EUROSLA KEYNOTE

Structure-sensitive constraints in non-native sentence processing

Claudia Felser

Studies examining the real-time application of structure-sensitive constraints in second-language (L2) sentence processing have shown that depending on the type of constraint under investigation, the constraint may be more likely, equally (un)likely, or less likely to be violated during L2 than during native (first-language, L1) processing. Several attempts have been made in the past to attribute L1/L2 processing differences to a specific underlying cause, including cognitive resource limitations, reduced sensitivity to grammatical information, or increased susceptibility to memory interference during L2 processing. Focusing on recent findings on the processing of referential and filler-gap dependencies, I argue that trying to reduce L1/L2 processing differences to a single cause is misguided. What is called for instead is a more careful investigation of how different types of constraint and information sources interact during L2 comprehension, taking into account what linguistic cues need to be extracted from the input or need to be re-accessed in order for a given constraint to be applied. This should provide us with a more nuanced picture of how the relative weighting or timing of constraints or information sources might differ in L2 in comparison to L1 processing.

1. Introduction

The primary goal of language processing research is to uncover the mental mechanisms that allow us to derive complex meanings during reading or listening from stimuli which, at the physical level, bear no obvious resemblance to these meanings at all. In language comprehension sequences of speech sounds, strings of orthographic units, or sequences of manual and facial signs need to be segmented and analysed in a way that allows comprehenders to extract relevant grammatical, semantic, and discourse-level information for establishing accurate form-meaning mappings. Native speakers of any language can normally accomplish this feat without any conscious effort and within a matter of a few hundred milliseconds. Considering the complexity of this task, this is remarkable. Even more remarkable is the fact that people are also able to compute accurate meaning representations in a language they did not acquire as a native language (first language, L1) during infancy and childhood but which they started learning later in life as a second or additional language (L2). Research now acknowledges that language acquisition and the ability to process the input are closely intertwined (see, for example, Phillips & Ehrenhofer, 2015), and a better understanding of language learners' processing skills might also provide us with the better understanding of how languages are acquired.

In real-time language comprehension written or spoken utterances are analysed incrementally, with information extracted from the linguistic stimulus itself (bottom-up information) quickly being integrated with contextual or other top-down information (e.g., Tanenhaus et al., 1995). Incremental analysis and interpretation have also been attested for L2 comprehension (e.g., Roberts & Felser, 2011; Williams, 2006), and L2 processing has been found to be guided by top-down information, such as contextual information (e.g., Pan & Felser, 2011) and probabilistic biases (e.g., Dussias & Cramer Scaltz, 2008).

Additionally, studies of L2 sentence processing have revealed differences between L1 and L2 speakers' processing patterns (see Roberts, 2013, for a review). L1/L2 performance differences have traditionally been attributed to a lack of relevant L2 knowledge, reduced L2 experience, L1 influence, and/or age-of-acquisition effects. Recent attempts to account for L1/L2 differences in sentence-level processing include the idea that, in comparison to L1 processing, L2 processing is impeded by cognitive resource limitations (McDonald, 2006) or slower lexical access (Hopp, in press), a reduced ability to predict (Grüter et al., 2017), reduced sensitivity to grammatical information (Clahsen & Felser, 2006, 2018), or increased susceptibility to memory interference (Cunnings, 2017). These hypotheses are often difficult to disentangle empirically and are not necessarily mutually exclusive.

University of Potsdam, DE
felser@uni-potsdam.de

Based on recent findings from L2 processing studies, what I argue for in the current study is that trying to reduce L1/L2 processing differences to any specific single cause is probably misguided. A more realistic and nuanced picture of L2 learners' processing abilities can be obtained by taking into account how the ability to extract relevant grammatical and semantic properties from the input interacts with processing mechanisms such as those involved in structure-building and memory search, as well as with processing economy and probabilistic constraints.

2. Constraint interaction in language comprehension

Our ability to integrate different information sources during processing is captured by theoretical models which assume that real-time language comprehension is guided by interacting constraints of different types (Gibson & Pearlmutter, 1998). On the one hand, there are universal or language-specific linguistic constraints, including grammatical, semantic and discourse-level constraints. On the other hand, there are processing constraints which reflect general properties of our cognitive system or of the architecture and workings of the language-processing system (e.g., Gibson, 1998). Given that our processing resources are finite, these constraints will typically take the form of computational economy constraints. Yet another type of constraint are probabilistic constraints, which are shaped through language use and exposure (MacDonald et al., 1994). Probabilistic constraints may, for example, help us predict how a sentence fragment is likely to continue or decide what the most probable interpretation is for an ambiguous word or string of words, based on previous experience. Although the precise way in which our linguistic knowledge system and the language processor (or parser) interact, and the degree to which they are in fact separable, is still under debate (for a discussion, see Lewis & Phillips, 2015), I assume here that real-time parsing decisions are guided by our linguistic knowledge in conjunction with probabilistic and processing economy constraints.

The constraints that guide comprehension may differ in their relative weighting and/or timing, and, as I argue below, differences in processing performance between populations might reflect between-population differences in constraint weightings, or differences regarding the point in time at which an information source affects processing decisions. Non-native language processing may additionally be influenced by linguistic properties of comprehenders' native or prior languages (e.g., Erdocia & Laka, 2018), as well as by maturational constraints. That is, the age at which the L2 has been acquired may affect whether or not L2 comprehenders can demonstrate native-like processing performance (see Birdsong, 2018, for a review and discussion).

In the following I focus on constraints that are structure-sensitive, which is to say that they make reference to constituent structure and to hierarchical relationships between sentence constituents (such as c-command, Reinhart, 1983).

3. Structure-sensitive constraints on discontinuous dependencies

Structure-sensitive constraints play an important role in restricting the formation of intra-sentential dependencies such as referential or movement ('filler-gap') dependencies. In example (1), for instance, a constraint known as binding Condition A (Chomsky, 1981) requires that reflexive pronouns such as *herself* must be bound locally, which precludes a reading according to which *herself* is coreferential with the matrix subject *Mary*.

(1)   *Mary$_i$ feared that Sue might have hurt herself$_i$.

Similarly, when interpreting *wh*-questions, as in example (2), a constraint that belongs to a family of constraints collectively referred to as island constraints (Ross, 1967) normally prevents us from construing the fronted *wh*-pronoun *who* (the filler) as the direct object of the verb *invite*; the potential object gap is indicated by underscores.

(2)   *Who did Mary wonder why Sue had invited __ ?

Both referential and filler-gap dependencies involve constituents that are grammatically or semantically underspecified in some way and thus require licensing by another element in the current sentence or discourse in order to become fully interpretable. It is important to note that from a left-to-right processing perspective, dependency types differ, among other things, in terms of the search direction for a licenser. In the case of referential dependencies that involve anaphoric constituents (such as reflexive pronouns), encountering an anaphor triggers a backward-looking memory search for a suitable licenser (henceforth, antecedent), as indicated in (3).

(3)   Mary feared that Sue might have hurt *herself*.
      ⟵

When processing filler-gap dependencies, in contrast, encountering a fronted *wh*-element triggers a forward-looking or predictive search for a suitable licenser (as indicated in 4), with the licenser usually being a verb or preposition further downstream that lacks a complement.

(4)   *Who* did Mary say that Sue has invited __ ?
      ⟶

Structure-sensitive constraints may apply across potentially very large chunks of the sentence currently being processed. Successfully applying such constraints during real-time processing requires (i) sufficiently fast and detailed structure-building, (ii) functioning search mechanisms, including both prediction ability and the ability to re-access and navigate previously built representations, and (iii) the constraint being active in that it prevents grammatically unsuitable licensers from being considered.

Even highly proficient L2 speakers sometimes have problems with the real-time application of structure-sensitive constraints, especially where non-local dependencies are involved (e.g., Boxell & Felser, 2017;

Felser & Cunnings, 2012; Keating, 2009; Kim et al., 2015; Marinis et al., 2005). Not all types of discontinuous dependency are necessarily processed differently by L2 and L1 comprehenders, however (e.g., Aldwayan et al., 2010; Felser et al., 2012; Felser & Drummer, 2017; Jessen et al., 2017; Omaki & Schulz, 2011). Systematically comparing how L2 speakers with similar language profiles process different dependency types in parallel experimental settings and assessing whether and how their performance differs from that of native speakers might bring us closer towards identifying the likely sources of L1/L2 processing differences.

Focusing on the processing of referential and *wh*-dependencies, in the next section I examine how different types of theoretical approaches might be able to account for the observed differences and similarities between L1 and L2 processing.

4. Constraint application in L1 and L2 processing: Similarities and differences
4.1. Overview and rationale
Here I consider a selection of sentence processing studies that my collaborators and I have carried out over the course of the past decade that have all used the same experimental methodology and similar experimental designs. Together with the fact that in all of these studies we investigated post-childhood L2 learners at similarly high proficiency levels (Level B2 or above of the Common European Framework of Reference), this should make the results from these studies fairly comparable. All investigations focus on learners who were able to demonstrate native-like knowledge of the relevant constraints in complementary offline tasks. This should allow us to rule out the simplest possible explanation for L1/L2 processing differences: A lack of relevant linguistic knowledge. The testing languages were either English or German, and we always chose L1/L2 combinations that should help minimise the possibility of negative L1 influence.

The method we used for measuring participants' processing was eye-movement monitoring during reading (Clifton et al., 2007). Recording participants' eye movements during reading allows us to capture both their initial reaction to a critical word or phrase in the stimulus materials, as well as later reactions, thus providing us with fine-grained and highly time-course sensitive reading profiles. The technique allows for fairly natural reading and has proven very suitable for studying non-native language comprehension (Keating, 2014; Roberts & Siyanova-Chanturia, 2013). Interpreting eye-movement data is based on the assumption that reading times reflect processing time (Just & Carpenter, 1980); hence elevated reading times at the particular word or sentence region, relative to a control condition, are taken to indicate processing difficulty at that region.

The linguistic phenomena to be focused on include binding Conditions A–C (Chomsky, 1981), island constraints (Ross, 1967), and the hybrid phenomenon of strong crossover (Postal, 1971). To determine whether a structure-sensitive constraint is successfully applied

during comprehension, we have used experimental designs in which an inappropriate licenser is provided which serves as a bait in that it has a property or properties that might make it an attractive candidate for licensing the dependent element. In the case of referential dependencies, we provide a grammatically inappropriate but feature-matching antecedent, and in the case of filler-gap dependencies, we provide an inappropriate lexical subcategoriser. For illustration, consider the minimal sentence pair in (5), which differs in the gender conventionally associated with the proper names *Mary* versus *Brian*, the matrix subjects.

(5)   a.   *Mary* feared that Sue might have hurt herself.
      b.   *Brian* feared that Sue might have hurt herself.

The only legitimate antecedent for *herself* in both (5a) and (5b) is the local one, the proper name *Sue*. If readers link the reflexive to *Sue* without considering the matrix subject as a possible antecedent, we would expect to see no differences in reading times at the reflexive in (5a) versus (5b). However, if any reading-time differences are observed, then these can only possibly stem from participants' considering the inappropriate antecedent (*Mary*/*Brian*) and reacting to the perceived gender (mis-)match between *Mary* or *Brian* and the reflexive.

For each of the main findings to be reviewed below, I compare how well different accounts for L1/L2 processing differences fare in explaining the observed processing pattern, focusing on three proposals:

· processing resource limitation accounts (e.g., McDonald, 2006),
· the memory interference account (Cunnings, 2017), and
· the revised shallow structure hypothesis (Clahsen & Felser, 2018).

This makes it necessary to operationalise these accounts in ways that may not always do full justice to them. For example, the hypothesis that L1/L2 processing differences reflect processing resource limitations in L2 comprehension (e.g., McDonald, 2006) might predict that L2 speakers should show a general delay in constraint application as they take longer or have more difficulty than L1 speakers to process the input and compute the kind of representations over which the constraint is defined. Another prediction is that L2 comprehenders should generally strive to keep dependencies as short as possible, and potentially at the cost of violating a structure-sensitive constraint. A preference for minimising dependency length (MDL) is well attested in the monolingual processing literature and is usually attributed to computational resource limitations (e.g., Gibson, 1998). On the assumption that processing resources are more easily exhausted in L2 than in L1 comprehension, we might expect L2 speakers to have more difficulty establishing non-local dependencies and to be more likely to consider inappropriate local licensers than L1 speakers.

From the perspective of the memory interference account (Cunnings, 2017) the direction of search for a licenser would be expected to be a crucial factor. If L2 comprehenders are particularly vulnerable to memory retrieval interference, then they should have more difficulty applying constraints on backward-looking than on forward-looking dependencies, where a licenser is actively predicted (compare also Felser, 2015). The third account under consideration, the shallow structure hypothesis, proposes that the weighting and/or timing of grammatical and non-grammatical information differs between L1 and L2 comprehension (Clahsen & Felser, 2006, 2018; Felser, 2016). It predicts that constraint application should be successful during L2 comprehension only if the relevant grammatical cues can be processed and used effectively for structure-building, prediction, or memory search.

4.2. Referential dependencies

Within formal linguistics, structure-sensitive constraints on intra-sentential referential dependencies form part of the binding theory (Chomsky, 1981). To examine the real-time application of binding Condition A by proficient L1 German-speaking learners of English, Felser and Cunnings (2012, Experiment 1) used stimulus materials such as (6) below. We manipulated the conventional or stereotypical gender match both between the reflexive (*herself* versus *himself*) and its legitimate antecedent *the nurse* and between the reflexive and a grammatically inappropriate competitor antecedent (*Susan/she* versus *Peter/he*).

(6)   Susan (Peter) waited for ages in the county hospital. She (He) knew that the nurse had prepared herself (himself) for the operation in the evening. It was going to be a long and complicated procedure.

The analysis of the eye-movement data revealed that unlike the L1 English-speaking controls, who linked the reflexive to its correct antecedent *the nurse* as soon as they read the reflexive, the L2 participants initially tried to link it to the competitor antecedent (see also Felser et al., 2009). In a follow-up experiment, Felser and Cunnings (2012) found this to be the case even if the competitor antecedent did not c-command the reflexive, indicating that the referential dependency that the L2 speakers attempted to form did not involve syntactically mediated binding. Only at later processing stages did the L2 group home in on the correct antecedent *the nurse*. This finding is consistent with the hypotheses that L2 comprehenders are more prone to memory interference than L1 speakers and that they are less sensitive than L1 speakers to grammatical cues such as c-command. It also shows that L2 comprehenders do not necessarily prioritise on the minimise dependency length constraint when trying to establish referential dependencies, as this would have favoured the correct (local) antecedent.

Binding Condition B constrains the interpretation of personal pronouns such that a pronoun may not be bound by a local antecedent, as indicated through co-indexation in (7).

(7)   Simon$_i$ was happy that Peter$_k$ had reserved him$_{i/*k}$ a seat on the train home.

Pronouns in some syntactic environments seem to be exempt from Condition B, however. For example, pronouns appearing with spatial prepositions in sentences such as (8) can optionally be linked to a local referent (here, *Ryan*).

(8)   Andy$_i$ noticed Ryan$_k$ place a chair next to him$_{i/k}$ at the front of the hall.

Examining online pronoun resolution in sentences such as (7) and (8), Patterson et al. (2014) found that both English native speakers and L1 German-speaking learners of English behaved in accordance with binding Condition B in that they considered only the non-local antecedent (*Simon*) in sentences such as (7). However, L1/L2 differences were seen in participants' eye-movement patterns for sentences such as (8). Here the L2 group again only considered the non-local antecedent (*Andy* in example 8), whereas the native speakers also considered the local one as a potential antecedent. On the face of it, this looks like the L2 speakers over-extended Condition B to sentences in which this constraint does not in fact apply.

Puebla et al. (in preparation) report eye-movement evidence indicating that Condition B incompatible antecedents might be more likely to be considered in native than in non-native processing. Their materials included German sentences such as (9) in which an embedded object pronoun (*ihn* 'him') can be linked to the matrix subject (*Florian*) but not to the subject of the embedded clause. The inappropriate local antecedent was rendered more salient than was the case in Patterson et al.'s (2014) study through modification (*der Kollege aus Frankreich* 'the colleague from France').

(9)   *Florian glaubte, dass der Kollege   aus*
       F.   believed that the colleague$_{masc}$ from
       Frankreich ihn bald vorstellen würde.
       France   him soon introduce would
       'Florian believed that the colleague from France would introduce him soon.'

The authors found that native German speakers – but not L1 Russian-speaking L2 learners of German – initially violated Condition B by considering the inappropriate local antecedent.

L2 speakers' ability to ignore inappropriate local antecedents is unexpected from the point of view of the memory interference hypothesis. It also provides further evidence that the MDL constraint does not carry any particularly strong weight in L2 processing. Our findings on both Condition A and Condition B can be reconciled by the hypothesis that referential dependency formation is more strongly guided by discourse-level cues, such as topic-hood, and less by structural cues such as c-command, in L2 compared to L1 comprehension (Clahsen & Felser, 2018; Cunnings, 2017; Felser, 2016). That is, the most accessible antecedent in the current discourse representation is likely to be initially retrieved

by L2 comprehenders, which may or may not also be a grammatically appropriate antecedent.

Let us now turn to cataphoric pronouns, whose interpretation in certain syntactic configurations is thought to be constrained by binding Condition C. This constraint prohibits coreference between a referring expression (such as a proper name) and a pronoun that c-commands it, as in the German example (10a). However, in the absence of a c-command relation (as in 10b, where the cataphoric pronoun serves as a possessive adjective), coreference between a cataphoric pronoun and a following name should be permitted.

(10) a. *Er$_i$ war begeistert, weil   Georg$_{*i}$ so gut*
        he   was excited   because G.   so well
        spielen konnte.
        play   could

     b. *Sein$_i$ Lehrer war begeistert, weil   Georg$_i$ so*
        his   teacher was excited   because G.   so
        gut spielen konnte.
        well play   could
        'He (his teacher) was excited because George was able to play so well.'

Encountering a cataphoric pronoun has been shown to trigger a forward-looking search for a suitable antecedent both during L1 and L2 comprehension (Drummer & Felser, 2018, Experiment 1). Investigating the real-time application of Condition C in German, Drummer and Felser (2018, Experiment 4) found that both native speakers and L1 Russian-speaking learners of German only applied this constraint at later processing stages. Both groups initially considered the named referent (here, *Georg*) as a potential antecedent for the pronoun regardless of whether or not the pronoun c-commanded the name.

This finding suggests that the MDL constraint guides forward-looking referential dependency formation in the same way in both L1 and L2 comprehension, with the constraint on interpretation commonly referred to as Condition C serving as a later filter that leads to inappropriate coreference relationships being discarded.

4.3. Wh-movement dependencies
In *wh*-movement languages such as English, *wh*-elements can potentially be fronted across considerable distances. *Wh*-fronting is however restricted by so-called island constraints, which prohibit extracting a *wh*-element from a sentence region under certain conditions (Ross, 1967). For illustration consider example (11), where encountering the relative complementiser *that* will trigger a search for a suitable lexical licenser for the relativised noun phrase (NP) *the bike*, such as a transitive verb, within the relative clause (RC) that modifies *the bike* (RC1).

(11) This is *the bike* [$_{RC1}$ that the cyclist [$_{RC2}$ who rode extremely quickly] saw __ today].

From the perspective of the MDL constraint, the first available potential licenser is the verb *ride*; however, since

this is embedded within another relative clause (= RC2), and RCs are considered to be extraction islands, construing *the bike* as the object of *ride* should not be possible here. Note that despite containing an island region, example (11) is perfectly grammatical and interpretable, as long as comprehenders respect the RC island and manage to establish a link between *the bike* and its true subcategoriser (here, the verb *see*) instead.

Using stimulus materials similar to (11), Felser et al. (2012) examined the timing of constraint application in English native speakers and L1 German-speaking learners of English. In one of their eye-movement experiments, semantic fit was manipulated as an experimental diagnostic for object-dependency formation, such that construing the relativized NP as the direct object of *ride* was either plausible or not (e.g., *the bike …rode* vs. *the pond … rode*). In this experimental paradigm, elevated reading times in the implausible (*the pond … rode*) compared to the plausible direct-object condition (*the bike …rode*) would reveal that dependency formation was attempted, in violation of the RC island constraint. The results however showed that both the L1 and the L2 group respected RC islands during processing, as plausibility effects were observed only for control sentences that did not contain an RC island (e.g., *the bike/pond that the cyclist rode…*); see Omaki and Schulz (2011) for similar findings from self-paced reading. This suggests, once again, that the MDL constraint is not weighted any more strongly in L2 than in L1 comprehension and that both groups were able to recognise the island boundary signalled by the presence of the *wh*-pronoun *who* in (11).

Somewhat different findings have been reported by Boxell and Felser (2017) for another type of island. Here we examined whether native and non-native speakers of English would respect complex subject islands during processing. Extraction from complex subjects such as the bracketed NP in (12) is not normally permitted.

(12) *The policeman knew *which prisoners* [$_{NP}$ the activities that inspired __] would …

This constraint should prevent comprehenders from trying to construe *which prisoners* as the object of the embedded verb *inspire*. Boxell and Felser's (2017) study exploits the phenomenon of parasitic gaps, which under certain conditions allows for subject islands to become permeable (Engdahl, 1983). As was first demonstrated by Phillips (2006) for online processing, native English-speaking comprehenders allow for *which prisoners* to be linked to the verb *inspire* if the embedded finite RC in (12) (*that inspired…*) is replaced by a non-finite complement clause as in (13) (*to inspire…*), but not otherwise (Given space limitations, I do not discuss how parasitic gaps may be formally represented and accounted for.).

(13) The policeman knew *which prisoners* [$_{NP}$ the activities to inspire __] would help __.

Again, a plausibility manipulation was used (*which prisoners … inspire* versus *which houseplants … inspire*) to test

for attempted dependency formation between the fronted *wh*-phrase and the verb inside the island region. While Boxell and Felser (2017) could replicate Phillips' (2006) finding for native speakers, L1 German-speaking learners of English were found to violate the subject island constraint during their initial reading of the critical post-gap sentence region. Unlike English native speakers, they initially tried to link *which prisoners* to the verb *inspire* irrespectively of the finiteness manipulation. The L2 group's eye-movement patterns revealed sensitivity to the constraint during somewhat later processing stages, however. Their regression-path times at the post-gap region revealed a plausibility effect for permissible parasitic gaps as in (13) only, but no plausibility effect at the corresponding sentence region in finite environments as in (12). No between-group differences were observed in rereading and total reading times either, suggesting that the L2 group's violation of the subject island constraint was indeed only fleeting.

How might the apparent discrepancy between Felser et al.'s (2012) and Boxell and Felser's (2017) findings be explained? We saw earlier that L2 processing does not appear to be guided any more strongly than L1 processing by the MDL constraint. Yet, in Boxell and Felser's (2017) study, L2 comprehenders did initially go for the first available potential subcategoriser (the verb *inspire* in examples 12 and 13), even where this should have been prohibited by the subject island constraint. The difference in timing between the two types of island constraint we examined cannot be accounted for by the memory interference hypothesis, which claims that L2 comprehenders are able to build syntactic representations of the kind required for island constraints to be applied and will only attempt filler-gap dependency formation if the dependency is grammatically licensed (Cunnings, 2017, p. 667).

The hypothesis that L2 comprehenders have a reduced sensitivity to grammatical cues in the input (Clahsen & Felser, 2006, 2018) offers an explanation for the observed timing differences. A reduced or delayed ability to extract relevant grammatical information from the input may make L2 comprehenders more likely than L1 comprehenders to initially compute incomplete or shallow syntactic representations. Note that for the kind of subject islands examined by Boxell and Felser (2017), recognising the island region requires processing and integrating several grammatical cues in the input. Comprehenders must realise that the NP headed by *activities* in both (12) and (13) functions as a subject, whilst *which prisoners* serves as an object (even though the *wh*-phrase may initially have been mistaken for a subject). Crucially, in addition to this, comprehenders need to recognise that in example (12), *activities* is modified by a finite RC whose *wh*-operator is not expressed overtly. Together these properties render the bracketed NP in (12) impermeable.

Considering this fairly complex set of islandhood-inducing cues (recall that replacing the finite RC with a non-finite complement clause does in fact make the island region permeable), it does not seem particularly surprising that the L2 group should have taken somewhat longer than the L1 group to identify the island region and to realise that linking *which prisoners* to *inspire* in (12) is

inappropriate. If islandhood cues are not recognised or processed quickly enough, the MDL constraint will lead the parser to form a link between the fronted *wh*-phrase and the first potential licenser it comes across (the verb *inspire*), which is precisely what we observed. In contrast, in the materials used by Felser et al. (2012) to examine the timing of the RC island constraint, the start of the island region was overtly signalled by the appearance of the *wh*-pronoun *who*, a comparatively obvious cue which we might expect L2 comprehenders to be able to notice easily.

4.4. A hybrid phenomenon: Strong crossover
In formal linguistics, the term crossover refers to phenomena where the fact that syntactic movement has crossed a pronoun affects the possibility of coreference between the fronted constituent and the pronoun (Postal, 1971). In so-called weak crossover (WCO) configurations as in (14a), coreference between *which girl* and the pronoun *her* tends to be considered as less acceptable than in the corresponding non-crossover configuration in (14b).

(14) a. $^?$ [Which girl]$_i$ did her$_i$ teacher praise __ ?
      b.   Her$_i$ teacher praised [the new girl]$_i$.

Strong crossover (SCO) configurations are created if the pronoun c-commands the fronted constituent's canonical or base position, as is the case in (15a).

(15) a.   * [Which girl]$_i$ did she$_i$ praise __ ?
      b.   * She$_i$ praised [the new girl]$_i$.

The unavailability of a coreferential reading in (15a) has been attributed to binding Condition C (but cf. Chomsky, 1982), on the assumption that the fronted constituent is reconstructed at its base position at some level of syntactic representation. A Condition C violation is more obvious in (15b), the corresponding non-crossover configuration, where the cataphoric pronoun c-commands the referring expression. The current majority view, though, seems to be that crossover and Condition C effects reflect semantic or pragmatic constraints (Huang, 2000; Schlenker, 2005). Note that from a processing point of view, crossover configurations constitute a hybrid case as they involve both the resolution of a *wh*-dependency and pronoun resolution, with the former process potentially affecting the latter.

When testing native German speakers' and L1 Russian-speaking learners' sensitivity to crossover constraints during online processing, Felser and Drummer (2017) found that both the L1 and L2 comprehenders immediately ruled out coreference between the pronoun and a fronted *wh*-constituent in SCO (16a) but not in WCO configurations (16b).

(16) a. Über *welchen Musiker   aus   Amerika er*
          about which   musician$_{masc}$ from America he
                bald schreiben würde, das   erzählte Georg
          soon write   would that told   G.
                aufgeregt.
          excitedly   (SCO configuration)

b. Über *welchen Musiker   aus   Amerika sein*
about which musician~masc~ from America his
Biograph bald   schreiben schreiben
biographer soon write   write
würde, das   erzählte Georg aufgeregt.
would that told   G.   excitedly
                                    (WCO configuration)
'Which musician from America he/his bio-
grapher would soon write about, this George
related excitedly.'

This indicates that both groups were sensitive to the configurational difference between (16a) and (16b), and that both were able to apply the SCO constraint when encountering the subject pronoun *er* 'he' in sentences like (16a), whilst allowing for coreference in (16b). At no point during processing did Felser and Drummer's (2017) participants seem to consider the *wh*-expression within the fronted prepositional phrase in (16a) as an antecedent for a following subject pronoun. This finding is unexpected if constraint application during L2 processing were affected by processing resource limitations, as the MDL constraint would have favoured linking the pronoun to the *wh*-phrase in both WCO and SCO configurations.

Given that both L1 and L2 comprehenders only applied binding Condition C with some delay in Drummer and Felser's (2018) study, it seems surprising that both participant groups immediately ruled out inappropriate coreference relationships in SCO configurations in Felser and Drummer's (2017) study. Recall that in both studies, the critical experimental manipulation concerned the pronoun's relative prominence (or c-command domain), allowing possessive but not subject pronouns to enter into a coreference relationship with another sentence participant. The stimulus materials differed, however, in that the pronoun was cataphoric in Drummer and Felser's (2018) study since it preceded its potential antecedent, as indicated in (17a), whereas it followed a potential antecedent phrase in Felser and Drummer's (2017) crossover study, as indicated in (17b).

(17) a. pronoun … name
     b. *wh*-expression … pronoun … *wh*-gap

The relative ease with which both L1 and L2 comprehenders were able to rule out inappropriate referential dependencies in configurations like (17b) is less surprising if we view SCO and Condition C effects as reflecting a pragmatic constraint against coreference between co-arguments, such as Huang's (2000) disjoint reference presumption (DRP). This constraint tells us to interpret co-arguments as referring to distinct entities unless a co-argument is marked as reflexive.

Note that in both configurations (17a) and (17b), a subject pronoun should be easily recognisable as such and also easy to distinguish from pronouns used as possessive adjectives, even for non-native comprehenders. Unlike subject pronouns, possessive adjectives are not arguments of the verb in the clause they occur in. When processing crossover sentences from left to right, the parser's assumption will be that the fronted *wh*-phrase in (16) denotes a participant in the event described by the (yet-to-be-received) verb. The DRP will then prevent the parser from attempting to form a referential link between the *wh*-phrase and the subject of the same verb. In contrast, in order to rule out coreference in standard Condition C configurations as in (10a), the potential antecedent's referential status needs to be ascertained before the DRP can be applied, which in Drummer and Felser's (2018) study was reflected in delayed constraint application.

4.5. Summary
The above set of findings show that even if the choice of experimental method and design and non-native participants' L2 proficiency level are kept the same, structure-sensitive constraints may be either more likely, equally (un-)likely, or less likely to be violated during L2 than during L1 processing (see **Table 1** for an overview of the findings). The observed L1/L2 processing differences and similarities show that the application of structure-sensitive constraints is not necessarily more problematic in L2 than in L1 processing.

In the following, I discuss to what extent different hypotheses about L1/L2 processing differences are able to account for the observed pattern of findings.

5. Discussion
The results from the above selection of eye-movement monitoring studies revealed that both L1 and L2 comprehenders' ability to suppress unsuitable licensers

Table 1: Overview of the findings discussed in section 4.

Constraint	Violated in L1 processing	Violated in L2 processing
Condition A	no	yes
Condition B	yes	no
Condition C	yes	yes
Strong crossover	no	no
RC islands	no	no
Subject islands	no	yes

Note: Constraint violation here refers to experimental effects that indicate that an inappropriate licenser for a dependent element was initially considered during processing.

during processing varies depending on the type of dependency involved. How can the observed L1/L2 differences be accounted for?

Let us first consider the hypothesis that L1/L2 processing differences come about due to cognitive resource limitations disproportionally affecting L2 comprehension (McDonald, 2006). If L2 processing were generally slower or less efficient than L1 processing we might have expected the application of structure-sensitive constraints to be generally delayed during L2 in comparison to L1 processing. This was not the case, however. As we saw above, some constraints did not differ in their timing between L1 and L2 speakers, and effects of the RC island and the SCO constraints were already visible in early processing measures in both groups. From the perspective of resource limitation accounts, we might also have expected the MDL constraint to be more strongly weighted in L2 than in L1 processing, as keeping dependencies short is thought to save processing resources (e.g., Gibson, 1998). This should have made binding Condition A easier for L2 comprehenders to apply than any of the other constraints under investigation, because Condition A is the only constraint that we tested that requires the dependent element to be linked to a local licenser. Again, this is not what we observed.

Next, consider the hypothesis that L2 comprehension is more prone to memory interference than L1 comprehension, a problem which we would expect to primarily affect backward-looking dependencies, where an appropriate licenser needs to be selected from the set of potential licensers previously encountered (or presupposed if not explicitly mentioned). Whilst this hypothesis would account for the delayed application of binding Condition A, the lack of early interference effects in our Condition B studies is somewhat unexpected. The memory interference account offers no straightforward explanation for the observed difference in timing between L2 comprehenders' application of the RC island constraint and the subject island constraint either. The interference hypothesis does not specifically seem to predict any L1/L2 differences in the application of binding Condition C, which is consistent with what we observed. Its predictions regarding L2 comprehenders' sensitivity to crossover constraints in configurations such as (17b) are less clear. Given that during the processing of crossover configurations, pronoun resolution interacts with *wh*-dependency resolution, we might have expected L2 comprehenders to experience interference from the *wh*-expression preceding the pronoun. However, this is not what we observed. Both L1 and L2 speakers showed immediate sensitivity to the SCO constraint and considered a potential antecedent NP within the *wh*-phrase only in WCO configurations, where coreference was independently shown to be permitted.

Finally, let us turn to the hypotheses that comprehenders' sensitivity to grammatical cues in the input is reduced in L2 in comparison to L1 processing and that this may be compensated for by a correspondingly increased reliance on semantic and discourse-level cues to interpretation. This would account for our observation that when resolving reflexives or anaphoric pronouns, L2 comprehenders are initially drawn to the most prominent potential antecedent – the matrix subject or discourse topic – before homing in on, in the case of reflexives, the grammatically appropriate local antecedent. An initial preference for prominent antecedents, reflecting increased reliance on discourse-level cues in L2 processing, will disfavour a (less-prominent) local antecedent. This would explain why the L2 speakers tested by Patterson et al. (2014) failed to consider a local antecedent for pronouns even in Condition B exempt environments, despite allowing for pronouns in such environments to be linked to a local antecedent in a complementary offline task.

Given that respecting island regions during processing requires comprehenders to process the relevant islandhood-inducing cues, the hypothesis that L2 speakers may have a reduced sensitivity to grammatical cues in the input also offers a possible account for the observed timing differences between the RC and the subject island constraint. When taking into account the complexity and salience of the respective islandhood cues, we noted that the RC island boundary in Felser et al.'s (2012) study was signalled by a comparatively salient cue, an overt *wh*-pronoun. In contrast, subject islands of the kind examined by Boxell and Felser (2017) are signalled through a combination of several cues which cannot easily be read off the incoming string of words, so that the presence of these cues (and thus, the presence of an island region) may not be immediately obvious to L2 comprehenders. If relevant islandhood-inducing cues are initially missed, other constraints such as the MDL and the presence of a potential lexical licenser in the current input will guide the decision of when to initiate dependency formation.

Regarding the processing of cataphoric pronouns, we found both L1 and L2 comprehenders to behave very similarly. Both groups of comprehenders applied binding Condition C with a slight delay in Drummer and Felser's (2018) study, and both applied the SCO constraint immediately in Felser and Drummer's (2017) study. I suggested above that these seemingly inconsistent findings can be accounted for if we assume that referential processing is guided by a pragmatic constraint against forming referential links between (potential) co-arguments and that a possible violation of this constraint is easier to anticipate – and thus to avoid – in SCO than in standard Condition C environments.

While the hypothesis that relative cue weightings might differ in L1 versus L2 processing can account for several of the above findings, a global statement to the effect that L2 comprehenders lack sensitivity to grammatical information in the input would be too simplistic. Sensitivity to c-command, for example, seems to be reduced during L2 reflexive processing but not when applying the SCO constraint. Thus, it might be that re-accessing previously built syntactic representations (and retracing command paths) during memory search is more difficult than determining the structural prominence or discourse function of the constituent currently being processed, for example. As was noted above, the relative salience of grammatical information in the input may also influence the likelihood of this information being processed quickly enough for a constraint violation to be avoided.

Assuming that incoming strings of words are transformed into representations of different kinds (e.g., syntactic, semantic, and discourse-level) during processing, it is conceivable that detailed structure-building – but not the computation of discourse-level representations – lags behind in L2 versus L1 comprehension, or that the syntactic representations computed during L2 processing are more prone to instability than during L1 processing. This may then lead L2 comprehenders to try and establish backward-looking dependencies at the discourse-representational level rather than through navigating hierarchical phrase-structure representations. This hypothesis would account for the findings on L2 anaphor resolution discussed above, and it is also consistent with the finding reported by Trompelt and Felser (2014) that L2 speakers are more likely than L1 speakers to link an ambiguous anaphoric pronoun to a non c-commanding coreference antecedent than to a c-commanding variable binder during real-time processing. While variable binding is mediated by hierarchical structural configurations, discourse-based coreference assignment is not (e.g., Reuland, 2001).

In short, what the above pattern of findings shows is that applying structure-sensitive constraints during L2 processing may be problematic if the structural representation over which the constraint is defined has not been computed accurately or fast enough, or has faded from memory, at the point during processing when the constraint becomes relevant. With multiple constraints continuously interacting over time during processing, structure-sensitive constraints may be temporarily overridden by discourse-level or computational economy constraints. Note that this scenario is by no means specific to L2 processing; it may merely be more likely to happen during L2 compared to L1 processing due to reduced automaticity of grammatical processing routines in the L2, interference from the L1 (which the above studies did not investigate), or other factors. Investigating language processing from the perspective of interacting gradient constraints might also provide a useful way of capturing individual differences among language learners, an issue that has been attracting growing attention in recent years (see Kidd et al., 2018, for a review and discussion).

Nevertheless, a much larger sample of processing studies will be needed before we can draw any firm conclusions about L1/L2 differences and similarities in constraint or cue weighting. The studies reviewed above were not designed to examine L2 comprehenders' ability to predict, for example, or the extent to which L2 processing is guided by probabilistic information. Moreover, since we only tested L2 speakers who had demonstrated sensitivity to the constraint under investigation in offline tasks, and because we only tested language combinations for which there was no conflict between the L1 and L2 with regard to how a constraint was instantiated, the above results do not tell us anything about how or when a given constraint is acquired.

6. Conclusion

Examining how and to what extent L2 speakers can use their linguistic knowledge during real-time processing can provide us with a more comprehensive picture of their L2 mastery than can be obtained from offline or metalinguistic tasks alone. Focusing on L2 comprehenders' ability to apply structure-sensitive constraints during processing, I showed that in both L1 and L2 processing, structure-sensitive constraints may be violated initially and that L1/L2 differences in real-time constraint application cannot easily be reduced to general processing resource limitations or to any specific processing problem. A more fruitful approach to understanding both L1/L2 processing differences and similarities, in my view, is to carefully consider, for each phenomenon under investigation, the interplay between relevant linguistic cues in the input, the processing mechanisms involved (such as cue-based memory search, or prediction), processing economy constraints, and probabilistic constraints. Some of the reported findings point towards different weighting and/or timing of grammatical and discourse-level information in L1 versus L2 processing, and theoretical models that allow for gradience and variable constraint weightings (e.g., Smolensky et al., 2014) might prove useful for capturing this kind of cross-population variability.

References

Aldwayan, S., Fiorentino, R., & Gabriele, A. (2010). Evidence of syntactic constraints in the processing of wh-movement: A study of Najdi Arabic learners of English. In B. VanPatten, & J. Jegerski (Eds.), *Research in second language processing and parsing* (pp. 65–86). Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/lald.53.03ald

Birdsong, D. (2018). Plasticity, variability and age in second language acquisition and bilingualism. *Frontiers in Psychology*, *9*, 81. DOI: https://doi.org/10.3389/fpsyg.2018.00081

Boxell, O., & Felser, C. (2017). Sensitivity to parasitic gaps inside subject islands in native and non-native sentence processing. *Bilingualism: Language and Cognition, 20*(3), 494–511. DOI: https://doi.org/10.1017/S1366728915000942

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht: Foris.

Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding.* Cambridge, MA: MIT Press.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(1), 3–42. DOI: https://doi.org/10.1017/S0142716406060024

Clahsen, H., & Felser, C. (2018). Some notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition, 40*(3), 693–706. DOI: https://doi.org/10.1017/S0272263117000250

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P.

G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: a window on mind and brain* (pp. 341–372). Amsterdam: Elsevier. DOI: https://doi.org/10.1016/B978-008044980-7/50017-3

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, *20*(4), 659–678. DOI: https://doi.org/10.1017/S1366728916000675

Drummer, J.-D., & Felser, C. (2018). Cataphoric pronoun resolution in native and non-native sentence comprehension. *Journal of Memory and Language*, *101*, 97–113. DOI: https://doi.org/10.1016/j.jml.2018.04.001

Dussias, P. E., & Cramer Scaltz, T. R. (2008). Spanish-English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta Psychologica*, *128*, 501–513. DOI: https://doi.org/10.1016/j.actpsy.2007.09.004

Engdahl, E. (1983). Parasitic gaps. *Linguistics and Philosophy*, *6*, 5–34. DOI: https://doi.org/10.1007/BF00868088

Erdocia, K., & Laka, I. (2018). Negative transfer effects on L2 word order processing. *Frontiers in Psychology*, *9*:337. DOI: https://doi.org/10.3389/fpsyg.2018.00337

Felser, C. (2015). Native vs. non-native processing of discontinuous dependencies. *Second Language*, *14*, 5–19.

Felser, C. (2016). Binding and coreference in non-native language processing. In A. Holler, & K. Suckow (Eds.), *Empirical perspectives on anaphora resolution* (pp. 229–253). Berlin, Boston: De Gruyter Mouton.

Felser, C., & Cunnings, I. (2012). Processing reflexives in English as a second language: The role of structural and discourse-level constraints. *Applied Psycholinguistics*, *33*(3), 571–603. DOI: https://doi.org/10.1017/S0142716411000488

Felser, C., Cunnings, I., Batterham, C., & Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition*, *34*(1), 67–98. DOI: https://doi.org/10.1017/S0272263111000507

Felser, C., & Drummer, J.-D. (2017). Sensitivity to crossover constraints during native and non-native pronoun resolution. *Journal of Psycholinguistic Research*, *46*(3), 771–789. DOI: https://doi.org/10.1007/s10936-016-9465-8

Felser, C., Sato, M., & Bertenshaw, N. (2009). The on-line application of Binding Principle A in English as a second language. *Bilingualism: Language and Cognition*, *12*(4), 485–502. DOI: https://doi.org/10.1017/S1366728909990228

Gibson, E. (1998). Syntactic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1–75. DOI: https://doi.org/10.1016/S0010-0277(98)00034-1

Gibson, E., & Pearlmutter, N. J. (1998). Constraints on sentence comprehension. *Trends in Cognitive Science*, *2*, 262–268. DOI: https://doi.org/10.1016/S1364-6613(98)01187-5

Grüter, T., Rohde, H., & Schafer, A. J. (2017). Coreference and discourse coherence in L2: The roles of grammatical aspect and referential form. *Linguistic Approaches to Bilingualism*, *7*(2), 199–229. DOI: https://doi.org/10.1075/lab.15011.gru

Hopp, H. (in press). The bilingual mental lexicon in L2 sentence processing. *Second Language*.

Huang, Y. (2000). *Anaphora. A cross-linguistic study.* Oxford: Oxford University Press.

Jessen, A., Festman, J., Boxell, O., & Felser, C. (2017). Native and non-native speakers' brain responses to filled indirect object gaps. *Journal of Psycholinguistic Research*, *46*(5), 1319–1338. DOI: https://doi.org/10.1007/s10936-017-9496-9

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *85*, 109–130. DOI: https://doi.org/10.1037/0033-295X.87.4.329

Keating, G. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, *59*, 503–535. DOI: https://doi.org/10.1111/j.1467-9922.2009.00516.x

Keating, G. (2014). Eye-tracking with text. In J. Jegerski, & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 69–92). New York: Routledge.

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. DOI: https://doi.org/10.1016/j.tics.2017.11.006

Kim, E., Baek, S., & Tremblay, A. (2015). The role of island constraints in second language sentence processing. *Language Acquisition*, *22*, 384–416. DOI: https://doi.org/10.1080/10489223.2015.1028630

Lewis, S., & Phillips, C. (2015). Aligning grammatical theories and language processing models. *Journal of Psycholinguistic Research*, *44*(1), 27–46. DOI: https://doi.org/10.1007/s10936-014-9329-z

MacDonald, M. C., Pearlmutter, M., & Seidenberg, M. (1994). The lexical nature of ambiguity resolution. *Psychological Review*, *101*, 676–703. DOI: https://doi.org/10.1037/0033-295X.101.4.676

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, *27*(1), 53–78. DOI: https://doi.org/10.1017/S0272263105050035

McDonald, J. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*, 381–401. DOI: https://doi.org/10.1016/j.jml.2006.06.006

Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second language sentence processing. *Studies in Second Language Acquisition*, *33*, 563–588. DOI: https://doi.org/10.1017/S0272263111000313

Pan, H.-Y., & Felser, C. (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua*, *121*, 221–236. DOI: https://doi.org/10.1016/j.lingua.2010.08.003

Patterson, C., Trompelt, H., & Felser, C. (2014). The online application of binding condition B in native

and non-native pronoun resolution. *Frontiers in Psychology* 5:147. DOI: https://doi.org/10.3389/fpsyg.2014.00147

Phillips, C. (2006). The real-time status of island phenomena. *Language, 82,* 795–823. DOI: https://doi.org/10.1353/lan.2006.0217

Phillips, C., & **Ehrenhofer, L.** (2015). The role of language processing in language acquisition. *Linguistic Approaches to Bilingualism, 5*(4), 409–453. DOI: https://doi.org/10.1075/lab.5.4.01phi

Postal, P. (1971). *Cross-over phenomena.* New York: Holt, Rinehart and Winston.

Puebla, C., Patterson, C., & **Felser, C.** (in preparation). The application of binding Condition B during native and non-native pronoun resolution in German. Ms. University of Potsdam.

Reinhart, T. (1983). *Anaphora and semantic interpretation.* London: Croom Helm.

Reuland, E. (2001). Primitives of binding. *Linguistic Inquiry, 32,* 439–492. DOI: https://doi.org/10.1162/002438901750372522

Roberts, L. (2013). Sentence processing in bilinguals. In R. P. G. van Gompel (Ed.), *Sentence Processing* (pp. 221–246). Hove: Psychology Press.

Roberts, L., & **Felser, C.** (2011). Plausibility and recovery from garden-paths in second-language sentence processing. *Applied Psycholinguistics, 32,* 299–331. DOI: https://doi.org/10.1017/S0142716410000421

Roberts, L., & **Siyanova-Chanturia, A.** (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition, 35*(2), 213–235. DOI: https://doi.org/10.1017/S0272263112000861

Ross, J. (1967). *Constraints on variables in syntax.* (Unpublished doctoral dissertation). Cambridge, Massachusetts: Massachusetts Institute of Technology.

Schlenker, P. (2005). Non-redundancy: Towards a semantic reinterpretation of binding theory. *Natural Language Semantics, 13*(1), 1–92. DOI: https://doi.org/10.1007/s11050-004-2440-1

Smolensky, P., Goldrick, M., & **Mathis, D.** (2014). Optimization and quantization in gradient symbol systems: a framework for integrating the continuous and the discrete in cognition. *Cognitive Science, 38*(6), 1102–1138. DOI: https://doi.org/10.1111/cogs.12047

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. DOI: https://doi.org/10.1126/science.7777863

Trompelt, H., & **Felser, C.** (2014). Variable binding and coreference in non-native pronoun resolution. In W. Orman, & M. J. Valleau (Eds.), *BUCLD 38: Proceedings of the 38th annual Boston University Conference on Language Development* (pp. 471–483). Somerville, MA: Cascadilla Press.

Williams, J. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition, 9*(1), 71–88. DOI: https://doi.org/10.1017/S1366728905002385