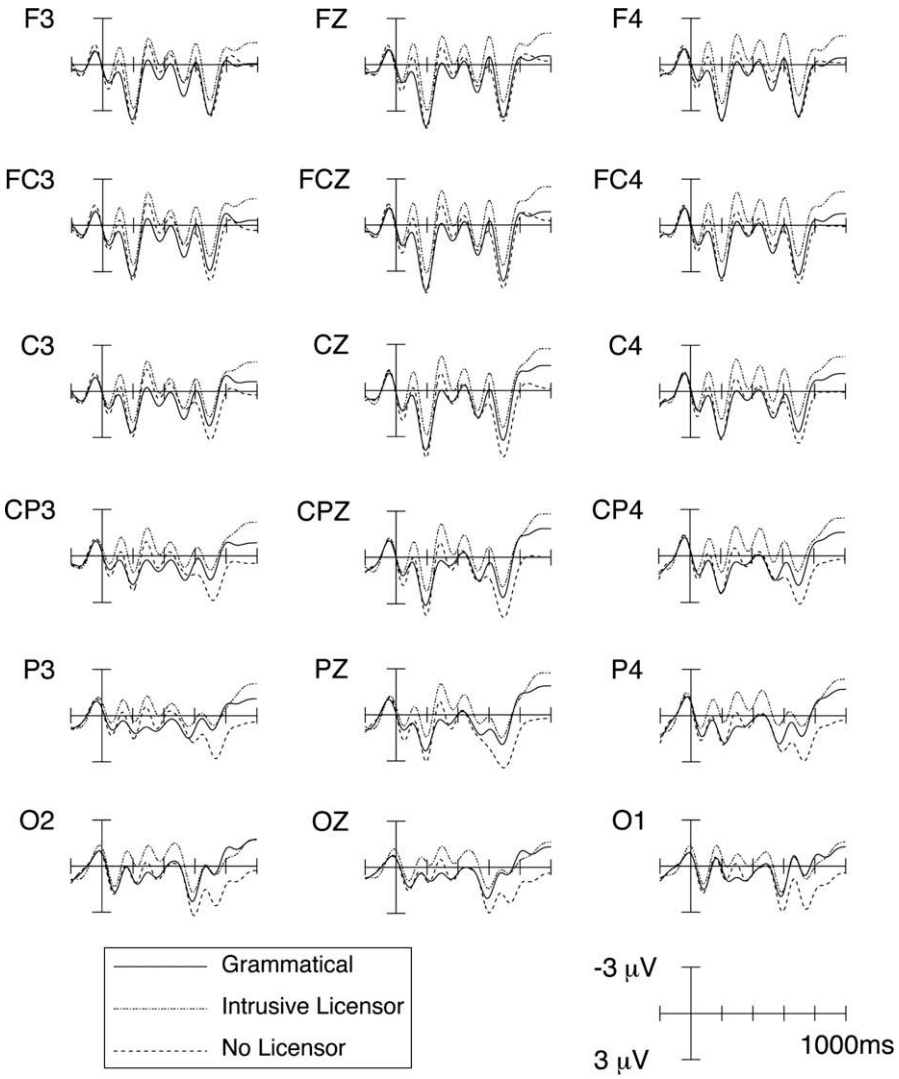

Fig. 2. Grand average waveforms at the negative polarity item *ever* for conditions with the licensor *very few*. ERPs are displayed relative to a 200 ms pre-stimulus baseline interval.

Table 3
ANOVA *F*-values at the critical word (*ever*) for all time intervals, with the factors *condition* and *region*, for the NPI conditions.

Overall ANOVA	0–200 ms	200–400 ms	400–600 ms	600–800 ms	800–1000 ms
Condition (4,108)	—	4.0**	—	2.9*	2.3†
Region (2,54)	—	—	—	—	—
Condition × region (8,216)	—	—	—	2.5*	3.5*

* $p < .05$.
** $p < .01$.
† $p < .1$.

The intrusive licensor condition showed a posterior positivity in the same interval that placed it between the two other conditions, but this condition did not differ significantly from either condition. The same pattern extended to the end of the epoch, but was not statistically significant in the 800–1000 ms interval. For the conditions with the NPI licensor *very few* the no licensor condition again showed a P600-like positivity in the 600–800 ms interval, relative to the grammatical control condition, ($F(1,27) = 7.8$, $p < .05$ at the posterior region; $F(1,27) = 4.7$, $p < .05$ at the central region). The no licensor condition also showed a P600 relative to the intrusive licensor condition ($F(1,27) = 4.2$, $p < .05$ at the posterior region; $F(1,27) = 4.7$, $p < .05$ at the central region). The positivity continued to the 800–1000 ms interval in the posterior region, where the difference between the no licensor condition and the grammatical control was marginally significant ($F(1,27) = 3.8$, $p = .06$), and the difference between the grammatical control and the intrusive licensor condition was significant ($F(1,27) = 5.1$, $p < .05$). Since the two different baselines did not impact the overall pattern of P600 effects, we conclude that the intrusion effect seen in the P600 is a genuine reflection of processing the NPI *ever*.

In order to investigate whether the intrusion effect found in the ERP results differed between the two NPI licensors we conducted a further ANOVA for the time intervals where we found a P600 effect. This ANOVA included the two grammatical conditions and the two intrusive licensor conditions in a 2×2 analysis with the factors *licensor* and *position* of the licensor. The analysis revealed no significant main effects or interactions involving the licensor factor at any region of interest.

4.1.3. Discussion

In summary, presentation of a negative polarity item in the absence of a preceding licensor elicited a P600-like component, relative to grammatical control conditions in which the NPI was licensed by a c-commanding NP containing *no* or *very few*. More

Table 4
ANOVA *F*-values for pairwise comparisons involving the *no* conditions at the critical word (*ever*) for all time intervals and each ROI.

No (1,27)	0–200 ms	200–400 ms	400–600 ms	600–800 ms	800–1000 ms
Grammatical vs. intrusive					
Anterior	–	–	–	–	–
Central	3.3†	4.2*	–	–	–
Posterior	3.6†	5.2*	–	–	–
Grammatical vs. ungrammatical					
Anterior	–	–	–	–	–
Central	–	–	–	3.6†	–
Posterior	–	–	–	5.1*	–
Intrusive vs. ungrammatical					
Anterior	–	–	–	–	–
Central	6.3*	3.8†	–	5.5*	–
Posterior	5.8*	6.6*	–	4.8*	–

* $p < .05$.
† $p < .1$.

Table 5
ANOVA *F*-values for pairwise comparisons involving the *very few* conditions at the critical word (*ever*) for all time intervals and each ROI.

Very few (1,27)	0–200 ms	200–400 ms	400–600 ms	600–800 ms	800–1000 ms
Grammatical vs. intrusive					
Anterior	5.5*	6.5*	5.5*	4.0†	3.2†
Central	6.2*	7.5*	5.8*	–	–
Posterior	6.2*	6.5*	3.7*	–	–
Grammatical vs. ungrammatical					
Anterior	–	–	–	–	–
Central	–	–	–	8.1**	–
Posterior	–	–	–	13.9**	6.1*
Intrusive vs. ungrammatical					
Anterior	4.7*	3.1†	–	8.4**	6.8*
Central	8.8*	9.0**	5.4*	15.4**	14.6**
Posterior	8.8*	9.4**	6.3*	12.9**	12.3**

* $p < .05$.
** $p < .01$.
† $p < .1$.

Table 6
Grand average and standard errors for the two critical time windows for the NPI conditions.

NPI conditions mean voltages (μV)	600–800 ms	800–1000 ms
Anterior		
No		
Grammatical	1.36 (0.59)	0.21 (0.46)
Intrusive	1.04 (0.50)	−0.50 (0.52)
Very few		
Grammatical	1.46 (0.57)	−0.01 (0.51)
Intrusive	0.22 (0.46)	−1.42 (0.57)
No licensor (ungrammatical)	1.98 (0.55)	0.10 (0.56)
Central		
No		
Grammatical	0.98 (0.58)	−0.49 (0.47)
Intrusive	0.63 (0.45)	−0.88 (0.50)
Very few		
Grammatical	0.92 (0.54)	−0.88 (0.49)
Intrusive	0.13 (0.50)	−1.92 (0.52)
No licensor (ungrammatical)	2.29 (0.53)	0.33 (0.60)
Posterior		
No		
Grammatical	0.84 (0.53)	−0.53 (0.41)
Intrusive	0.63 (0.44)	−0.61 (0.51)
Very few		
Grammatical	0.46 (0.50)	−1.17 (0.43)
Intrusive	0.27 (0.57)	−1.49 (0.55)
No licensor (ungrammatical)	2.34 (0.57)	0.78 (0.66)

importantly for the purposes of this study, this P600 effect was either absent or substantially attenuated in ungrammatical sentences where the NPI was preceded by a non-c-commanding potential licensor. This reduction of the P600 is an ERP counterpart of the intrusive licensing from grammatically inaccessible licensors observed in other studies of NPI processing (e.g., Drenhaus et al., 2005). Moreover, the fine-grained temporal sensitivity of ERPs showed that there was no delay between the onset of the grammaticality effect and the onset of the effect of intrusive licensing. This suggests that the intrusion effect is not a consequence of processes that are triggered only after the parser fails to identify a grammatically accessible licensor, e.g., reanalysis processes. Rather, the intrusion effect appears to reflect representations or processes that are already in place at the earliest stages of licensing an NPI. This time course information is consistent with earlier studies of intrusion in NPI licensing, and it is particularly relevant to the comparison of NPI licensing and reflexive binding presented here.

The study tested both the high-frequency licensor *no* and the lower-frequency licensor *very few*, in order to address the possibility that intrusive licensing effects might be specifically related to the presence of high-frequency licensors. Our analyses found no evidence to suggest that intrusive licensing effects differ as a function of licensor type.

However, although the relative timing of the ungrammaticality effect and the intrusive licensing effect is consistent with the one previous ERP study on intrusive NPI licensing (Drenhaus et al., 2005), the ERP response elicited in the no licensor condition contrasts with some previous ERP studies of NPIs. In these studies a central or anterior negativity was observed in response to ungrammatical NPIs, in addition to a P600 effect: a LAN effect was reported in Shao and Neville (1998) and an N400 effect in Drenhaus et al. (2005) and Saddy, Drenhaus, and Frisch (2004). No similar LAN or N400 effect was observed in this study. Unfortunately, a direct comparison between these studies and the current one is not straightforward, due to methodological and cross-linguistic differences. In the study by Shao and Neville (1998), the critical comparison involved the difference between unlicensed *ever* and the different word *never* in the control condition (e.g., *Max says he has ever/never been to the party*). In the study by Drenhaus et al. (2005), participants received a training session on sentences similar to the critical experimental conditions, which might have caused them to engage task-related strategies for error detection. In both studies, participants were asked to judge the grammaticality of sentences during the EEG recording. The present study instead used comprehension questions, and participants were not alerted to the appearance of ungrammatical sentences prior to the study. Some previous studies (e.g., Urban, Gunter, Friederici, & Bormann, 2000) have shown that the choice of behavioral task may have a significant impact on the ERP components elicited during a language task. For instance, Hahne and Friederici (2002) also reported that a task that explicitly asked participants to judge the semantic coherence of the sentence modulated the N400 effect. Their results led them to suggest that the N400 associated with semantic aspects of sentence comprehension reflects controlled processes. A recent study by Steinhauer, Drury, Portner, Walenski, and Ullman (2007) tested a wider range of English NPIs, and similar to our results they found no effects of NPI (non-)licensing earlier than the P600 for *ever* and *any*. Interestingly, they found an N400 for the unlicensed NPI *at all*, suggesting that the response profile to different NPIs may vary. If this is the case, then direct cross-linguistic comparison between the German results presented by Drenhaus and colleagues and the results presented here might not be possible. However, despite the differences with respect to the earliest component elicited in these studies, our results and previous results are similar in the respect that the earliest effect

of the intrusive licensor appears at the same latency as does the effect of ungrammaticality of the no licensor conditions (see also Drenhaus, beim Graben, Frisch, & Saddy, 2006, for an analysis of the Drenhaus et al. (2005) data using Symbolic Resonance Analysis). In the current study, this is reflected as a reduction of P600 in the intrusive licensing condition, whereas in the German studies by Drenhaus and colleagues, this is reflected as a reduction of the N400 component. This pattern of results is consistent with the interpretation that NPI intrusion effects reflect an immediate illusion of grammaticality, rather than the result of a later process initiated after detection of an anomaly.

4.2. Reflexive binding

4.2.1. Comprehension questions

Overall average accuracy for all trials that included a comprehension question was 89.4%, with average accuracy for target items of 86.8%. Individual condition averages were 88.0% (*SD* 11%) for congruous reflexives, 86.7% (*SD* 11%) for the intrusive condition, and 85.6% (*SD* 9%) for the incongruous condition. Paired-samples *t*-tests revealed no significant differences among the conditions for both participants and items analyses.

4.2.2. Event-related potentials

Fig. 3 displays grand average ERPs at the 18 electrodes used in statistical analyses. ERPs elicited by reflexives with a congruent antecedent are plotted against responses in the incongruent and intrusive antecedent conditions. Visual inspection reveals that both conditions with a reflexive that mismatched the stereotypical gender of the accessible antecedent elicited a posterior positivity relative to the matched congruent condition, with an onset at around 400 ms and a peak at around 600 ms. The onset, amplitude and distribution of the positivity for these two conditions was identical initially, but the positivity was longer lasting in the intru-

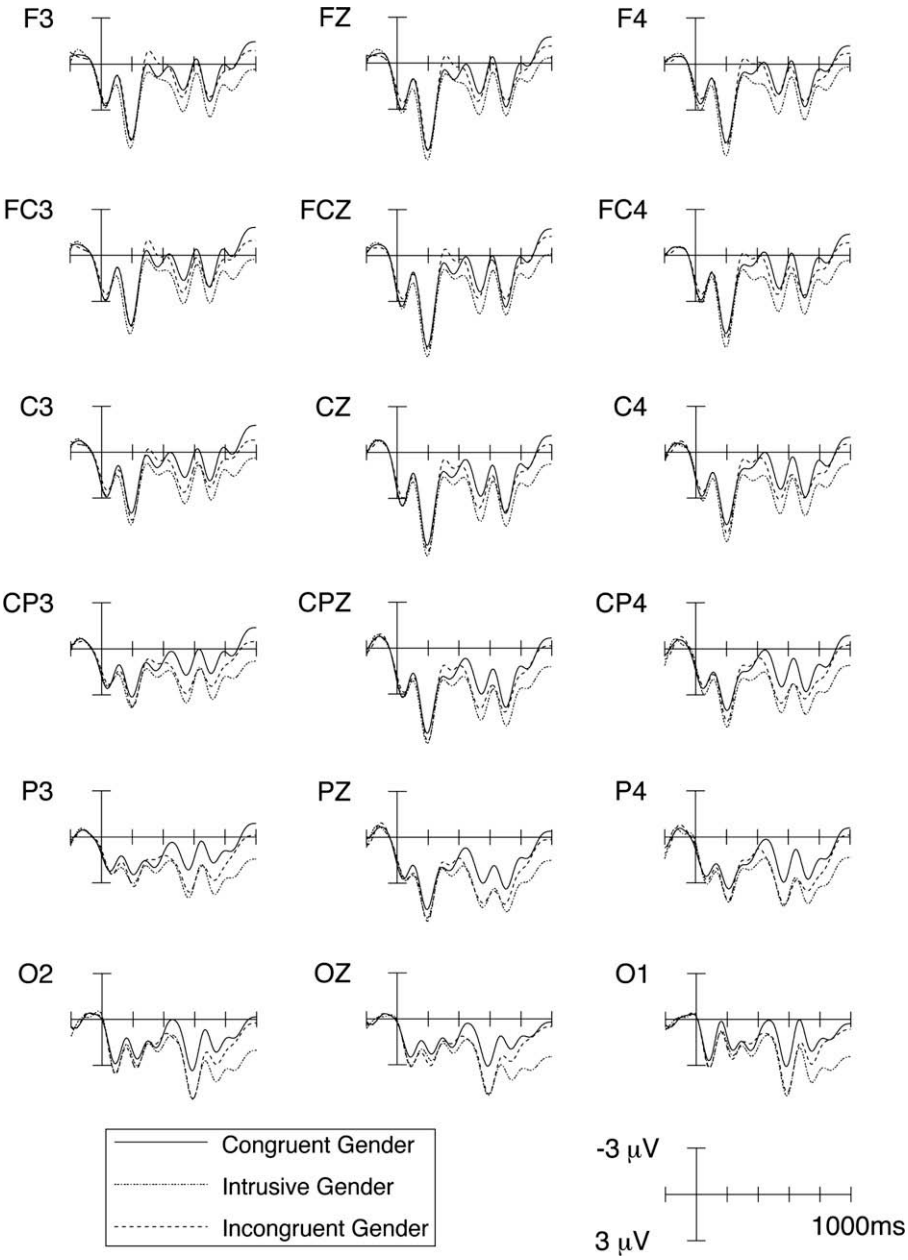

Fig. 3. Grand average waveforms at the reflexive (*himself*, *herself*) in the reflexive conditions. Waveforms are displayed relative to a 200 ms pre-stimulus baseline interval.

Table 7
ANOVA *F*-values at the critical word *himself/herself* for all time intervals in the reflexive conditions.

Overall ANOVA	0–200 ms	200–400 ms	250–350 ms	400–600 ms	600–800 ms	800–1000 ms
Condition (2,54)	—	—	—	3.1†	—	—
Region (2,54)	8.7**	5.8*	15.0***	6.8**	10.6**	9.9**
Condition × region (4,108)	—	—	—	—	3.1*	—

* $p < .05$.
** $p < .01$.
*** $p < .001$.
† $p < .1$.

sive condition, although the latter part of the positivity was not reliable. However, visual inspection also suggested early differences between the incongruent and intrusive conditions, in the form of a central-anterior negativity between 250 and 350 ms. Consequently, in addition to the standard analysis intervals used for the NPI data, an additional interval lasting from 250 to 350 ms was included in the analyses.

Table 7 shows the results from the omnibus ANOVA with the factors *condition* and *region* in the reflexive conditions. It revealed a condition × region interaction in the 600–800 ms interval, a marginal main effect of condition in the 400–600 ms interval and main effects of region in every time interval. In order to assess the source of this result, planned comparisons were performed within each ROI. These comparisons compared the congruent condition, the incongruent condition, and the intrusive condition. The results of these comparisons are shown in Table 8. In Table 9, we present the mean and SE values for the three critical windows. Both visual inspection and the values presented in Table 9 indicate that during the late time window (800–1000 ms), there was a strong trend for a greater positivity in the intrusive condition than in the other two conditions. However, these differences did not reach significance: intrusive vs. congruent, $F(1,27) = 2.4$, $p = .13$; intrusive vs. incongruent, $F(1,27) = 2.4$, $p = .13$.

Finally, we also compared the size of the basic ungrammaticality/incongruency effect elicited by the no licensor NPI condition and the incongruent reflexive condition, respectively. The size of the effect was determined based upon the mean voltage differences between the no licensor and the grammatical NPI conditions during the 600–800 ms interval, and between the incongruent and the congruent reflexive conditions during the 400–600 ms interval. Using data from the six posterior electrodes where the P600 effect was strongest, we entered voltage differences into a 2 [NPI vs. reflexive] × 6 [channels] ANOVA. We found no main effect of dependency type or of channels, and no interaction. Therefore, the basic P600 effect appears to be no different in the NPI and reflexive violation conditions, excepting the apparent latency differences between the two.

Table 8
ANOVA *F*-values for pairwise comparisons of the reflexive conditions.

Reflexive conditions (1,27)	0–200 ms	200–400 ms	250–350 ms	400–600 ms	600–800 ms	800–1000 ms
Congruent vs. intrusive						
Anterior	—	—	—	—	—	—
Central	—	—	—	4.0†	—	—
Posterior	—	—	—	4.9*	5.5*	—
Congruent vs. incongruent						
Anterior	—	—	—	—	—	—
Central	—	—	—	3.9†	—	—
Posterior	—	—	—	9.1**	3.8†	—
Intrusive vs. incongruent						
Anterior	—	—	3.9†	—	—	—
Central	—	—	3.6†	—	—	—
Posterior	—	—	—	—	—	—

* $p < .05$.
** $p < .01$.
† $p < .1$.

Table 9
Grand average and standard errors for the three critical time windows for the reflexive conditions.

Reflexive conditions mean voltages (μV)	400–600 ms	600–800 ms	800–1000 ms
Anterior			
Congruent	0.99 (0.69)	1.16 (0.89)	−0.43 (0.80)
Intrusive	2.43 (0.50)	2.30 (0.73)	1.20 (0.86)
Incongruent	1.39 (0.64)	1.40 (0.71)	0.61 (0.61)
Central			
Congruent	1.36 (0.68)	1.61 (0.91)	0.12 (0.86)
Intrusive	2.98 (0.51)	3.16 (0.64)	1.96 (0.80)
Incongruent	2.35 (0.60)	2.42 (0.71)	0.53 (0.64)
Posterior			
Congruent	1.47 (0.62)	1.80 (0.89)	0.76 (0.89)
Intrusive	2.96 (0.44)	3.89 (0.54)	2.89 (0.72)
Incongruent	2.78 (0.57)	3.19 (0.70)	1.12 (0.71)

4.2.3. Discussion

Analysis of the ERPs for the reflexive conditions revealed a posterior positivity with an onset at around 400 ms in both the intrusive and the incongruent conditions. This resembles the P600 component found in previous studies of reflexive violations (Harris, Wexler, & Holcomb, 2000; Osterhout & Mobley, 1995; Osterhout et al., 1997). There were no differences between the incongruent and intrusive antecedent conditions, and both were significantly different from the congruent antecedent condition. Additionally, a marginal central-anterior negativity was observed in the 250–350 ms interval, suggesting a difference between the intrusive and incongruent conditions. (Expanding the analysis to broader time intervals, such as 300–500 ms, yielded less reliable differences.) This marginal effect suggests a difference between the intrusive condition and other two conditions at an interval prior to the P600. Although the exact source of this difference is unclear, we suggest that it should not be interpreted as the first index of anomaly for two reasons. First, the basic comparison between the congruent and incongruent conditions failed to show any reliable difference in the earlier time intervals. Second, previous ERP results on reflexive binding have consistently observed P600 responses to incongruent binding relations. To our knowledge, no effects of an earlier negativity have ever been reported. The observed P600 pattern suggests that participants were immediately sensitive to the incongruence of the sentence, and that there was no effect of intrusive licensing in the early detection of the anomaly.

Interestingly, although visual inspection of the 800–1000 ms time interval suggests that at the posterior region the intrusive condition showed a greater positivity than the two other conditions, these differences were not reliable (to repeat: intrusive vs. congruent, $F(1,27) = 2.4$, $p = .13$; intrusive vs. incongruent, $F(1,27) = 2.4$, $p = .13$). However, if this trend reflects a real difference, it might be interpreted as a late intrusion effect. That is, the longer-lasting positivity seen in the intrusive condition could indicate that it is at this later stage that the presence of intrusive noun phrases with matched gender features can exert an influence. This intrusion effect could lead to increased processing cost and potentially to erroneous interpretations. Note that this trend goes in the opposite direction from the intrusion effect found in the NPI conditions, where the intrusive condition showed a reduced P600 effect, instead of an enhanced

one, relative to the no licensor conditions. This trend is consistent with the patterns observed in Sturt (2003). Sturt used eye-tracking measures and a similar design to this study, although the structural position of the inaccessible antecedent was different in Sturt's studies than in ours. Sturt consistently found no intrusion effect in first pass eye-movement measures, but in the first of his two experiments he found late effects of inaccessible antecedents. In particular, second pass reading times on the reflexive were slowest in the accessible mismatch/inaccessible match condition that is most similar to our intrusive condition. Sturt's Experiment 2 found no corresponding intrusion effect. Relatedly, Badecker and Straub (2002, Experiment 3) found a delayed slowdown effects in self-paced reading times two words after a reflexive that matched both an accessible and an inaccessible antecedent. They found no corresponding effect in two subsequent experiments with reflexive or reciprocal anaphors (Experiments 4 and 5). Taken together, these findings suggest that grammatically inaccessible antecedents for reflexives may lead to greater processing cost, although the effect is not consistent across studies.

5. General discussion

5.1. Summary of main results

The ERP experiment reported here compared reflexive binding and NPI licensing using maximally similar manipulations. In both cases, the correct licensor/antecedent was the subject of the main clause, and the intrusive licensor/antecedent was embedded within a relative clause that modified the subject. The absence of a suitable antecedent/licensor led to a P600 effect in both cases, but the effect of the intrusive licensor/antecedent was not the same for reflexives and NPIs. In the NPI conditions the ERPs in the intrusive condition diverged from the no licensor condition as soon as the effect of ungrammaticality started (around 600 ms), and appeared as a reduction in the P600 effect in the intrusive condition. This suggests that the intrusive licensor in some way disrupts the earliest stages of grammaticality detection for NPIs. In contrast, the P600 in the reflexive conditions was initially identical in the incongruent and intrusive conditions. This suggests that intrusive antecedents do not impact the initial stages of reflexive binding in the same way that intrusive licensors do for NPIs.

This study thus establishes a clear contrast between two different domains. One shows early effects of intrusive licensing whereas the other does not. In addition, we saw a trend in the latest time interval in the reflexive conditions that could potentially reflect a late intrusion effect. Importantly, however, this intrusion effect had the opposite polarity profile from the early NPI intrusion. To our knowledge, this is the first side-by-side comparison of intrusion in different grammatical domains, using the same structural configurations in a within-subjects design. The contrast between reflexives and NPIs presents a challenge for any theory that assumes (i) that NPI intrusion effect is the consequence of general-purpose retrieval mechanisms, and (ii) that reflexive binding and NPI licensing reflect similar item-to-item structural dependencies. The contrast is consistent with our hypothesis that reflexive binding and NPI licensing reflect fundamentally different types of relations. However, one wonders how the two linguistic domains give rise to the specific ERP effects that we observe, and of what this implies for theories of sentence processing in general. In the following two sections we elaborate on possible reasons for (non-)effects of intrusion in reflexive binding and NPI licensing, respectively.

5.2. Accurate retrieval in building syntactic dependencies

The lack of an intrusion effect for reflexive binding is consistent with the results of previous behavioral studies that have found no intrusive effects in early eye-tracking measures (Sturt, 2003). Theoretical accounts of reflexive binding diverge on the issue of whether constraints on reflexives should be stated in syntactic or semantic terms, but there is little controversy over the assumption that a reflexive is specifically dependent on an antecedent for its interpretation.

These findings require a system that has representations and retrieval mechanisms that are sufficiently robust to prevent spurious retrieval of an inaccessible licensor for a reflexive, and that can search for an antecedent that has specific semantic features (e.g., feminine gender) without being misled by a structurally inappropriate antecedent that bears those features. This effect can be achieved straightforwardly in any architecture where it is possible to directly address structural notions like 'subject of the current clause' upon processing a reflexive (We set aside here the well-known fact that reflexives in many languages, including English, may take non-subjects as antecedents, such as *John showed Mary herself in the mirror*). Although this is a structural notion that is easily expressed, it is not a notion that can be straightforwardly implemented in just any architecture. It requires either that the parser have a sufficiently rich control mechanism such that the property 'item in the current clause' can be used as a retrieval cue, or that the syntactic structure of a sentence be represented faithfully enough that this notion can be reconstructed in an online search of a complete tree structure.

In models that rely on the assumption that chunks of structure are retrieved from content-addressable memory it is less straightforward to capture notions such as 'subject of the current clause', since being part of the current clause is not an inherent property of any element. Such models can only capture notions like this by relying on the structure of their control mechanism. For example, in the ACT-R-based architecture of Lewis and Vasishth the insensitivity of reflexive binding to intrusive antecedents is surprising if semantic and structural retrieval cues are assumed to be similarly weighted (Lewis & Vasishth, 2005). It is, of course, possible to adjust the model parameters such that structural cues are more strongly weighted, but this leaves open the question of how to implement a structural cue like 'subject of the current clause'. Although this is not encoded in extant versions of the model, it could perhaps be expressed through an elaboration of the 'goal buffer', which is a key part of the control mechanism in the architecture. The immediate constituents of each clause could be marked with a clause index that encodes which clause the constituent is a part of, expanding on the binary distinction between embedded and non-embedded clauses used in existing versions of the model. The clause index of each constituent would remain unchanged as the parser's state changes. The goal buffer could then keep track of the index of the clause that the parser is working on at a given moment. Using these mechanisms, an incoming reflexive could find an appropriate antecedent based on retrieval cues that specify that the antecedent must be a subject and must also bear the clause index currently specified in the goal buffer. This is one possible implementation, although it should be clear that this involves different model parameters from those that yield predictions of intrusive licensing effects, and that have sometimes been used to motivate the ACT-R architecture of sentence processing (e.g., Vasishth et al., 2008).

Additionally, one must also take into account the fact that reflexive conditions potentially showed a late intrusion effect with properties that are the opposite of the early NPI intrusion pattern, as discussed in Section 4.2.3. This pattern is consistent with some previous studies using other techniques (Badecker & Straub, 2002;

Sturt, 2003). Crucially, any reflexive intrusion effects are delayed and tend to have the opposite effect from intrusive NPI licensing. Intrusive NPI licensors facilitate processing, leading to an illusion of grammaticality, whereas intrusive antecedents for reflexives tend to cause increased processing difficulty. As such, the divergent pattern of results for the two dependencies is inconsistent with the predictions of the current ACT-R model.

We should caution that our conclusions about reflexives do not necessarily generalize to other cases of anaphora, since in other types of anaphora the empirical evidence for structurally accurate retrieval is more equivocal, and the theoretical challenges for creating an accurate retrieval mechanism are greater. As discussed in Section 1, pronouns are subject to an anti-locality constraint that prevents a pronoun from taking a clause-mate as its antecedent (e.g., **Sue fears that Bob criticizes him too much*; Principle B, Chomsky, 1981). Previous language processing studies offer an unclear answer to the question of whether speakers successfully exclude clause-mate arguments from consideration when searching for an antecedent for pronouns like *him*, *her*, and *them* (Badecker & Straub, 2002; Clifton et al., 1997; Kennison, 2003; Nicol & Swinney, 1989). At the same time as the empirical evidence is rather mixed, it is also more difficult to specify models of grammatically accurate retrieval once we move beyond the domain of reflexives. Once a structural notion like 'subject of the current clause' is defined, it is straightforward to use that notion to target an appropriate antecedent for a reflexive. In contrast, the constraints on pronoun antecedents are stated in negative terms (e.g., *not in the current clause*), and potential antecedents can occur in a far broader range of positions. Therefore, it may be more difficult to avoid intrusion effects for pronouns. Similar considerations arise for reflexives that may have logophoric interpretations (e.g., Runner et al., 2006).

5.3. Licensing NPIs online

Current leading accounts of negative polarity claim that the licensing condition results from an interaction between the lexical semantic features of NPIs and the semantic and pragmatic properties of the linguistic environments that host the NPIs. Under these accounts the traditional c-command requirement on NPI licensing can be viewed as a by-product rather than an explicit component of the licensing theory. Although there are differences among theories, most current accounts share two critical insights (Chierchia, 2006; Fauconnier, 1975; Israel, 2004; Kadmon & Landman, 1993; Krifka, 1995; Ladusaw, 1992). First, the lexical meanings of NPIs correspond to extreme values along a scale of contextually determined alternatives. For example, the NPI *ever* refers to a broad time interval, contrasting with shorter intervals such as *a day* or *a week*; similarly, the NPI *lift a finger* is a predicate that describes a minimal degree of effort, contrasting with notions such as *make an effort* or *try one's hardest*. Second, NPIs have the pragmatic effect of strengthening the statements that they occur in, and hence a successful licensing environment must have the property that the NPIs inclusion leads to a strengthened statement. The strengthening effect can be observed in examples like (8), where the inclusion of the NPI *ever* in (8a) precludes the possibility of exceptions, unlike in (8b).

(8) a. The kids won't ever eat carrots. #But maybe they will if you ask nicely.
b. The kids won't eat carrots. But maybe they will if you ask nicely.

Importantly, the inclusion of an NPI yields a strengthened statement only when appears in a type of semantic context described as *downward entailing* (DE, Ladusaw, 1979). Negation is the prototypical example of an operator that creates a downward entailing environment. Compared to the statement *The kids won't eat carrots this week*, a stronger claim is that *The kids won't eat carrots this month*, and it is stronger yet to claim that *The kids won't ever eat carrots*. In the absence of negation the relations reverse. The statement *The kids will eat carrots this month* is a weaker claim than *The kids will eat carrots this week*. The variant of the sentence with the NPI *ever* is, by hypothesis, excluded in the positive context because it fails to yield a strengthened statement. Like negation, most NPI licensors share the property of contributing the property of downward entailment to material in their semantic scope. As noted in Section 2.3, an explicit licensor is not always required to create a DE context. Nonetheless, when a licensor is present, an NPI must appear within the semantic scope of the licensor in order for the interaction of NPI semantics and licensor properties to obtain the strengthening effect. Since semantic scope relations often mirror c-command relations in syntactic structure, the commonly observed c-command requirement can be derived as a by-product, rather than as a formal licensing requirement on NPIs. Accordingly, in those few cases where semantic scope does not align with c-command relations, the c-command requirement disappears (see example (7) in Section 2.3).

Under this account, licensing of NPIs and reflexives involves fundamentally different mechanisms. The structural c-command requirement is a formal licensing condition for reflexive binding, but is an emergent property of the semantics and pragmatics of NPIs. More importantly, reflexive binding entails retrieval of an antecedent, whereas the semantic/pragmatic account of NPIs predicts that NPI licensing does not entail retrieval of a licensor. Since our results did not show immediate intrusion effects for reflexive binding we conclude that at least this retrieval process is immune to similarity-based interference. Meanwhile, we suggest that the intrusion effect in NPI licensing is a consequence of indirect licensing, with its roots in the inferences that speakers generate during online semantic and pragmatic processing.

It has long been noted that in addition to direct semantic licensing through DE environments, pragmatic inferences play an important role in NPI licensing. For instance, it has been argued that speakers can draw negative inferences from a positive statement that are sufficient to license NPIs (Giannakidou, 2006; Linebarger, 1987). The role of negative implicatures provides a way to account for contrasts between pairs of sentences like *I am surprised that we have any sugar.* vs. *?I am sure we have any sugar* (modified from Linebarger, 1987). By hypothesis, *any* is licensed in the first of these sentences because it gives rise to a negative implicature: *I expected that we do not have any sugar*. Although such 'indirect' licensing effects are not yet fully understood, they may provide a clue to the source of intrusive licensing effects. Specifically, intrusive licensing may reflect indirect licensing that is triggered by erroneous negative implicatures that are generated during the course of sentence interpretation. Restrictive relative clauses that contain negative quantifiers may be particularly prone to erroneous implicatures.

Our proposal relies upon two key assumptions. First, we follow much recent evidence in psycholinguistics that comprehenders are sensitive to the contrastive function of restrictive modifiers, and can use the presence of a restrictive modifier to rapidly infer an intended contrast, e.g., the expression *the tall cup* implies a contrast with a non-tall cup (Altmann & Steedman, 1988; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Restrictive relative clauses are good examples of modifiers that invite inferences about contrasting referents. To our knowledge, all demonstrations of intrusive NPI licensing to-date have involved sentences with restrictive

relative clause modifiers. Second, when an assertion is made about one member of a contrastive set, the comprehender is invited to consider whether the speaker intends to convey that the assertion is not true of the other members of the contrastive set. In some circumstances this negative inference is almost mandatory. For example, the contrastively focused sentence *The RED apples are sweet* implies that non-red apples are not sweet, or at least that the speaker does not know the status of the non-red apples. In other circumstances, however, a negative inference is possible but not required. Although the factors that determine whether a negative inference is generated are not well understood, we suggest that speakers may be more likely to generate such inferences if the contrasting referents are made very salient in the discourse. Negative quantifiers can do exactly this. Results from previous psycholinguistic studies suggest that they accomplish this by shifting focus to the complement of the quantified set (e.g., Moxey, Sanford, & Dawydiak, 2001; Sanford, Moxey, & Paterson, 1996).

These assumptions together provide a possible account of how relative clauses and negative quantifiers conspire to yield intrusive NPI licensing effects. During the processing of a modified noun phrase such as *the bills that no democratic senators have voted for* - a comprehender may reasonably infer that the set denoted by this complex expression has some property *P*, i.e., the predicate that has yet to be encountered. Importantly, he may also infer that the speaker intends to contrast this set with another, such as *the bills that some democratic senators have voted for*, and he might further infer that this set does not have the property *P*. If this negative inference is generated, it may suffice to license NPIs. Importantly, the negative inference is pragmatically sensible but by no means logically necessary. Consequently, we predict that there should be trial-by-trial variation in whether speakers generate the negative implicature, consistent with the probabilistic nature of intrusive NPI licensing.

This approach to intrusive NPI licensing differs from the account proposed by Vasishth et al. (2008). Vasishth and colleagues attribute intrusive licensing to partial cue matching during retrieval of a licensor for the NPI, thereby treating the effect as representative of a far broader phenomenon. Under their approach, intrusive NPI licensing is a natural consequence of an architecture for language processing like ACT-R that represents syntactic structures as sets of disconnected chunks and relies on retrieval from a content-addressable memory to build relations between those chunks. The account predicts that intrusive licensing effects should generalize to a broader set of structural environments and a wider class of syntactic phenomena. In contrast, our account does not involve specific retrieval of a licensor, and it depends on pragmatic inferences that are generated in specific contexts. It predicts that intrusive NPI licensing effects are not representative of linguistic dependency formation in general, and also predicts that the strength of intrusive licensing should be modulated by manipulation of the semantic/pragmatic properties of the context. Our account is therefore consistent with the contrasting results in the NPI licensing and reflexive binding conditions in the current experiment, but many further tests will be required in order to determine the generality of intrusive licensing effects.

Since we have offered a semantic/pragmatic account of NPI licensing and intrusive licensing errors it is natural to ask why violations elicited a P600 component, which is typically associated with syntactic and morphological anomalies, rather than modulation of the N400, which is widely described as a reflection of semantic processing. However, there are a number of reasons why this conflict is probably only apparent. First, the kinds of semantic processes indexed by the N400 are probably different from the semantic processes involved in NPI licensing. A number of studies have found that N400 amplitudes are modulated by plausibility based on world knowledge (e.g., Hagoort, Hald, Bastiaansen, & Petersson, 2004), or by lexical associations independent of the truth of the proposition. Fischler and colleagues found no N400 difference to the final word in *A robin is a bird* and *A robin is not a bird*, but did observe a larger N400 in sentences with thematically unrelated target words, such as *A robin is not a vehicle* (Fischler, Bloom, Childers, Roucos, & Perry, 1983; see also Kounios & Holcomb, 1992; Noveck & Posada, 2003). In contrast, NPI licensing involves higher level semantic/pragmatic relations that are independent of lexical associations and world knowledge, and instead rely on notions of logical entailment and pragmatic scales. These logical relations are computed over structural configurations, and some have argued that NPI licensing involves a special type of pragmatic function that is grammaticalized/syntacticized (Chierchia, 2006). From this perspective, NPI licensing may have more in common with the types of structural relations that are characteristically associated with the P600.

Second, it is possible that unlicensed NPIs initiate a type of revision/re-checking process that has often been associated with the P600 in ERP studies. In studies of garden path sentences and syntactic anomalies P600 effects are typically associated with revision or reanalysis processes (Friederici, 2002; Osterhout & Holcomb, 1992). More recently, P600 effects have been observed in a number of studies of syntactically well-formed but semantically anomalous sentences, typically involving thematic anomalies or argument reversals (e.g., *For breakfast the eggs would eat...*; Kim & Osterhout, 2005; Kolk, Chwilla, van Herten, & Oor, 2003; Kuperberg, Holcomb, Sitnikova, Greve, & Dale, 2003), and it has been suggested that this P600 reflects monitoring processes triggered when a comprehender detects uncertainty over syntax-semantics compatibility (Van Herten, Kolk, & Chwilla, 2005). The P600 elicited by unlicensed NPIs may reflect similar processes, since it involves a lexical item that appears in a locally appropriate position but that is judged to be incompatible with the broader semantic context. It is also interesting to note that Drenhaus and colleagues found in an analysis of their ERP data using Symbolic Resonance Analysis that the P600 elicited by unlicensed NPIs may correspond to two temporally distinct components, which they associate with diagnosis and reanalysis, respectively (Drenhaus et al., 2006; see also Friederici, Mecklinger, Spencer, Steinhauer, & Donchin, 2001). This proposal is in line with our results, which suggest that the presence of spurious pragmatic inferences may create immediate illusions of acceptability for NPIs, thus making error detection and reanalysis less likely.

6. Conclusion

In this article we tested the presence and the time course of intrusive licensing effects in the processing of two distinct types of dependency: a syntactic dependency involving reflexive pronouns such as *himself*/*herself*, and the licensing of negative polarity items (NPIs) such as *ever*, which is widely considered to involve semantic and pragmatic relations. ERP recordings demonstrated a robust electrophysiological correlate of intrusion effects for NPI licensing, in the form of a reduction in the P600 elicited by ungrammatical NPIs. But no corresponding modulation of the P600 was observed for reflexive binding, where syntactic constraints appeared to prevent intrusive antecedents from influencing the initial stages of anaphor resolution. Also, we found that there is no time difference between the P600 elicited by ungrammaticality in the absence of an NPI licensor and the reduction of the P600 in the presence of an intrusive licensor. This suggests that the intrusive licensing effect for NPIs is not a consequence of repair-driven processes that are triggered after a violation is detected. We suggest that the contrast between the two phenomena lies in the representational differences that underlie

each dependency. Instead of attributing NPI intrusion effects to a domain-general mechanism of (partial) cue matching and retrieval, we propose that intrusion effect arises in the NPI domain as a consequence of calculating semantic/pragmatic information.

Acknowledgments

The authors thank Esther Chung, Ed Kenschaft, Ellen Lau, Paul Pietroski, Clare Stroud, and Matt Wagers for their help with the study presented here and for useful discussion of the issues, and thank Shravan Vasishth and Rick Lewis for valuable discussions and for making available their ACT-R code, which helped our understanding of the theoretical issues discussed here. Preparation of this paper was supported in part by NSF DGE-0801465 to the U of Maryland.

References

Altmann, G. T., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*, 191–238.

Anderson, J. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science, 29*, 313–341.

Anderson, J., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 748–769.

Chierchia, C. (2006). Broaden your views. Implicatures of domain widening and the 'logicality' of language. *Linguistic Inquiry, 37*, 535–590.

Chomsky, N. (1981). *Lectures on government and binding*. Berlin: Mouton de Gruyter.

Clements, G. N. (1975). The logophoric pronoun in Ewe: Its role in discourse. *Journal of West African Languages, 10*, 141–177.

Clifton, C., Kennison, S. M., & Albrecht, J. E. (1997). Reading the words her, his, him: Implications for parsing principles based on frequency and on structure. *Journal of Memory and Language, 36*, 276–292.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*, 21–58.

Cowart, W., & Cairns, H. S. (1987). Evidence for an anaphoric mechanism within syntactic processing: Some reference relations defy semantic and pragmatic constraints. *Memory & Cognition, 15*, 318–331.

Drenhaus, H., Blaszczak, J., & Schütte, J. (submitted for publication). *An ERP study on the strength of licensers in negative polarity constructions*. Ms, University of Potsdam.

Drenhaus, H., beim Graben, P., Frisch, S., & Saddy, D. (2006). Diagnosis and repair of negative polarity constructions in the light of symbolic resonance analysis. *Brain and Language, 96*, 255–268.

Drenhaus, H., Saddy, D., & Frisch, S. (2005). Processing negative polarity items: When negation comes through the backdoor. In S. Kepser & M. Reis (Eds.), *Linguistic evidence—empirical, theoretical, and computational perspectives* (pp. 145–165). Berlin/New York: Mouton de Gruyter.

Fauconnier, G. (1975). Polarity and the scale principle. In *Proceedings of the 11th Chicago Linguistics Society* (pp. 188–199).

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology, 20*, 400–409.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*, 78–84.

Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., & Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: A spatio-temporal analysis of event-related brain potentials. *Cognitive Brain Research, 11*, 305–323.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Brain Research: Cognitive Brain Research, 1*, 183–192.

Giannakidou, A. (1998). *Polarity sensitivity as (non)veridical dependency*. Amsterdam, Philadelphia: John Benjamins.

Giannakidou, A. (2006). Only, emotive factives, and the dual nature of polarity dependency. *Language, 82*, 575–603.

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1304–1321.

Graff, D. (2003). *English gigaword*. Philadelphia: Linguistic Data Consortium.

Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95–112.

Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*, 439–483.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*, 438–441.

Hagoort, P., Wassenaar, M., & Brown, C. M. (2003). Syntax-related ERP-effects in Dutch. *Brain Research: Cognitive Brain Research, 16*, 38–50.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis. Early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11*, 194–205.

Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Brain Research, 13*, 339–356.

Harris, T., Wexler, K., & Holcomb, P. (2000). An ERP investigation of binding and coreference. *Brain and Language, 75*, 313–346.

Israel, M. (1998). *The rhetoric of grammar: Scalar reasoning and polarity sensitivity*. Ph.D. Thesis, U.C., San Diego.

Israel, M. (2004). The pragmatics of polarity. In L. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 701–723). Oxford: Blackwell.

Jackendoff, R. (1992). Mme. Tussaud meets the binding theory. *Natural Language & Linguistic Theory, 10*, 1–31.

Kadmon, N., & Landman, F. (1993). Any. *Linguistics and Philosophy, 16*, 353–422.

Kazanina, N., Lau, E., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The effect of sysntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language, 56*, 384–409.

Kennison, S. M. (2003). Comprehending the pronouns her, him, and his: Implications for theories of referential processing. *Journal of Memory and Language, 49*, 335–352.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 52*, 205–225.

Kolk, H., Chwilla, D., van Herten, M., & Oor, P. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language, 85*, 1–36.

Kounios, J., & Holcomb, P. (1992). Structure and process in semantic memory: Evidence from event-related brain potentials and reaction times. *Journal of Experimental Psychology: General, 121*, 459–479.

Krifka, M. (1995). The semantics and pragmatics of weak and strong polarity items. *Linguistic Analysis, 25*, 209–257.

Kuperberg, G., Holcomb, P., Sitnikova, T., Greve, D., & Dale, A. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience, 15*, 272–293.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203–205.

Ladusaw, W. A. (1979). *Negative polarity items as inherent scope relations*. Ph.D. Dissertation, University of Texas at Austin.

Ladusaw, W. A. (1992). Expressing negation. In C. Barker & D. Dowty (Eds.), *Proceedings of the second conference on semantics and linguistic theory (SALT)*. Ohio State University (pp. 237–259).

Lau, E. F., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of prediction in rapid syntactic analysis. *Brain and Language, 98*, 74–88.

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

Lewis, R., Vasishth, S., & Van Dyke, J. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science, 10*, 447–454.

Linebarger, M. (1987). Negative polarity and grammatical representation. *Linguistics and Philosophy, 10*, 325–387.

Moxey, L. M., Sanford, A. J., & Dawydiak, E. J. (2001). Denials as controllers of negative quantifier focus. *Journal of Memory and Language, 44*, 427–442.

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3*, 151–165.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18*, 5–19.

Noveck, I., & Posada, A. (2003). Characterizing the time course of an implicature and evoked potentials study. *Brain and Language, 85*, 203–210.

Oldfield, J. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97–113.

Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition, 25*, 273–285.

Osterhout, L., & Holcomb, P. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785–804.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language, 34*, 739–773.

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

Phillips, C. (2006). The real-time status of Island phenomena. *Language, 82*, 795–823.

Pollard, C., & Sag, I. A. (1992). Anaphors in English and the scope of the binding theory. *Linguistic Inquiry, 23*, 261–303.

Reinhart, T., & Reuland, E. (1993). Reflexivity. *Linguistic Inquiry, 24*, 657–720.

Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2006). Processing reflexives and pronouns in picture noun phrases. *Cognitive Science, 30*, 193–241.

Saddy, D., Drenhaus, H., & Frisch, S. (2004). Processing polarity items: Contrastive licensing costs. *Brain and Language, 90*, 495–502.

Sanford, A. J., Moxey, L. M., & Paterson, K. B. (1996). Attentional focusing with quantifiers in production and comprehension. *Memory & Cognition, 24*, 144–155.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual interpretation. *Cognition, 71*, 109–147.

Sells, P. (1987). Aspects of logophoricity. *Linguistic Inquiry, 18*, 445–479.

Shao, J., & Neville, H. (1998). *Analyzing semantic processing using event-related brain potentials. The newsletter of the center for research in language*. San Diego, La Jolla, CA: University of California.

Steedman, M. (1997). *Surface structure and interpretation*. Cambridge, MA: MIT Press.

Steinhauer, K., Drury, J., Portner, P., Walenski, M., & Ullman, M. T. (2007). *Brain potentials for formal semantics: Licensed vs. unlicensed negative polarity items (NPIs) poster presented at CNS (Cognitive Neuroscience Society)*. New York: Cognitive Neuroscience Society.

Stowe, L. (1986). Evidence for online gap creation. *Language and Cognitive Processes, 1*, 227–245.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language, 48*, 542–562.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language, 35*, 454–475.

Urban, S., Gunter, T. C., Friederici, A. D., & Bormann, T. (2000). Verb information im Satzverstehen. In D. F. A. Vorberg et al. (Eds.), *Experimentelle Psychologie—Beitraege zur 42. Tagung experimentellarbeitender Psychologen*. Lengerich: Pabst Science Publications.

Van Herten, M., Kolk, H., & Chwilla, D. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research, 22*, 241–255.

Van Valin, R. D., & LaPolla, R. J. (1997). *Syntax: Structure, meaning, & function*. Cambridge University Press.

Vasishth, S., Brüssow, S., Lewis, R., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science, 32*, 685–712.

Vasishth, S., Drenhaus, H., Saddy, D., & Lewis, R. (2005). Processing negative polarity. In *Talk presented at* the 18th CUNY sentence processing conference. University of Arizona.

Xiang, M., Dillon, B. W., & Phillips, C. (2006). Testing the strength of the spurious licensing effect for negative polarity items. In *Talk presented at the 19th annual meeting of the CUNY conference on human sentence processing*, March. New York.

Zribi-Hertz, A. (1989). Anaphor binding and narrative point of view: English reflexive pronouns in sentence and discourse. *Language, 65*, 695–727.