

PART II
SENTENCE STRUCTURE

Heritage language gaps

Maria Polinsky

3.1 Introduction

Researchers of heritage languages (HLs) seek to understand and model the linguistic representations (mental grammars) of heritage speakers. In pursuing this goal, we compare a particular HL to its baseline, the language of exposure that constituted the heritage speakers' input. This comparison should eventually allow us to understand the main mechanisms driving the construction of mental grammars in natural language.

To discuss the nature of underlying representations in HLs, I will focus here on the alternation between overt and silent elements in HLs. In a nutshell, the problem is as follows: as compared to the baseline, HL speakers have difficulty in detecting and interpreting silent grammatical elements. Thus, their comprehension reflects what Laleko and Polinsky (2017) characterize as the Silent Problem. Similar problems are observed in production, but at the same time, the production of HLs can be characterized by a number of missing elements, some of which are not attested in the baseline. Most commonly such omissions in production are found in morphophonology or inflectional morphology, but free-standing clausal constituents also seem to be affected.

Before I go into the details of the Silent Problem, let me present some introductory remarks that have to do with the *inventory of silent elements* and with the *divide between comprehension and production*.

Not all missing material is the same from a theoretical standpoint. Some silent material arises under deletion of individual segments which form part of a movement chain (using the traditional terminology, I will refer to such deleted material as a *trace*). Likewise, the material deleted under identity in ellipsis results in silence, and as I have shown elsewhere, such instances of ellipsis undergo reanalysis in HL (Polinsky 2016, 2018: 263–268). Silent elements can also correspond to dedicated lexical items: null pronominals, which

constitute part of the lexical inventory of a given language. Next, we recognize implied missing subjects of control clauses; depending on the theoretical outlook and particular analysis, these are identified either as PRO, which is a universal lexical category without a corresponding overt counterpart; as material deleted under movement, or as a null pronominal (see Landau 2013 for an overview). In addition to phrasal silent elements, there are also empty (functional) heads, which I will not consider here.

In what follows, I will concentrate on null pronouns (*pro*) in subject and object position, and on gaps in relative clauses. The latter can be interpreted as material deleted under A-bar movement or as null pronouns, depending on the language and analysis of the relevant relative clause (see Bianchi 2002, Cinque 2020, de Vries 2018 for overviews of prominent analyses).

In preparing this chapter, I was astonished to see that as a field, we are missing quite a few data points with respect to HLs (and corresponding baselines). Crucially, as we proceed to fill the empirical gaps in HLs, a key strategy should be to use phenomena rather than languages as a starting point. Volumes of work have been produced on individual HLs and comparisons between them, but it is not uncommon for us (present company included) to lose track of what these empirical observations are supposed to show. Knowing exactly what we do not know is the first step toward plugging information gaps, and it is my hope that some of the points made below can serve as goalposts for future work.

Turning to language performance, linguistic knowledge is deployed in two main ways: people use language to convey their thoughts to others or to themselves (production) and to understand what is said by others (comprehension). In both situations, they must assemble the building blocks of language in a systematic fashion. Accordingly, as we evaluate HL data, it is sensible to combine the data from HL production and comprehension. Of course, this brings in additional complications for a researcher. On the practical plane, more data have to be collected (and indeed, most of the existing work on HLs focuses exclusively on just one of the two aspects of language knowledge and use). On the theoretical plane, paying attention to production and comprehension alike commits one to taking a stand on the relationship between the production (generation) and comprehension (parsing) of language. That relationship is a complex problem, one that has been at the forefront of linguistic theorizing, especially as experimental work feeding back into linguistic theory has become a norm in linguistics (see Momma and Phillips 2018 for an overview). Roughly, we can identify two main approaches to production and comprehension:

- (i) Production mechanisms and comprehension mechanisms are distinct, but they interact heavily during a single act of understanding or speaking (see Huettig and Hartsuiker 2010 for an overview).
- (ii) Production mechanisms and comprehension mechanisms share the same underlying representations (Momma and Phillips 2018).

Approach (ii) represents a more nuanced variation on what is commonly referred to as a *single-mechanism view*. Note that both types of approaches, (i) and (ii), allow for interaction between production and comprehension but differ in the understanding of the root mechanisms of these two processes. Here, I adopt approach (ii), both in light of growing evidence in its favor and because of analytical parsimony: unless we have evidence to the contrary, it is more economical to assume that production and comprehension are based on the same mechanisms of structure building.

At the same time, there are interesting differences between production and comprehension in general and in HLs in particular (see, for example, Bowles 2011, Frasson et al. 2021, Montrul et al. 2008, with respect to silent elements). Taken at face value, such differences can be evidence in support of approach (i), which implies differences between comprehension and production. If so, such discrepancies in data need to be examined closely.

The structure of the chapter is as follows. In Section 3.2, I consider the use of null pronouns in HLs, and in Section 3.3, I examine gaps (or absence thereof) in relative clauses. In considering both types of silent elements, I distinguish between comprehension and production data. In Section 3.4, I offer my main conclusions and identify areas where further work is needed.

3.2 Null and overt pronouns

3.2.1 Comprehension

The fate of null pronouns, primarily in the subject position, has been a popular topic in bilingual research. Null pronouns need to be licensed and identified/recovered (Rizzi 1986b). *Licensing* ties a null pronoun to a particular structural position, which in principle could include the subject or object position. *Recovery* allows the reference of a null pronoun to be identified in one of two ways: (i) through agreement (in a morphologically rich language; see Rizzi 1986b and subsequent literature), or (ii) through an interpretive rule linking the null element with a current discourse topic, especially in languages like

Mandarin or Japanese (compare Huang 1984: 550: “there is a rule of coindexation, in the discourse grammar of a discourse-oriented language, which coindexes an empty topic ... with an appropriate preceding topic”).

Since null pronouns do not fully disappear from HLs (an issue I will return to below), the question arises as to whether “bilinguals have native-like knowledge of the syntax of null and overt subjects more generally” (Keating et al. 2016: 39), that is, whether bilinguals retain the baseline licensing and recovery conditions for null arguments. In considering this question, researchers often ask whether the null/overt contrast is sufficiently represented in the input to a given heritage language and whether heritage speakers of that language actually produce null forms.

Few studies address the question of whether bilinguals maintain the contrast between overt and null pronouns (many researchers concentrate on possible biases in the linking of a pronoun to an antecedent). Keating et al. (2011, 2016) asked whether bilinguals maintain the null/overt contrast and respond differently to null versus overt pronouns during online processing. Using a series of comprehension experiments, they investigated this question for heritage speakers of Spanish, a full pro-drop language, who are dominant in American English. They examine the bilinguals’ knowledge of impersonal null pronouns as well as their knowledge of the contrast between referential null and overt pronouns. In comparing their bilinguals to the Spanish baseline, they conclude that “both groups were similar in that they responded differently to null versus overt pronouns during online processing” (Keating et al. 2016: 45). In other words, the bilinguals’ lexicon still includes null pronominals, arguably with the same specifications and licensing requirements as in the monolingual lexicon.

Rinke and Flores (2018) examined the interpretation of null/over subjects by German-Portuguese and Spanish-Portuguese bilingual children (thus, heritage speakers of Portuguese living in Germany and in Andorra respectively).¹ Their results are similar to the ones by Keating and co-authors, with the main conclusion being that heritage speakers had the same specifications and licensing requirements as in the baseline.

In a similar vein, Dubinina and Polinsky (2013) noted sensitivity to the null/overt pronoun contrast in the subject position for heritage speakers of Russian. The preservation of this contrast is particularly striking because in Russian, unlike in Spanish, pro-drop is limited and occurs only in certain embedded clauses, something which is unlikely to be common in the input.

¹ In this chapter, I discuss some data from European Portuguese and Brazilian Portuguese. Unless noted otherwise, *Portuguese* refers to the European variety.

Laleko and Polinsky (2017) noted that Heritage Korean speakers also maintain the null/overt pronoun contrast. Thus, although the empirical data are still limited, the balance of evidence suggests that the null/overt contrast is maintained in both rich-agreement pro-drop languages (Spanish) and languages with topic-determined pro-drop (Korean).

Given that null pronominals are in principle available in the heritage grammar, how are they recovered and identified? With respect to null and overt pronouns, interpretive biases have been explored extensively, in both the theoretical and psycholinguistic literature (see Keating et al. 2016 for an overview). The overall conclusion is that monolinguals preferentially link null pronouns to subject antecedents and link overt pronouns to antecedents in lower structural positions (Carminati 2002, 2005; Keating et al. 2011, 2016). Note that there are two correlations here: one is between subjects as antecedents of null pronouns, and the other between non-subjects as antecedents of overt pronouns.

To illustrate, in the Spanish example below, the null pronoun in the second clause is preferentially interpreted as referring to the subject, and the overt pronoun *él*, as referring to the object:²

- (1) a. Juan_i pegó a Pedro_k. *pro*_{i > k} está enfadado. [Spanish]
 Juan hit PRP Pedro be.PRS.3SG angry.M
- b. Juan_i pegó a Pedro_k. Él_{k > i} está enfadado.
 Juan hit PRP Pedro He be.PRS.3SG angry.M
- ‘Juan hit Pedro. He is angry.’ (Keating et al. 2016: 38)

However, it is worth noting that trends observed in the baseline are only preferences, not a hard rule. Depending on the test, these preferences are observed in 50–75% of cases (Carminati 2005; Keating et al. 2016, and references therein).

How do heritage speakers deal with these preferences? It appears that the two-way association between subjects as antecedents of null pronouns and non-subjects as antecedents of overt pronouns is not pronounced in HLs (compared to the baseline). In their extensive comparison between monolingual Mexican Spanish speakers and English-dominant bilinguals, Keating et al. (2016) observed that “only the monolingual group showed a referential bias with overt pronouns,” linking these pronouns to non-subject antecedents (Keating et al. 2016: 45).

² The interpretive preferences are indicated by subscripts.

Heritage speakers appear to generalize the recovery strategy that links any pronoun, whether null or overt, to the highest structural position available. To reformulate the hypothesis proposed by Carminati (2002: 57) for Italian, we can then observe the following principle:

(2) *Position of antecedent strategy (PAS)*

Choose the highest structural argument as the antecedent.

Three factors are likely to play a role in the observed generalization of PAS by heritage speakers: (i) the reduction in the scope and use null pronouns in HLs; (ii) the *scalar nature* of PAS, and (iii) the *long-distance nature* of the anaphoric dependency.

It is not a big surprise that heritage speakers may have more restricted lexical resources than baseline speakers. In particular, their inventory of null pronominals seems to be more limited than the inventory in the corresponding baseline. On the flip side, looking at the heritage grammar, we observe some changes in agreement, both on goals and probes. In terms of the content of phi-features, their array is less rich than in the baseline and/or they may become more opaque (see Scontras et al. 2018 which shows that number and gender features get bundled together in heritage Spanish, but not in the baseline). It may also be the case that fewer structural elements probe for phi-features in HLs, as compared to the baseline, which again contributes to a more restricted set of agreeing elements. These changes in agreement contribute to the decline in null pronominal elements, at least in rich-agreement languages. The result is the reduced range of null pronominals in an HL, and ultimately the loss of these elements.

Perceptual salience, or rather lack thereof, may also contribute to these putative changes in the grammatical system. Functional material is often unstressed and low in perceptual salience. If heritage speakers do not perceive functional elements, they may also produce them less and less, which is yet another pathway to the loss of agreement. To replace the material that is low in perceptual salience, heritage speakers may start producing more overt forms. The combination of licensing restrictions and rearrangement of vocabulary items in favor of the more salient elements creates a perfect storm that leads to changes in the agreement system and the abundance of overt pronouns.³

Let us now turn to factor (ii). In the baseline languages, the linking of overt pronouns to non-subject antecedents is a preference, not a categorical

³ I am grateful to Tanja Kupisch for suggesting to me the role of salience in this instance.

rule. Heritage speakers disfavor scalar principles more generally, not just in application to PAS. This aversion to scalar principles in HL gets realized as the elimination of privative oppositions in favor of equipollent ones. In a privative opposition, one of the members is fully specified, and the other either lacks the relevant feature or does not have it fully specified (while the absence of a feature vs. lack of specification is generally important, it is not crucial for the discussion here). In an equipollent opposition, both members of a grammatical opposition are equally specified (see Polinsky 2018 for an overview, and Laleko 2011, 2015, for the elimination of privative oppositions in favor of equipollent ones in the aspectual grammar of HLs). It is tempting to view the difference between the two types of opposition in terms of economy. Full specification appears to be less economical, but at the same time, lack of specification leads to greater uncertainty, something that heritage grammars tend to avoid.⁴ This avoidance of uncertainty underlies the switch from scalar to discrete principles. If this is on the right track, the erosion of the scalar linking principle associated with pronouns is unsurprising.

With respect to factor (iii) above, associating pronouns (null or overt) with discourse antecedents requires significant processing effort. This ties into a general challenge for heritage speakers. Lower proficiency heritage speakers (and other speakers differing from the monolingual baseline, as explored in the work of Sorace and her colleagues) have problems connecting elements of structure that are separated by distance. If the lexical content is more clearly specified, that allows for an easier construction of long-distance dependencies. Here I would like to underscore that factoring in processing difficulty requires a transition from processing to grammar. Building long-distance dependencies may start out as a challenge in real-time processing only, but if strategies deployed to deal with such a task become entrenched, they become part of the new grammar. How such entrenchment happens needs to be worked out, otherwise the connection between processing effort and grammar is merely a desideratum.

In sum, the problems heritage speakers encounter with null subjects arise from the confluence of three factors: difficulty with elements that are low in perceptual salience – nothing is perceptually lower than a silent element;

⁴ A question may immediately arise here: how can we reconcile the avoidance of underspecification with the lack of full feature specification on some probes that I discussed in relation to factor (i)? There is no contradiction here, because the lack of full feature specification on probes eventually leads to the loss of particular features, something that can be amply observed in the loss of gender features or clusivity (which, according to Harbour 2016, is a privative feature, hence another challenge to heritage speakers) under contact.

difficulty with scalar representations; and difficulty integrating linguistic material at a distance.

In the next section, I will examine the influence of these (and additional) factors in a different grammatical domain: relative clauses. But before I do that, let me also consider the comprehension of null objects.⁵ Superficially, languages with null objects can be divided into two types: languages where null objects alternate with overt noun phrases and clitic pronouns and languages without object clitics, where the alternation is only between overt noun phrases and null objects. Structurally, not all null objects are the same, and they may receive different analyses (see Flores et al. 2020, Rinke et al. 2019 for a short overview).

The comprehension of null objects in HLs has not received much attention. Chou et al. (2020) stands out as a study where such comprehension was closely investigated with an eye to a comparison between monolingual and English-dominant heritage speakers of Mandarin. In their picture-verification experiment, they provided subjects with a context and then asked them to evaluate if the test sentence, with a transitive predicate and a null object, appropriately described the context. Consider the following test stimuli:

(3) a. Context sentence (with accompanying picture):

‘It is a cold day. Cuiyi has just received a box in the mail.’

b. Test sentence (with accompanying picture):

Cuiyi yaohuang-le *ec* yixia. [Mandarin]
Cuiyi shake-PRF once

‘Cuiyi shook (the box).’ (Chou et al. 2020: 234–235)

The authors did not find any difference between monolinguals and heritage speakers; both groups were able to correctly interpret the null objects in context. However, as the authors themselves point out, “being able to accept null objects in contextually appropriate contexts is different from being able to reject null objects in prohibited contexts” (Chou et al. 2020: 252). The latter include, for example, situations of A-binding, where Mandarin null objects are impossible in the baseline, possibly for minimality reasons:

⁵ I will only consider definite null objects here, since the licensing and interpretation of indefinite (weak) null objects is less clear to begin with.

- (4) Zhangsan_i shuo [Lisi_j bu renshi $ec_{k/i/*j}$] [Mandarin]
 Zhangsan say Lisi not know
 ‘Zhangsan_i said that Lisi_j does not know ($him_{k/i/*j}$).’ (Chou et al. 2020: 252)

At this stage, it is not known whether baseline and heritage speakers respond differently to such structures. If both baseline and heritage speakers have the same interpretation of (4), that may be a signal of the shared grammar of null objects, and concomitantly, of the fact that heritage Mandarin allows such null objects. If heritage speakers are accepting of sentences like (4), and at the same time are accepting of null objects more generally, that would be an indication of a different grammatical representation in heritage Mandarin.

3.2.2 Production

In principle, monolinguals and bilinguals may differ in both their licensing and recovery of null pronouns, but there is no evidence that the licensing conditions differ across the two populations. In particular, we do not find null pronouns in adult HLs appearing in structural positions that are different from the structural positions found in the baseline. As there is no evidence of different licensing conditions in HLs, compared to the baseline, by elimination, all the issues in the variability of null pronouns have to do with interpretation and use. The general intuition is as follows: in cases where an overt form alternates with a null one, the overt form is easier to interpret, which may make it preferable. That would explain the greater number of overt expressions in HL in those positions where the baseline variety has null ones.

However, the preference is not categorical; heritage speakers still drop overt elements, and the absence of categorical differences between the baseline and HLs is a particularly noteworthy characteristic of heritage production. Further still, as I mentioned in the introduction, some instances of heritage production are ridden with omissions, both at the morphological and even clausal level, which creates an intriguing tension between the apparent need for an overabundance of overt material for comprehension and the observed dearth of overt material in production.

Numerous studies document the increased use of overt subjects in pro-drop languages, like Spanish and Italian, when these languages come into contact with non-pro-drop languages, such as English or German (Lipski 1996; Montrul 2002, 2008, 2016; Müller and Hulk 2001; Otheguy et al. 2007;

Serratrice 2007; Serratrice et al. 2004; Silva-Corvalán 1994; Sorace and Filiaci 2006; Sorace et al. 2009; Tsimpli et al. 2003, 2004; among others). This is not surprising, as overt subjects are not ungrammatical in the baseline, and they allow heritage speakers to avoid any ambiguity in production. The overall use of these subjects (which is never categorical) is consistent with the general tendency for overmarking observed in heritage language (see Polinsky 2018 for an overview).

The use of null pronouns may already be diminished in the speech of first-generation immigrants, whose language serves as input to heritage speakers, although the empirical data vary. Several researchers document a smaller proportion of null subjects in first-generation immigrant Spanish (Montrul 2016; Otheguy et al. 2007; Raña-Risso and Barrera-Tobón 2018). For Spanish, one could argue that the overuse of overt subjects is due to transfer from English (Otheguy et al. 2007). On the other hand, in their production study of two generations of Portuguese speakers in Germany, Flores and Rinke (2020b) did not find any decline in the use of null subjects by first-generation immigrants from Portugal. Note that in both cases the societally dominant language (English, German) is not *pro*-drop, but it is hard to estimate how much of the dominant language is actually used by first-generation immigrants in each case study. Some first-generation immigrants may have very low proficiency in the societal language, and if so, transfer from that language is unlikely. Additionally, immigrant speakers converse with interlocutors of different backgrounds and may feel uncertain about how well they are understood, which may lead to the need for explicitness. The use of overt pronouns is a particular case of explicitness, and this speech behavior may become entrenched.

A more telling test case for the emergence of overt pronouns in heritage language would be a setting not driven by transfer: a bilingual environment where both the dominant and baseline language allow null arguments, particularly null subjects. The Spanish-Catalan dyad presents such a case: de Prada Pérez (2009, 2019) examined the contact between these two languages in Minorca and found no significant difference in the rate of overt subjects among Spanish monolinguals, Catalan monolinguals, Spanish-dominant bilinguals, and Catalan-dominant bilinguals. She did, however, observe a difference between the omission of first- and third-person pronouns; the bilinguals in her study had about the same rate of null first-person pronouns accompanied by a decline in *pro*-drop with third-person pronouns, an issue I will revisit shortly.

The case I now turn to, heritage Friulian, is more nuanced. Heritage Friulian is a northern Italian variety spoken in Latin America (Frasson et al. 2021). Frasson and colleagues examine Friulian as spoken by first-generation immigrants and next-generation heritage speakers in Argentina and Brazil.

While Brazilian Portuguese has restricted pro-drop, Spanish is a fully pro-drop language, so it is not unreasonable to expect the growth in pro-drop in the Friulian of Argentina, supported by both languages in the dyad.

Heritage Friulian presents an intricate pattern, one that requires a separate consideration of each person. Second-person pronouns do not drop at all, and are in fact excluded from consideration by Frasson and colleagues. In the third person, pro-drop seems to be more restricted than in the baseline, and lexical subjects are widely expressed – a predictable pattern consistent with the other observations presented here.

In the first person (singular and plural), however, overt subject pronouns appear to be omitted more than in the homeland variety. This result may appear surprising, on a closer investigation, what appears to be the rise of pro-drop is actually due to the realignment in the pronominal system. As Frasson et al. (2021) argue, subject clitics in the Friulian immigrant varieties have been changing from agreement markers to weak pronouns.⁶ If a subject clitic is in fact reanalyzed as a pronoun, it is expected to be infelicitous in combination with another pronoun.

The result is a new system with three options for first-person pronouns: strong tonic pronouns (possibly used contrastively); weak overt pronouns, the ones that develop from clitics, and regular null pronouns. Strong pronouns (*jo* for 1sg, *on/noaltris* for 1pl) are not used together with weak overt pronouns. However, the drop of a strong pronoun does not entail pro-drop; instead, that space is taken by overt pronouns of a different type.

On a more general level, the Friulian data show that the change in the heritage pronominal system does not proceed at the same pace in all grammatical persons. In the heritage Friulian case, the development of a new first-person pronoun is ahead of potential changes in the other persons. This is important for considerations of pro-drop in other heritage languages, where more fine-grained changes should at the very least be expected. As I already mentioned, the examination of Spanish-Catalan bilinguals also points to staggered development, with a stronger decline in third-person pro-drop as compared to first-person pro-drop, where more fine-grained pronominal distinctions can be observed (de Prada Pérez 2019).

Since the data from heritage Friulian and Spanish-Catalan bilinguals are based on production, it remains to be seen if the staggered reanalysis of the pronominal system is also observed in comprehension.

⁶ Such a change is consistent with the characteristic growth in analyticity observed across different HLs. It also constitutes a particular way of sidestepping the problems with inflectional morphology that are typical in HLs as well. It is common to see inflectional morphology just disappear, but Friulian offers an example of reanalysis, not loss.

The structure of relative clauses is not uniform across languages (see Bhatt 2002; Bianchi 2002; Cinque 2020; Hulsey and Sauerland 2006; Keenan and Comrie 1977; Lehmann 1984; Reinhart 1998; among others). The first cut one could make is in terms of the syntactic analysis: between those relative clauses whose gap results from movement and those where the gap is a null pronominal bound by an operator; I will address some diagnostics of these analyses below. The use of a null pronominal (as opposed to a copy deleted under movement) is a particularly likely option in languages that independently have null pronominals in their inventory (consider Matsumoto 1989 or Reinhart 1998, scholars who approach the issue from very different perspectives but nevertheless reach similar conclusions).

Relative clauses are relevant for processing studies because they force the speaker to create a connection between the antecedent and the gap at a distance. The long-distance dependency established in relative clauses is potentially challenging from two different standpoints: relating elements of structure (including the silent ones) and a general memory load.

3.3.1 Comprehension

3.3.1.1 Current state of knowledge

In numerous studies, using different techniques and sampling different populations, researchers have shown that the comprehension of relative clauses with a subject gap (subject relatives; SR) is more accurate and faster than the comprehension of those with an object gap (object relatives; OR);⁸ consider Lau and Tanaka (2021) for a recent overview of the subject advantage and all the factors that may contribute to it – their number only keeps growing.

Subject advantage is quite visible in the comprehension of relative clauses by HL speakers (and also by L2 learners and some child populations). O'Grady et al. (2001) investigated Korean heritage speakers' comprehension of subject and object relative clauses, comparing the heritage speakers (16 learners in an accelerated second-semester Korean course at a US university) to 45 non-heritage L2 learners of Korean in second-semester and fourth-semester courses at the same university. Participants completed a picture-selection task

⁸ The gap in a relative clause may correspond to other constituents, not just subject or object (see Keenan and Comrie 1977), but there has been little comparative work including other types (but see, for example, Diessel and Tomasello 2005, who do consider other types). In experimental studies, the main emphasis has been on comparing SRs and ORs, and other gap positions are still awaiting investigation.

in which they had to circle the picture corresponding to the sentence they heard; all the pictures represented reversible actions. The comparison was between subject relatives, (6a), and object relatives, (6b):

(6) a. SR [Korean]

[____i namca-lul cohaha-nun] yeca_i
 man-ACC like-ADN woman
 ‘the woman who likes the man’

b. OR

[namca-ka ____i cohaha-nun] yeca_i
 man-NOM like-ADN woman
 ‘the woman whom the man likes’

HL speakers and L2 learners performed similarly; everyone did better with subject relatives than object relatives and often misunderstood the latter as the former. Similarly, heritage Russian speakers show a strong subject bias in relative-clause comprehension (Polinsky 2011), regardless of the order in the relative clause (SV or VS).⁹ The four possibilities, all of which are licit in Russian, are presented below:

(7) a. SR, VX order [Russian]

sobak-a_i [kotor-aja ____i ukusila košk-u]
 dog-NOM which-NOM bit cat-ACC

b. SR, XV order

sobak-a_i [kotor-aja ____i košk-u ukusila]
 dog-NOM which-NOM cat-ACC bit
 ‘the dog that bit the cat’

(8) a. OR, VX order [Russian]

sobak-a_i [kotor-uju ukusila ____i košk-a]
 dog-NOM which-ACC bit cat-NOM

b. OR, XV order

sobak-a_i [kotor-uju košk-a ukusila ____i]
 dog-NOM which-ACC cat-NOM bit
 ‘the dog that the cat bit’

⁹ The syntactic derivation of these word orders is complex, but the details are not critical to the points made here. See Polinsky (2011) for more details.

The native-speaker adults, monolingual children, and bilingual children (average age about 7;0) performed extremely well on all types of relative clauses, surpassing 90% accuracy on each. Difficulties appeared only in the heritage-adult group; these speakers did well with subject relatives but performed below chance on object relatives with both word orders. The fact that bilingual children performed indistinguishably from their monolingual peers suggests that the performance by adult heritage speakers cannot be due to a fossilization of the child language. Instead, this is a striking instance of reorganization later in life.

Transfer from English cannot explain the data, since the heritage Russian speakers performed equally well (or equally poorly) on all word orders and did not prefer orders that mirror English (i.e., subject relatives in VO order and object relatives in SV order). Instead, the explanation seems to follow from the intersection of two effects: First, heritage speakers seem to lose their sensitivity to case morphology, the essential clue that guides native speakers (and, arguably, bilingual children) to determine who did what to whom. This decline in sensitivity to morphological distinctions mirrors the general neglect of inflectional morphology so typical of HLs (Lohndal 2021; Putnam et al. 2021). Second, in the absence of case-differentiating morphology, the universal preference for the subject relative interpretation, which can be interpreted as a version of PAS (presented above in (2)) kicks in, causing heritage speakers to perform perfectly on subject relatives and below chance on object relatives. The result is a rearrangement of relative-clause syntax in heritage Russian: it remains in the heritage grammar, but it is limited to the highest structural position in the clause.

In heritage Spanish, the subject advantage is also present, but with an additional effect from word order: ORs with the embedded subject in preverbal position are easier to comprehend than those with the embedded subject in the postverbal position, both for children and adults (Sánchez-Walker 2012), and this pattern seems to mirror what is observed in the comprehension of relative clauses by child language acquirers (e.g., Prévost 2009; Sevcenco et al. 2013).

The target-like comprehension of relative clauses by bilingual children has also been noted for Turkish (see Marinis and Özge 2020, on Turkish–English bilinguals, age range 4;7–9;2) and Mandarin (see Jia and Paradis 2018, on Mandarin Chinese–English bilinguals, age 6;7).¹⁰ In other words, the bilingual

¹⁰ In trilingual children, however, the results are not as even; for example, Chan et al. (2017) show that Mandarin–English–Cantonese speaking children have difficulties identifying non-SRs in Cantonese, despite the slight preference for ORs demonstrated by monolingual Cantonese children (Chan et al. 2018).

children are similar to the monolinguals (whose data I will discuss shortly later in this section). But adult heritage speakers seem to deviate from this target-like comprehension.

Although more data are needed, these comprehension results point to a coherent picture, one that is consistent with the observations about coreference presented in Section 3.2.1. We again see a strong preference for interpreting the gap in the relative clause as the subject, and this preference is stronger than what is found in the baseline. This is what transpires with the data on heritage Russian (Polinsky 2011) and heritage Korean; heritage Spanish data show more variation (Sánchez-Walker 2012), which could be related to differences in the proficiency and HL exposure levels within the study pool.

Another factor that may determine the preference for the subject interpretation deserves mention: literacy, which in turn may best be understood as the stand-in for quality of exposure. Heritage speakers rarely have experience with literacy (assuming the relevant language has written registers at all), and they often lack schooling in their HL. Complex structures, including relative clauses, are of course more common in written language (or in more elaborate registers if no written language exists for the baseline). “To the extent that increased exposure results in stronger representations, speakers with more exposure to written language might be expected to acquire stronger representations of these more complex structures” (Dąbrowska et al. 2022).

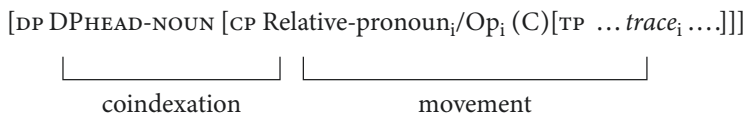
While there is no arguing that literacy/exposure plays an important role in the setting of grammatical representations, it is worth noting that the SR/OR distinction is well acquired in pre-literate children; their production may be skewed in favor of SRs, a point that I will revisit in the section below, but the comprehension of relative clauses is quite target-like around age 4;0 in monolinguals. For instance, Diessel and Tomasello (2005) consider relative clauses in child English and German and show that the main difficulties arise with the relativization of indirect objects and possessors. Hu et al. (2016) show target-like performance on Mandarin relative clauses among monolingual children, and Chan et al. (2018) document a slight preference for ORs in monolingual Cantonese children. Recall also the data from bilingual children addressed earlier in this section; bilingual children, who lack literacy in the home language, perform at ceiling in the comprehension of SRs and ORs. Thus, exposure-qua-literacy is an unlikely culprit in the deficits in relative clauses that we observe in HL speaking adults.

3.3.1.2 Gaps in our knowledge

One of the analytical uncertainties concerning relativization in HLLs has to do with the nature of the gap in the relative clause and its relationship to the head noun and relative pronoun (if available). On the analytical level, we can recognize two main strategies of relative-clause formation, one with a syntactic dependency (created by movement), the other with an anaphoric dependency between the head noun, the binding operator, and the position of the relativized constituent in the relative clause.

Simplifying things somewhat, the movement analysis of an externally-headed relative clause looks like this: An operator or the pronominal representation of the head is fronted and forms a movement chain with the element in the base position; the head noun is coindexed with these but it is generated externally to the relative clause:¹¹

(9) Movement analysis of relative clause



This analysis is not the only one that posits movement; the head-raising analysis of relative clauses provides an alternative (Bianchi 1999; Cardoso 2017; De Vries 2002; Kayne 1994; among others). On that analysis, the head noun is generated inside the relative clause (in its thematic base position) and then undergoes A-bar movement to the CP-domain; the relative pronoun, overt or abstract, is an operator, just as in (9). Next, the head NP is moved to its surface position left of the relative pronoun, where it can be associated with the external determiner.

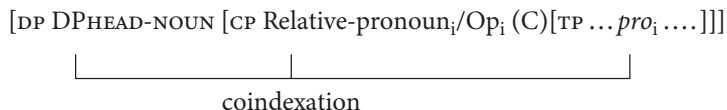
For the purposes of this discussion, the difference between the A-bar movement analysis and the A-movement analysis is not critical; therefore, I will refer to this family of analyses generically as *movement*. Taking this more general view allows for a more productive review of the data, since the two possible analyses retain a number of unresolved issues, and may not even be mutually exclusive (see de Vries 2018; Hulsey and Sauerland 2006).

In addition to the movement analyses, however implemented, another alternative has been proposed: the operator inside the relative clause binds an element in the base position, and these two pieces of structure are coindexed

¹¹ In the discussion below, I represent relative clauses as adjuncts to the head noun; for the purposes of this chapter, nothing hinges on this particular representation.

with the head noun (which is generated outside the relative clause, as in (9) above); consider Nishigauchi (1986); Reinhart (1998); a.o. Another flavor of this analysis, one that I will refer to as coindexation, is that there is simply coreference between the head noun, the relative pronoun (if any), and the gap or resumptive pronoun in the base position in the relative clause. The schematics for this analysis are as follows:

(10) Coindexation analysis of relative clause



Both movement and coindexation are compatible with the silent or overt realization of the relativized constituent in the base position inside the relative clause. In other words, this base position can be filled in two different ways: by what we can represent atheoretically as a gap, or by an overt constituent (a resumptive pronoun). The overtness criterion alone does not distinguish between A-bar movement and coindexation; in principle, both links in a syntactic chain can be spelled out, or the lower one can be spelled out exclusively (see Bošković 2007, 2014; Nunes 2004). Likewise, pronouns can come in overt or silent form as well (McCloskey 2017).

Since the differences between movement and coindexation may not be directly visible from the surface structure of a given relative clause, the learner/speaker needs to assess additional evidence in favor of one analysis or the other. Signs of A-bar movement include island sensitivity, reconstruction, parasitic gaps, or weak crossover effects (see McCloskey 2017 for an overview and some other diagnostics). Languages vary with respect to the availability of such effects, so the resulting syntactic accounts often have to rely on extremely subtle data.

Of course, all analyses are as good as the data that they account for and predict, but assuming these competing accounts are on the right track, the following question arises: does each individual HL maintain the syntactic structure of the relative clause that is available in the baseline, or do we observe a reanalysis? For example, relative clauses in the baseline may be built via movement, but does that entail that the corresponding HL is going to keep the same structure?

To the best of my knowledge, this question has not been addressed before. The main options, summarized in Table 3.1, include the preservation of the same strategy as in the baseline (movement, of the A-bar or A-type, or

Table 3.1 Relative-clause syntax, baseline vs. heritage language

Baseline	HL
Movement	Movement (no reanalysis) Coindexation (reanalysis)
Coindexation	Coindexation (no reanalysis) Movement (reanalysis)

coindexation via unselective binding) versus reanalysis. Under each of these options, the copy of the relativized constituent inside the relative clause can be silent or overt (to simplify the exposition, I do not mark the (non-)overtness of the base position in the table).

In addition, the syntax of relative clauses in a particular HL may also be affected by transfer from the bilingual's dominant language. If the home and the dominant language structure their relative clauses differently, it is reasonable to consider structural transfer effects (so-called abstract transfer, as opposed to the transfer of actual linguistic material) whereby the syntax of relative clauses in the dominant language gets imposed on the relative clauses of the corresponding HL. As with the general question raised above, we are simply on uncharted empirical territory here, but that is all the more reason why we should be aware of potential influences on the structure.

We still lack empirical facts needed to evaluate the possibilities outlined here with respect to HLs. Nevertheless, several conceptual and empirical considerations can still be offered. On the conceptual side, three properties are characteristic of HLs. The first of these is the preference for Merge over Move. This finds its reflection in the elimination of movement dependencies, which are often replaced by dependencies whose links are all base-generated and coindexed. At this stage of our knowledge of HLs, this preference for Merge over Move does not go beyond empirical observations; a principled explanation for this preference is still outstanding.

Two other characteristic properties observed in HLs have a rather direct connection to performance limitations, but it is worth noting that although these properties may be rooted in performance, they are represented in grammar. I have already touched upon the strong preference for overt forms over silent ones in HLs. That would favor all kinds of overt elements in relative clauses and could work well with the preference for coindexation between the head noun of a relative clause and an (overt) pronoun inside the relative clause. Accordingly, we may also expect the absence of island effects in such relative clauses. That in turn may skew the representation of relative-clause syntax in favor of coindexation, away from movement. And finally, we often

find that HLs, compared to the baseline, use shorter dependencies, something I return to in Section 3.3.2.2.

Turning to empirical facts, I conducted a pilot study, testing SRs and ORs in baseline and heritage Russian with respect to weak crossover (WCO), which has been claimed as a characteristic property of relative clauses in Russian and, according to some researchers, in English. WCO arises when a variable co-refers with a pronoun that does not c-command it (cf. Lasnik and Stowell 1991; Postal 1993; Ruys 2000; Safir 1984, 1986, 1996; a.o.). The reasoning behind the pilot study is as follows. If speakers are sensitive to WCO violations in relative clauses that would indicate that their grammar of relative clauses likely involves A-bar movement (other A-bar movement effects, such as reconstruction and parasitic gaps, are not observed in Russian). The absence of WCO effects in the respective relative clauses can either point to the suspension of crossover effects (as in Lasnik and Stowell 1991) or to the absence of movement. Although the absence of WCO effects may be equivocal, we stand to gain from comparing the baseline group and the heritage group; if the two groups are different with respect to WCO, that may help us understand the relevant underlying representations.

The choice of English as the dominant language for heritage speakers is also important; English shows WCO effects,¹² so we can expect that effect to be amplified in heritage Russian under abstract/structural transfer.

The following examples illustrate WCO in English restrictive relative clauses (where the relativized noun phrase cannot cross over the pronoun with which it is coindexed, in this case, a subject possessive pronoun), (11a), and in a *wh*-question, where the *wh*-expression cannot cross over the coindexed possessive pronoun, (11b):¹³

- (11) a. *the kid_i [who his_i sister called ____i a moron] (Postal 1993: 540)
 b. *Who_i does his_i mother love ____i?

WCO effects in Slavic, Russian in particular, are easy to observe in *wh*-questions, where the violation is very similar to what we see in the English examples such as (11b). Consider:

¹² As I already alluded to, this is a simplification. While WCO effects in *wh*-questions are robust, researchers disagree as to whether WCO effects in English relative clauses are present (see Lasnik and Stowell 1991, Safir 2017, for some discussion). If we go beyond the pilot study presented here, WCO in English relative clauses also needs to be tested in a systematic manner.

¹³ I represent the base position of the moved constituent as a simple gap with the relevant index. Both examples in (11) are grammatical without coindexation.

- (12) *Kogo_i ego_i mam-a ljubit ____i [Russian]
 who.ACC his mother-NOM loves
 ('Whom does his mother love?')

Outside *wh*-questions, however, WCO effects in Slavic are not as visible as in English; while generally present, they get suppressed under A-scrambling when the object linearly precedes the subject (Bailyn 2003, 2012; Slioussar 2007; Witkoś 2007; a.o.). Based on these additional considerations, only certain word orders inside a relative clause are informative with respect to WCO. In particular, if we need to test for WCO effects in Russian restrictive relative clauses with the object gap (ORs), the subject of that relative clause has to precede the verb, as in (13a):

- (13) a. OR, WCO context [Russian]
 ženščin-a_i [kotor-uju_i jejo_i sobak-a ukusila ____i
 woman-NOM which-ACC her dog-NOM bit
 za ruku]
 at hand
 'the woman whose dog bit her on the hand (lit.: which her dog bit on her hand)'
 b. OR, no WCO context
 ženščin-a_i [kotor-uju_i zlaja sobak-a ukusila ____i
 woman-NOM which-ACC angry dog-NOM bit
 za ruku]
 at hand
 'the woman whom the/a fierce dog bit on the hand'

In the mini-experiment, I asked baseline and heritage speakers to evaluate 10 pairs of such relative clauses, using the 1–7 Likert scale (1: impossible, 7: fine). The baseline speakers ($n = 15$) all left Russia as adults and settled directly in the United States (their average age is 48; average time outside the homeland, 12 years). The heritage speakers ($n = 14$), average age 24;1, were all born in the United States. Typically, heritage speakers receive their main input in the home language from their parents or other older caretakers, so the choice of older baseline speakers allows us to better approximate the language of home exposure.

The distribution of responses according to conditions and participants' groups is presented in Figure 3.1.

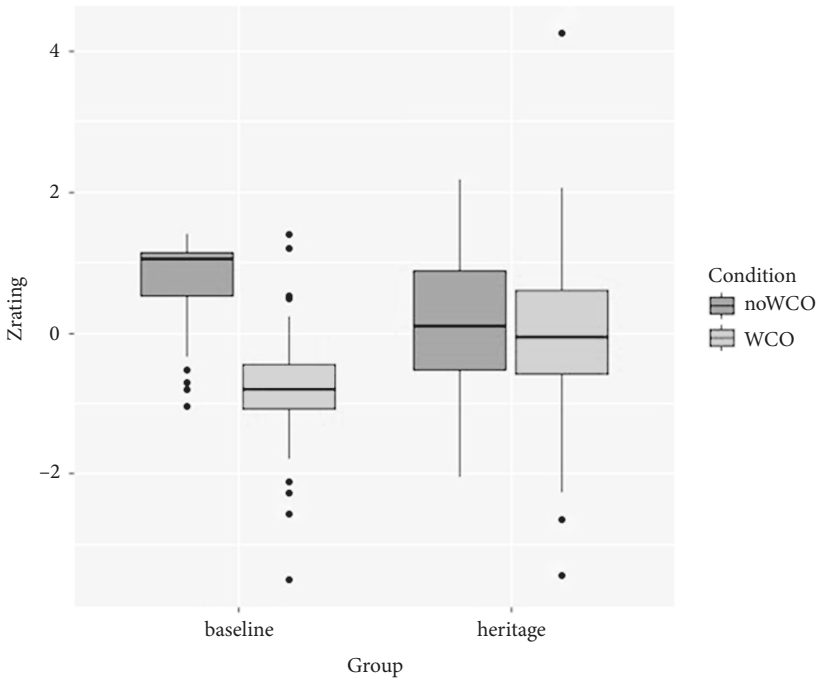


Figure 3.1 Ratings (z-score) by condition and participant group.

We observe the effect of both Group and Condition. The baseline group has a clear contrast between the WCO and non-WCO condition; the Bayesian model suggests a clear negative effect of the WCO condition ($\beta^{\wedge} = -1.41$, 95% CRI = $[-2.00, -0.76]$), such that participants rated sentences in which the object crossed over a possessor less natural than those in which the object crossed over an adjectival modifier. The WCO effect was reduced in heritage speakers, as suggested by the positive interaction between Group and Condition ($\beta^{\wedge} = 2.19$, 95% CRI = $[-1.99, -3.28]$).¹⁴ Thus, WCO violations were judged natural by heritage speakers as compared to the baseline.

Finally, the baseline speakers are not uniform; they present a striking bimodal distribution shown in Figure 3.2. Some speakers are clearly sensitive to WCO violations in ORs; for nine speakers, the raw rating of such relative clauses was 2.2, as compared to 5.8 on non-WCO condition. In the other group (six baseline speakers), WCO violations received the average raw score of 5.75, as compared to 6.1 on the non-WCO condition. Crucially, the ratings from

¹⁴ I am grateful to Zuzanna Fuchs and Utku Turk for help with the statistical analysis.

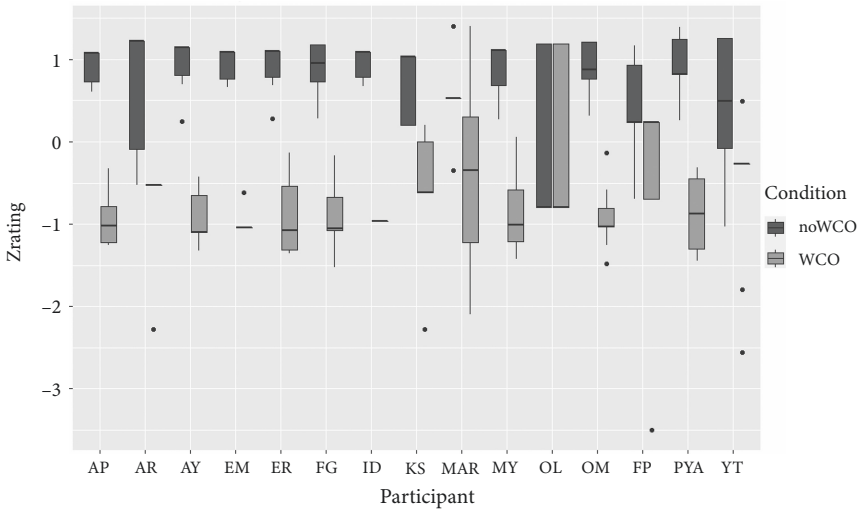


Figure 3.2 Ratings (z-scores), baseline speakers.

each individual baseline speaker are consistent; they are either about the same as the ratings of the non-WCO stimuli or they are quite low. In other words, there are no speakers who alternate their ratings for individual stimuli with WCO between high and low.

Let me first address the implications of these results for our view of *relativization in the baseline* and then for the understanding of *relativization in heritage Russian*.

It is standardly assumed that Russian restrictive relative clauses are built via A-bar movement, as diagrammed in (9) above (see Bailyn 2012: 106–108). However, the new empirical data suggest that this account is only partially correct. The bimodal pattern points to two grammars of relativization in baseline Russian, one with A-bar movement, (9), where WCO violations are found, and the other with coindexation (unselective binding), as in (10) above. In the latter case, there is no difference in rating between examples where the subject of the restrictive relative clause is modified by a possessive pronoun and those where it is modified by an adjective. Hence, WCO effects are absent.

Two other considerations lend support to the coindexation analysis. First, base-generated topics in Russian can be coindexed with null pronominals in the object position; thus, null pronominals in the object position are independently available. Second, data from the Russian National Corpus also provide additional evidence in support of the structure in (10); out of 100 restrictive object RCs with the feminine head noun, I found nine with WCO violations; for example,

- (14) PVO Iraka upravljalas' francuzskoj sistemoj_i,
 ADA Iraq.GEN was.controlled [French system].INS
 kotoruju_i ee_i "xozjaeva" smogli distancionno
 which.ACC her.GEN masters.NOM managed remotely
 otljučit'
 turn.off.INF

'The Iraqi air defense system was controlled by a French system that was successfully turned off remotely by its "owners."'

Thus, relativization in baseline Russian allows for at least two grammatical analyses, each observed in a subset of baseline speakers.

The co-occurrence of multiple grammars of relativization in one and the same language is not unheard of. For example, Barker et al. (1990) document two different strategies for the formation of Turkish relative clauses, distributed across speaker groups in a consistent manner. In an experimental study of English restrictive relatives, we also observed two different grammatical strategies, one with head raising, the other with operator movement (see Scontras et al. 2014; Tsai et al. 2014). Researchers have also documented variability compatible with multiple grammars in other aspects of language structure; for example, Han et al. (2007, 2016) uncover bimodal distribution among Korean speakers with respect to verb raising (as evidenced by scope readings under negation). Just as with the case presented here, speakers do not maintain such multiple grammars simultaneously; each speaker has a definite grammar and does not change it, but in the entire population of speakers, more than one option is present.

Unlike the baseline speakers, the heritage speakers were consistent in rating the WCO and non-WCO condition equally high (see Figure 3.1), which suggests that they all entertain the structure sketched out in (10): coindexation between the operator and the gap in the relative clause, rather than movement. The shift from syntactic to anaphoric dependency in heritage Russian is even more striking given that WCO effects are observed in the dominant language; the expected abstract transfer from English does not take place, at least not in this construction. It is a common observation that heritage speakers are a very varied group, ranging from highly proficient to rather weak speakers. Yet, contrary to all the expectations of variance among heritage speakers, we find that their relative clauses are built in a more uniform way than those in the baseline.

As striking as this result may seem, several cautionary notes are in order. The noteworthy difference between a subset of the baseline speakers and all the

heritage speakers may be due to a confluence of factors, non-discrimination of case morphology the first among them. If the heritage responses reflect the strong SR preference presented in Section 3.3.1, then these results may just mean that the speakers interpreted all the stimuli as having subject gaps, leaving us with no WCO to worry about. This explanation is unlikely since the stimuli did not denote reversible actions and had a clear plausibility bias in favor of the correct parse (e.g., “the woman that her dog bit”). Further still, such an explanation cannot be extended to baseline speakers, who are presumably sensitive to case morphology. One then needs two different accounts for the converging result.

Next, given the bimodal distribution in the baseline speakers, one can imagine that all the heritage speakers tested here were exposed exclusively to those baseline speakers whose relative clauses were built via coindexation, not movement. This is unlikely, but to rule out this possibility definitively, one has to test each heritage speaker together with the speakers who provided the primary input.

To build on this mini-experiment, one would need data on island effects in the relative clause.¹⁵ It would also be helpful to compare WCO effects among first-generation immigrant speakers and age-matched speakers in the homeland; however, WCO violations found in the Russian National Corpus (see above) suggest that this pattern is found in the homeland.

To summarize, the patterns observed with relative clauses are comparable to those observed with coreferential null pronouns and point to similar trends, with the strong preference to interpret the gap as the subject, and the particular ascent of this tendency in adult heritage speakers. We also see indications of

¹⁵ The expectation is that speakers who did not notice WCO violations would also accept scrambling (topicalization) out of the relative clause, as in (i), or would at least rate (i) higher than the group that followed the syntactic dependency. In (i), the scrambled expression and its base position are shown in bold, and the relative pronoun and the gap of the relativized argument, in italics.

- (i) **Takuju** **zadaču**_i ja ne znaju ni odnogo čeloveka_k [Russian]
 [such problem]_{ACC} 1SG not know not [one person]_{GEN}
 [*kotoryj*_k ____k by rešil ____i bez podgotovki]
 which_{NOM} SBJV solved without preparation

‘Such a math problem, I don’t know a single person that would have solved it without special preparation.’ (lit.: ... that would have solved *ec* without preparation)

I have tested such examples with two speakers from the WCO group, who rejected (i); two speakers who allowed WCO violations accepted it. However, not all Russian speakers accept extraction out of finite clauses. Complicating matters further, other diagnostics of A-bar movement are not reliable in Russian. Finally, examples such as (i) are quite complicated, which may make them degraded for independent reasons, and particularly difficult to test with heritage speakers.

a syntactic dependency (movement) being replaced with the anaphoric one. This finding is based on just one data point I presented here, but if confirmed, that type of change would be consistent with other observations on HLs where the range of syntactic dependencies is more restricted than in the baseline.

3.3.2 Production

3.3.2.1 Gaps vs. resumption

In Section 3.3.1.1, I discussed the subject gap processing advantage observed in the comprehension of restrictive relative clauses. This advantage is well-attested in monolingual populations and is even more pronounced in HL speakers. What about the production of relative clauses? As I already discussed in Section 3.2.2, most studies indicate that children acquire relative clauses relatively early, around age 4;0–5;0 (see papers in Kidd 2011; also Belletti and Guasti 2015: Chapter 5, Diessel and Tomasello 2005, Lakshmanan 2000, among others). Nevertheless, they continue to make occasional errors up to ages 7;0 or 8;0 (Friedmann and Novogrodsky 2004). Further still, several researchers note the asymmetry between the comprehension and production of relative clauses; for example, Friedmann and Novogrodsky (2004) observe that relative-clause comprehension lags behind the production of relative clauses. The general bias in child language is toward relative clauses with intransitive predicates (as in *the dog [that__ ran away]*). While this is technically a sign of subject advantage, there are no competing referents in such clauses, and the relativization is arguably easier than in clauses with a transitive predicate where one has to choose between the subject and the object.

Given these observations, we would expect the production of relative clauses in HLs to proceed in a relatively unencumbered manner. The data are still quite scarce, and more empirical work is needed, going in two directions. First, it would be helpful to establish quantitative data on the types of relative clauses in heritage speakers' production, both across languages and within a single language, especially if such a language forms a bilingual dyad with different dominant languages (such as Spanish in the United States, in Germany, or in Brazil). Second, structured production data, which would allow comparisons across subjects and across languages, are highly desirable, covering both adult and child bilingual populations. The methodology for structured production can involve a preference task, a picture-description task (cf. Novogrodsky and Friedmann 2006), or elicited production (the latter is accepted but has many critics).

The picture-description task by Novogrodsky and Friedmann (2006) was designed for special-language impairment populations, but such and similar tasks can be easily modified for typically developing children and adults. For instance, in the preference task the experimenter presents two options and asks the participants to choose which one they would like better. The task is constructed in such a way that the choice would have to be expressed as a relative clause, as shown in (15):

- (15) There are two children. The father combs one child, the barber combs another child. Which child would you rather be? Start with “I would rather be ...” or “The child that ...” (Novogrodsky and Friedmann 2006: 367)

In a pioneering study, Avram et al. (2022) applied this elicitation task in their study of French-dominant Romanian speakers ranging in age from 4;0 to 11;07, whom they then compared to an age-matched monolingual cohort. Both groups showed a preference for subject relative clauses, thus matching the results of comprehension studies. In addition, the older bilingual children produced a significant number of French-like subject relative clauses, which could be interpreted as a sign of structural transfer (as well as an indication that they are trying to avoid more complex relative-clause structures). On a more general level, the heritage speakers in this study were more likely to avoid using a relative clause than the monolingual group; instead, they used a free-standing root clause or a fragment.

The main finding that is relevant for the purposes of the current discussion had to do with the use of resumptive material in place of the gap in the relative clause, for example,

- (16) [...] copilu' [care încălzește dușu' **pe copil**] [Romanian]
 child.the that warms shower.the ACC child
 ‘the child for whom the shower warms up’ (Avram et al. 2022: ex. (12))

Such use of a resumptive phrase is unacceptable in baseline/homeland Romanian, and it was not attested in the monolingual group in this study. While still tentative, these production results point to the same tendency to avoid silent elements as noted in comprehension studies. The data presented by Avram et al. (2022) suggest that monolinguals are comfortable with silent gaps. Likewise, Diessel and Tomasello (2005), who examine the acquisition

of relative clauses primarily in English, comment on the virtual absence of resumption in their monolingual data. Consistent with the comprehension data, the Silent Problem is thus not limited to referential pronouns.

The production of resumptive expressions is probably motivated by several factors, some of which match the factors influencing comprehension (see Section 3.2). First off, establishing a long-distance dependency between the antecedent (the head noun, in this case) and the corresponding element in the relative clause requires computational effort, and it is arguably easier to connect two overt expressions than an overt and a silent element, for the simple reason that the overt expressions provide more specific instructions on accessing their denotations than the silent ones. Next comes the uncertainty of interpretation; such uncertainty is characteristic of different kinds of silent elements. It is typically observed in comprehension but it may be also implicated in production; thus, heritage speakers, whose own metalinguistic experience is more limited than that of more fluent speakers, use fully specified lexical material to avoid uncertainty and to make sure they can bring their message across to their interlocutor. And finally, the difficulty of establishing a long-distance dependency increases with structural and linear distance. Avram et al. (2022) note that resumption is more common in object relatives, confirming the effect of structural distance.¹⁶

3.3.2.2 Truncation?

So far, I have considered only those head-external relative clauses that have full sentential structure, from CP all the way down. However, head-external relative clauses can also be smaller than a CP: so-called truncated relatives. In addition to CP-relatives, the head noun can in principle be modified by a TP- or by a vP-level nominalization. Thus, the truncated option is provided by the grammar and supported by some cross-linguistic data. For instance, nominalized TPs as relative clauses have been proposed for Semitic languages (Ouhalla 2004), where they may have developed from paratactic constructions (Givón 2009). Although we do not have data on truncated relative clauses in HLs, *child language* facts and data from *endangered/shifting languages* point to the possible relevance of truncation in heritage grammars.

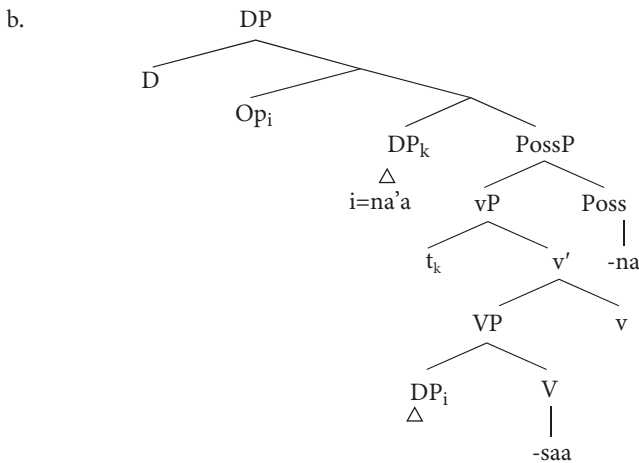
Truncated relative clauses have been documented in child language production. In particular, Pérez-Leroux (1993) found that Spanish children often

¹⁶ Frequency of different types of relative clauses is also likely to play a role, and this is where corpus data would be helpful.

produced truncated relatives after the main clause, paratactically connected to the main clause. Parallels between properties observed in child language and HLs are often brought up in research, and it is often the case that HL language development follows the L1 developmental route. On these assumptions, the presence of truncation in child language suggests that it should be explored in HLs as well.

Truncated relative clauses are also found in a number of endangered or shifting languages. For instance, a particularly clear and well-researched case of truncated relative clauses is reported for northern Paiute, a severely endangered Uto-Aztecan language (Toosarvandani 2014). Toosarvandani provides a careful syntactic and semantic analysis of these clauses showing that they are nominalized vPs (they can include vP-level adverbials but not TP- or CP-level ones). He also shows that the syntax of these nominalized clauses does not involve A-bar movement; rather, an operator in the D-field binds a (silent) resumptive pronoun (DP₂ below), which contributes the variable (the diagnostics he uses to distinguish deletion under movement from a null resumptive pronoun are the same as discussed in Section 3.3.1.2 above). The structure of one of his examples (presented in (17a)) is shown in (17b). The subject of the truncated clause still undergoes short movement to Spec, Poss, but there is no movement associated with relativization.

- (17) a. Nii ka=i=naa'a saa-na tika. [Northern
 1SG.NOM DEF.ACC=1SG.GEN=father cook-NMLZ eat Paiute]
 'I ate the thing that my father cooked.' (Toosarvandani 2014:
 809, ex. (60a))



(Toosarvandani 2014: 810)

What does this reduced structure have to do with representations of relative clauses in HLs? Just like heritage speakers, speakers of many endangered or shifting languages had limited exposure to the home language as compared to the monolingual baseline; for both types of speakers, the dominant language becomes their primary language, associated with greater societal value than the home language. In both bilingual situations, we observe the transition from early and naturalistic immersion in the ancestral language to the use of the societally dominant language. The result is different degrees of unbalanced bilingualism. Reduced exposure to the home language also leads to reduced knowledge of registers or lack of literacy. Socio-economic and socio-historical circumstances of language endangerment and HL development are different, but structural consequences of reduced exposure are quite similar in many respects (see Polinsky 2018: Chapter 7, Sasse 1992, for more discussion). Further still, for quite a few (not all) endangered languages, the only remaining speakers are heritage speakers, who are different from so-called “traditional speakers” (Bloomfield 1927; Kehayov 2017; Polinsky 2018: Chapter 1; a.o.).

In the case of HLs, the target language is spoken by homeland speakers or first-generation immigrants, and those varieties are the baseline. In the case of endangered languages, there is often no traditional language left. That’s why in examining structural consequences of unbalanced bilingualism, we find a richer empirical base in HLs, and findings drawn from HLs can be extended to endangered languages. However, in case of truncated relatives, things can go in the opposite direction: while we do not have direct evidence of relative-clause truncation in HLs, the possibility of such truncation should be considered in light of data from a different set of bilinguals.

I would like to underscore here that this scenario involves the use of truncated relative clauses as the only option for relativization; if a language under contact has both full and truncated relatives, that creates additional learning problems and may interfere with the economy of resources needed under reduced exposure. In other words, we would expect an HL to have either a fully tensed relative clause or a truncated relative clause, not both.¹⁷ Building on the parallel with endangered languages that I introduced above, it is worth noting that although truncated relative clauses are not unique to endangered languages, it is often the case that truncated structures are the only possible relative-clause type in an endangered language.

¹⁷ Elena Soare (p.c.) informs me that while baseline Romanian has both types, no truncated relative clauses are observed in heritage Romanian production.

For example, in widely spoken languages truncated relatives co-occur with other types of relatives, e.g. in Akan or in Quechua, whereas in endangered languages they seem to be the only possible type of relative clause – that appears to be the case in Uto-Aztecan, Chimariko (Jany 2008), and Wappo (Thompson et al. 2006).

If these considerations are on the right track, we can expect to find relative-clause truncation in HLs. Its use may be yet another strategy for creating a smaller size structure (and possibly one without A-bar movement), yet without sacrificing the internal complexity of the nominalized clause.

All told, the predictions we can make regarding the division of labor between full and truncated relative clauses reflect the pattern of the division of labor between anaphoric and syntactic dependencies in a restrictive relative clause (see Section 3.3.1.2), where several grammatical mechanisms are available in the baseline, but the respective HL narrows down the choices to fewer options.

3.4 Concluding remarks

This chapter presented and analyzed several cases where overt and null expressions can appear and alternate in heritage language. In earlier work (e.g., Laleko and Polinsky 2017; Polinsky 2018), we have referred to the avoidance of silent categories and some difficulties related to their interpretation as the Silent Problem. It is worth noting that the Silent Problem extends over different types of silent categories; in this chapter, I have considered null referential pronouns in the subject/object positions and gaps in relative clauses, which in turn could be traces of A-bar movement or (null) pronominals (depending on the language and on the particulars of a syntactic analysis). Several threads get woven together in the discussion of the Silent Problem, and in this conclusion, I will try to summarize each of them.

3.4.1 Lack of data

As we start to address new, more sophisticated questions related to HLs, gaps in our empirical database also become more apparent. In the discussion throughout this chapter, I have identified several areas where we lack certain basic empirical facts needed to go forward. The first has to do with the comprehension and production of null objects, especially in the contexts where such objects are impossible (see Section 3.2.1 on Mandarin Chinese). Null objects need to be evaluated both in languages where they are cross-referenced

on the verb (e.g., Romance, Georgian) as well as in so-called topic-prominent languages (e.g., Mandarin, Korean).

Next, we do not have enough information on the actual underlying representation of restrictive relative clauses in HLs; it has been a tacit assumption in the field that the size of relative clauses in a particular HL should be identical to their size in the baseline. However, the widespread use of truncated relative clauses in child language and in endangered languages, which bear a structural resemblance to HLs, points to the need to re-evaluate this tacit assumption. In a similar vein, I have shown that the underlying representation of relative clauses may differ between the baseline and the HL, which again points to the need to produce more work in this area. Relative clauses have been reasonably well studied, and syntacticians have a wealth of diagnostics at their disposal to tap into their structure in different populations.

3.4.2 Where's the change?

An obvious answer to this question is that silent categories should not be considered holistically. Yes, we see preference for overt material, but that alone does not necessarily reflect deep changes in the underlying representations. One could claim that the abundance of overt forms has to do with easing the conditions on interpretation and avoidance of ambiguity. The language of first-generation immigrants (which constitutes the main language of exposure for heritage speakers) appears to show some overextension of overt material, albeit on a smaller scale than the corresponding HLs.¹⁸ This points to possible extra-linguistic motivation for overt forms: speakers lack the linguistic confidence typical of a more homogenous speech community, and that compels them to avoid ambiguity or uncertainty as much as possible. There is no arguing against such speech behaviors, but they are a trigger, not the mechanism that underlies the changes we observe. I contend that the core issue of HL grammars has to do not with the licensing of silent elements, but with the dependencies in which they appear.

In at least two cases the grammar of silent categories undergoes restructuring in HLs. First, the data discussed here suggest that the subject advantage in long-distance dependencies is stronger in HLs than in the baseline. This in turn may be related to the more general property of HLs, which tend to

¹⁸ The observations on overextension of null elements in immigrant varieties are based on the use of overt subjects in contrast to null pronominals. No such observations are available for relative clauses.

favor equipollent oppositions, where each member of the opposition is fully specified, over privative ones (where one member of the opposition represents the elsewhere condition). That means that in the overt/null expression dyad, each comes with full featural specification, which then determines its use in the syntactic structure.

Another case discussed here where HLs manifest change in underlying representations has to do with the grammar and reanalysis of relative clauses. To begin with, the baseline group surveyed in the mini-experiment that I reported here was heterogeneous with respect to the syntax of relative clauses; some speakers build their Russian relative clauses via A-bar movement (as is claimed in most analyses, e.g. Bailyn 2012), while others show evidence of coindexation. This heterogeneity points to multiple grammars of relativization within a single language. At the same time, all the heritage speakers in this study had a more uniform grammar of relative clauses, one that avoids movement. This finding points to the leveling of the grammatical patterns in an HL. A larger question that looms behind these results has to do with the direction of grammatical reanalysis: do we expect reanalysis to go in both directions, from movement to coindexation, and from coindexation to movement, or are the pathways to reanalysis more constrained? And on a more general plane, what motivates the preference for Merge over Move that we observe in heritage grammars? Here, one could imagine two families of answers. On the one hand, as some researchers have proposed, the motivation may lie entirely in performance; HL speakers are trying to minimize working-memory load, and somehow merging items is easier in that regard than displacing them. However, whether one moves or merges, a long-distance dependency is the end result, and one then needs to explain why anaphoric dependencies are somehow easier for processing. An alternative explanation may rely on the lack of featural specification that drives movement. HL grammars demonstrate the opacity of grammatical features or their attrition (e.g., Fuchs et al. 2024; Scontras et al. 2018), and this explanatory route deserves to be explored further.

3.4.3 Beyond HLs

I have noted several parallels between HLs and patterns observed in endangered languages. The reason to pursue these parallels has to do with structural consequences of reduced exposure. If endangered-language grammars show similarities to those of the more familiar HLs, then we have evidence that

HLLs share universal properties that reflect restricted input in language acquisition, irrespective of the circumstances that shape them. In a similar vein, we can expect structures that have different underlying representations in heritage grammar to correspond to structures that appear late in acquisition (e.g., Montrul 2016). Assuming that different underlying representations in heritage grammar map to structures that appear late in acquisition, we can determine what properties are responsible for the late acquisition and whether they come from the limited input or from Universal Grammar.

The connection between HLLs, endangered languages, and child language can be seen as a two-way street; observations from one empirical domain can be tested on the others. These general considerations also suggest that HLLs are interesting not solely due to their unusual kind of bilingualism, but because they help us identify vulnerable aspects of grammar and eventually contribute to theories of learning in monolinguals.

3.4.4 Loose ends

The full list of unresolved questions would probably be longer than this entire chapter, but I wanted to zero in on one question in particular. The patterns discussed throughout this chapter are uniform in that heritage speakers tend to interpret overt forms more effectively than the corresponding silent ones. Some of these patterns also appear in production. However, heritage production is much more varied than comprehension and, crucially, is the showcase for the omission of material, especially in spontaneous speech. Whether or not this is due to real-time pressures still remains to be understood. In other words, the interplay between the overmarking and omission of material in HLLs remains a challenge for future studies. Accordingly, the variation in the production by heritage speakers (which in turn has to do with the great variability of proficiency in such populations) is rich material for constructing models of comprehension and production. I started this chapter by presenting two such models, dual and single. The data considered here may simply not be enough to fully support the single-model approach to comprehension and production, but they could clearly make an impact on our decisions about the models.

Acknowledgments

This work was supported in part by NSF grant BCS-1941733 and was partially completed during my visit to the Lab “Structures formelles du

language” at CNRS/Paris-8. I am grateful to Guglielmo Cinque, Cristina Flores, Zuzanna Fuchs, Matthew Hewett, Tanja Kupisch, Oksana Laleko, Terje Lohndal, Ora Matushansky, Eric Potsdam, Esther Rinke, Greg Scontras, Elena Soare, and two anonymous reviewers for helpful comments. All errors are my responsibility.