



---

2011

## Null Subjects in Heritage Languages: Contact Effects in a Cross-linguistic Context

Naomi G. Nagy  
*University of Toronto*

Nina Aghdasi  
*University of Toronto*

Derek Denis  
*University of Toronto*

Alexandra Motut  
*University of Toronto*

Follow this and additional works at: <https://repository.upenn.edu/pwpl>

---

### Recommended Citation

Nagy, Naomi G.; Aghdasi, Nina; Denis, Derek; and Motut, Alexandra (2011) "Null Subjects in Heritage Languages: Contact Effects in a Cross-linguistic Context," *University of Pennsylvania Working Papers in Linguistics*: Vol. 17 : Iss. 2 , Article 16.

Available at: <https://repository.upenn.edu/pwpl/vol17/iss2/16>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/pwpl/vol17/iss2/16>  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

## Null Subjects in Heritage Languages: Contact Effects in a Cross-linguistic Context

### Abstract

This paper presents an overview of the first variable examined in the Heritage Language Variation and Change in Toronto project (Nagy 2009), which strives to apply consistent methodology across multiple language-contact contexts and variables to advance our understanding of contact-induced change. It is principally comprised of sociolinguistic interviews conducted in Toronto with 40 speakers from each of six heritage languages (Cantonese, Faetar, Italian, Korean, Russian and Ukrainian). Participants are also asked about their ethnic identification, language use, and linguistic attitudes (Keefe & Padilla 1987, Hoffman & Walker 2010). Responses are translated into index scores to quantify each speakers' orientation toward their heritage language/culture and their English/"Canadian" culture.

Here we examine the effects of a constellation of factors (linguistic, typological, demographic, social) on a single linguistic variable: (pro-drop). Our Cantonese, Italian and Russian data, ~6,000 tokens, is contrasted with a sample from the Toronto English Archive (Tagliamonte & Denis 2010). For comparability with previous studies of pro-drop, we examine the effects of continuity of reference (Cameron 1995), contextual/formal ambiguity of the subject's referent (Paredes Silva 1993), clause type (Harvie 1998), priming by the preceding subject (Torres Cacoullos & Travis 2010), person and number of the subject, and tense of the following verb. Pro-drop rates and constraint hierarchies in each HL show no relationship to any indices of generation since immigration, ethnic identity or language use, suggesting that this variable is not used to construct ethnic identity and is not undergoing change as the heritage varieties of each language develop in Toronto.

# Null Subjects in Heritage Languages: Contact Effects in a Cross-linguistic Context

Naomi Nagy, Nina Aghdasi, Derek Denis and Alexandra Motut\*

## 1 Introduction

While many studies report on contact-induced language change, little progress in our theories of how languages vary and evolve can be made via their disparate methods. Inconsistencies among collection and analysis methods, and differences in contact situations and languages compared, have limited theoretical developments. The Heritage Language Variation and Change Project (HLVC, Nagy 2009) represents an innovation in applying consistent methodology across multiple language-contact contexts to advance our understanding of contact-induced change. The project examines a range of sociolinguistic variables in three generations of speakers from six heritage languages (HLs). Here we analyze the effects of a constellation of factors on a single linguistic variable: (null-subject), the presence or absence of an overt pronoun as the subject of a finite verb. Our data come from sociolinguistic interviews conducted in Toronto with speakers of three HLs. Some details of their communities are shown in Table 1, as is the place of origin targeted for each.<sup>1</sup> Analyses of Cantonese, Italian and Russian are contrasted with a small "baseline" sample from the Toronto English Archive (Tagliamonte and Denis 2010). We analyze linguistic factors that allow for comparability with previous studies. Generational changes since immigration and the effects of ethnic orientation (Hoffman and Walker 2011) are then examined.

For each of our range of measures of contact, we report a lack of effect of contact with English: the rate of null subjects does not differ significantly between generations of speakers born in the Toronto area and those born in the homeland, nor do any constraint rankings suggest a move toward the English grammar. A more nuanced search for contact effects related to quantity and quality of contact with English and speakers' attitudes toward their languages and their communities likewise does not reveal any patterns or effects suggesting contact.

Language	MT speakers <sup>2</sup>	Ethnic Origin	Established	Place of origin
Cantonese	166,650	537,000	1951 <sup>3</sup>	Hong Kong
Italian	185,765	466,155	1908	Calabria
Russian	65210	58,505	1916	St. Petersburg or Moscow
English	2,849,285	1,331,485	~1793	British Isles

Table 1. Demographic summary of Heritage Languages examined (see Nagy in press for sources).

The lack of evidence of contact effects may relate to how we define "heritage language." There are three non-overlapping categories of languages in Canada: indigenous, official (French and English) and heritage languages, spoken by immigrant groups more recent than the original French and British colonisers. Anyone who is a mother-tongue speaker of a language identified with their heritage, other than French or British, is thus a HL speaker. We do *not* use the term

---

\*We express our gratitude to the speakers who generously provided their time and voices to the HLVC Project, the many HLVC research assistants (listed at [http://individual.utoronto.ca/ngn/research/HLVC\\_personnel.htm](http://individual.utoronto.ca/ngn/research/HLVC_personnel.htm)) who collected and prepared data, students in LIN 1152 in Spring 2010 with whom we initiated this study, and the *CLAVIER 9*, *GALANA 4* and *NWAV 38 and 39* audiences who asked good questions.

<sup>1</sup>"MT speakers" is the number of mother tongue speakers in Toronto reported in Statistics Canada (2007). Ethnic origin data is from Statistics Canada (2009), also for Toronto. The "Established" date is when the first known church operating in each language was established in Toronto, a readily-available indicator of the existence of a community of speakers. "Place of origin" delimits the homeland of our first generation speakers, restricted to control the amount of regional variation in the sample.

<sup>2</sup>This category is referred to as "Chinese," and thus includes a number of people who speak languages other than Cantonese, but is the most comparable statistic available.

<sup>3</sup>Our research suggests that a Chinese-language church wasn't established until 1972, but there were already some 3,000 speakers by 1951 (City of Toronto).

“heritage language” with any implication of linguistic deficit. Thus, generalizations about impoverished systems often made about HLs are not relevant here. Future comparison to homeland varieties and the ethnic varieties of English developing in Toronto will enhance our understanding of whether there has been change that is not observable within the present sample.

## 2 Null Subjects

### 2.1 Null Subjects and Contact-Induced Change

We apply a consistent methodology cross-linguistically and cross-generationally to investigate a variable that has been claimed to exhibit contact-induced changes: null subject, also known as pro-drop. Toronto HLs are an ideal place to look for the effects of contact, since speakers of many HLs are in contact with English speakers. Subject pronoun realization is an ideal first variable to examine for effects of language contact in the HLs chosen for our study, since they are all null subject languages, and English is a non-null subject language. We can thus investigate the effect of contact between null subject languages and a dominant non-null subject language.

Such an effect has been reported in past literature. Otheguy et al. (2007) found that Spanish speakers who had arrived in New York City after the age of 16 and had been living in the city for less than six years had a significantly lower rate of overt subject pronouns than those who were born and raised in NYC (or who had arrived before age three). They concluded from this that contact with English resulted in a lower rate of null subjects. Polinsky (1995) also found evidence for a possible effect of contact with English on overt subject pronoun realization in six languages. Her study did not use variationist methods, and examined only a few speakers per language, but she found that the more attrited a speaker’s HL was, the more overt subject pronouns they used.

In contrast, Torres Cacoullous and Travis (2010) found that a putative contact effect was in fact due to priming. They examined variable *yo* (1<sup>st</sup> sg. pronoun) realization in New Mexican Spanish-English bilingual speakers, and found the same factors conditioning the realization of *yo* in these speakers as in varieties of Spanish with no English contact. Use of an overt subject pronoun was found to be conditioned by a “structural prime”: the use of an overt subject pronoun in the previous discourse favoured overt pronoun realization, whether that discourse was in English or Spanish. They concluded that this variable showed no evidence of contact-induced change.

### 2.2 Null subjects in generative grammar

To justify our selection of linguistic factors and HLs, we next present an overview of null subject realization from a theoretical perspective. A null subject language is a language where a clause may have a grammatical subject that is not realized overtly. Originally a binary parameter was proposed: languages were either +Null Subject or –Null Subject (Rizzi 1982, Perlmutter 1971). This approach has been refined to account for more typological variation (Biberauer et al. 2010): different kinds of null subject languages have different contexts where it is acceptable to “drop” the grammatical subject. Roberts and Holmberg (2010) present a typology of null subject languages, with four categories: consistent null-subject languages, expletive null-subject languages, radical pro-drop languages (or “discourse pro-drop languages”), and partial null subject languages. These typologically different groups display different properties in the distribution of null subjects.

For instance, consistent null-subject languages, like Italian, permit null subjects in all tenses and in all grammatical persons/number. Expletive null-subject languages (e.g., German) allow null expletive subjects but not referential ones. Radical pro-drop languages, like Cantonese, allow other nominal arguments, (e.g., objects) to be null, in addition to null subjects. (These languages also typically do not have person-agreement marking on the verb. Consistent null subject languages, on the other hand, typically have rich verbal inflection.) Partial null-subject languages, like Russian, limit null subjects to the 1<sup>st</sup> and 2<sup>nd</sup> person in finite clauses, and 3rd person pronouns “bound by a higher argument,” (a context that Holmberg (2005:539) admits is “rather poorly understood”). Generic pronouns are not realized overtly. Finally, non-null-subject languages, like English, bar null subjects in all finite clauses, except in specific discourse contexts (e.g., “diary drop”; see Haegeman 2000). English is used as our non-null subject comparison language.

These generalized distributions for the occurrence of null subjects are exactly that: generaliza-

tions, and, in some cases, idealizations. For instance, Roberts and Holmberg (2010:13, fn. 10) note “there is [...] considerable variation among the discourse pro-drop languages ... Chinese is apparently more restricted in this respect, making more use of overt pronouns, than, for example Japanese, and possibly more than many consistent null-subject languages.” And we provide evidence of null subjects in English below. Various approaches have been put forward to account for variation across null-subject languages. Roberts and Holmberg (2010) argue that an approach using a combination of micro- and macro-parameters can account for a wide range of differences between types of null-subject languages, with variation (within a Minimalist framework) being located primarily in features in the lexicon. Cole (2009) has suggested that “the syntactic licensing of thematic null subjects [but not expletive null subjects] is redundant,” and that recoverability of null subjects is heavily context-dependent, achieved through a combination of rich agreement morphology and the availability of a contextual antecedent. In this approach, an overt pronoun used in a null subject language is a kind of ‘last resort’ strategy where recoverability of the subject is not possible through either agreement morphology or a topic antecedent. Samek-Lodovici (1996) also found that the presence of a topic antecedent was crucial to the non-realization of an overt pronoun in Italian.

### 2.3 Null subjects in variationist sociolinguistics

Variationist studies investigate the social and linguistic factors that account for the variation that remains even within a particular language, both intra- and inter-speaker (cf. Bayley and Pease-Alvarez 1997; Otheguy, Zentella and Livert 2007; Paredes Silva 1993; Heap and Nagy 1998). First, a “subject continuity” effect is consistently reported for many languages including Spanish (Torres Cacoullous and Travis 2010, *inter alia*), Portuguese (Paredes Silva 1993), and Polish (Chociej 2010): tokens with the same referent as the subject of the previous clause favour null subjects, and tokens with a different referent from the subject of the previous clause disfavour them. This echoes Cole’s (2009) finding that a contextual antecedent is crucial for licensing null subjects.

Some studies report evidence for the “functional hypothesis” (Labov 1994:557–60), suggesting that that overt pronouns are introduced in null subject languages to clarify the discourse referent when information is unavailable in the morphology. In cases where ambiguity in the verbal paradigm makes the referent indeterminable from the verbal morphology if the subject is null, an overt pronoun will be used. Torres Cacoullous and Travis (2010:13) note that the evidence for this hypothesis is inconclusive. Their study found that morphological ambiguity was a significant factor, but it had the weakest effect of all significant linguistic factor groups. They also report that some studies have found morphological ambiguity to have a significant effect on subject pronoun realization (e.g., Bayley and Pease Alvarez 1997, Paredes Silva 1993), but other studies have reported no such effect (e.g., Ranson 1991, Bentivoglio 1987).

Other linguistic factors that have been reported are emphasis (Paredes Silva 1993 for Brazilian Portuguese), “discourse connectedness” (*ibid.*, Bayley and Pease Alvarez 1997 for Spanish), the position of the subject pronoun in the clause (Harvie 1998 for English), grammatical person and number (Bayley and Pease Alvarez 1997, Otheguy et al. 2007, Spanish; Paredes Silva 1993, Brazilian Portuguese). These factors are not cross-linguistically relevant but specific to certain null subject languages. We focus on the grammatical person and number effects.

The variety of null subject languages noted above and the array of factors that can contribute, across and within languages, to the realization of overt/null subject pronouns, create a rich opportunity to investigate the kinds of factors that hold cross-linguistically, using comparable and consistent methods across a variety of languages and generations. As described in Section 3.2, we categorize our tokens to examine predictions made by the theories outlined here. Space limitations prohibit extended discussion of their support in our data, given our focus on a search for contact effects.

## 3 Methods

### 3.1 Data Collection

These three HLLs have well-established communities in the Greater Toronto Area. For each lan-

guage, our corpus will soon have recordings of 40 native speakers, distributed across three generations. *First generation speakers* lived in the homeland until the age of 18, and have been in Toronto for >20 years. *Second generation speakers* have at least one parent who is a first generation speaker, and *third generation speakers* are those with at least one second-generation parent. Each generation is represented by four age groups: 12–18, 19–38, 39–59, and 60+.<sup>4</sup> Two male and two female speakers represent each age/generation cell. This analysis focuses on a subset of 16 Cantonese, 11 Italian and 12 Russian HL speakers. Pending availability of the full corpus, second and third generation speakers are collapsed in this paper. For comparison with English, we include eight speakers of comparable ages from the Toronto English Archive (Tagliamonte and Denis 2010) whose British Isles-origin families have been in Canada for several generations.

Fieldworkers who are fluent speakers of a HL recruit participants, starting in their own networks. They engage participants in three tasks to elicit naturally occurring speech in the HL, two of which produce the data used here. The first is a sociolinguistic interview containing questions adapted from Labov (1984) querying the speaker's background, their family's immigration history, and their observations on language, as well as other topics of interest to each speaker. This serves to collect demographic information and to elicit and record natural speech. The second is the Ethnic Orientation Questionnaire (EOQ), parallel to Hoffman and Walker's (2010) survey, used to investigate speakers' perceived degree of orientation toward the relevant ethnic group. The full EOQ and sociolinguistic interview questionnaires are on the project website (Nagy 2009).

### 3.2 Transcription, Data Extraction and Coding

Conventions for transcription have been developed for each language and are posted on the project's website (Nagy 2009). Fluent speakers of each language are trained to produce time-aligned transcriptions of the interviews using ELAN (Wittenburg et al. 2006). We exploit ELAN's capabilities by adding mark-up tiers to code each variable, all of which are time-aligned to the original recording. ELAN thus keeps all context intact. During any stage of analysis, the researcher can recover the broader context of a token, as all tiers are searchable. ELAN calculates basic statistics and produces transcriptions and coding easily exported to various analysis programs.

From the transcribed interviews, two kinds of data are collected and coded. One is self-reports in the EOQ, which are used to develop an EOQ index for each speaker. Answers to a subset of the questions, regarding language choice broadly, language choice for reading and writing, use of language with family, ethnic self-identification, and attitudes about ethnic discrimination, are coded on a scale of 0 to 2. Answers that indicate a strong identification with their ethnic identity and the HL are coded as 2. Answers showing a strong pull towards "Canadian" identity or English are coded as 0. Mixed responses are coded as 1.

The second type of data is taken from the sociolinguistic interviews, using the first 50100 main finite clauses with subjects consisting of overt pronouns or  $\emptyset$  forms. Verbs that occur in subordinate clauses, had nouns as subjects, or were part of discourse markers are excluded.

Each token is coded for properties of the verb: person, number and tense. Other internal factor groups are subject continuity of reference, clause type (main or conjoined), and ambiguity of the subject referent. Tokens are coded as ambiguous if the verb form used is homophonous with another form. For example, in standard English, "like" is ambiguous, as it occurs with several different persons, while "likes" is unambiguous: *-s* marks 3<sup>rd</sup> sg. For subject continuity, we code every token for whether its referent was the same as the referent of the subject of the previous clause, as in (1a), or different from the previous clause, as in (1b).<sup>5</sup>

- (1) a. **It** had the old red and gold F-W-Woolworth's sign right on the corner.  $\emptyset$  [**It**] had those little creaky wood, hardwood floors. (EXM37A)  
 b.  $\emptyset$  [**We**] used to bring a lunch with us, sandwiches and stuff.  $\emptyset$  [**I**] remember we used to go with Darryl, and Gary, and Jack-G. and all of us. (EXM47A)

Following Harvie (1998) we code for whether the token appeared as the subject of a main clause

<sup>4</sup>By definition, there are no first generation speakers younger than 38.

<sup>5</sup>Other speakers' utterances were included in the previous context as they contribute to discourse context.

or as the subject of the second (or later) conjunct of two or more conjoined sentences. Tokens coded as belonging to the main clause include the first subject (*io* ‘I’) in (2). Tokens coded as being in conjoined clauses include the second subject of that example ( $\emptyset$ ).

- (2) *Io sono stato sempre un santo finche  $\emptyset$  non ho conosciuto mia moglie* (I1M61B)  
 ‘I was always a saint until  $\emptyset$  (I) met my wife.’

We assume here that conjoined clauses are full sentences complete with subjects (whether null or overt) and not conjoined verb phrases. A conjoined sentence with overt subjects is shown in (3).<sup>6</sup> Additionally, Italian tokens are coded for presence of a preverbal object (e.g., *lo, me, te*).

- (3) **He’s** in the army and **he** goes to England three-or-four times a year. (EXM44A).

#### 4 Analysis and Results

First, the overall frequencies of null subjects ( $\emptyset$ -subjects) in the four languages are compared across generations. We then examine the factors constraining variable null subjects in each language. We employ mixed-effects logistic regression modeling using Rbrul (Johnson 2009). Third, we examine the speakers’ EOQ scores and their correlations with  $\emptyset$ -subject rates. Table 2 summarizes our sample.

	Cantonese	Italian	Russian	English
<b>First Generation</b>	800	377	1,337	n/a
<b>Second and Third Generation</b>	800	670	1,834	n/a
<b>Total</b>	1600	1,147	3,171	400

Table 2. Token count, by language and generation, for 57 speakers.

##### 4.1 Distributional and Multivariate Analysis

Figure 1 displays the frequency distribution of  $\emptyset$ -subjects in English and the three HLs. Error bars indicate 95% confidence intervals for these frequencies.

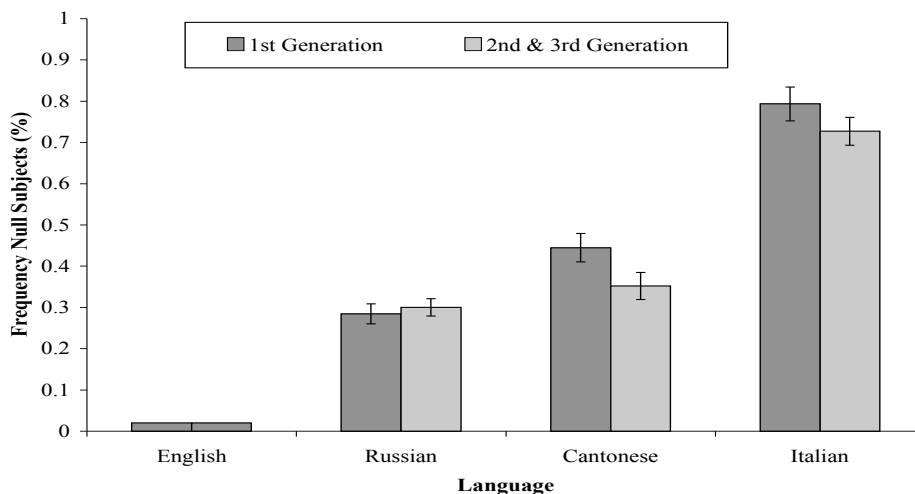


Figure 1: Percent of tokens with  $\emptyset$ -subjects across languages and generations (N = 6,216).

<sup>6</sup>It is possible that sentences with a null subject in the second conjunct are underlyingly conjoined verb phrases. However, on the surface it is ambiguous whether they are conjoined VPs or full conjoined clauses, and thus they could be interpreted by a listener as containing a null subject.

For the HLs, data is divided by generation, with first generation to the left of second/third generation. There is clear pattern for rate of  $\emptyset$ -subjects: English (a non-null subject language) < Russian (a partial-null subject language) < Cantonese (a radical pro-drop language) < Italian (a consistent null subject language). For our purposes, the generational comparisons are more relevant. In Italian and Russian, the error bars representing the 95% confidence limits indicate that there is no significant difference with respect to the rate of  $\emptyset$ -subjects between the speech of Italian- and Russian-Canadians who were born in Italy and Russia and those born in Canada. Although Cantonese shows a slight difference across generations, we will see that generation is not significant when included in a variable rule analysis. Crucially, none of the languages in either generation are close to the 2% rate of  $\emptyset$ -subjects in English. This is the first indication that contact with English is not causing a change in Toronto HLs, with respect to null subject variation.

In spite of similar cross-generational frequencies, the underlying grammar constraining the variation could still be undergoing change. If the HLs are changing through contact with English, then the variable grammar of our English speakers represents the model toward which the languages will change across generations. We thus begin our discussion of linguistic constraints by considering English. Table 3 presents a  $\emptyset$ -subject variable grammar for English, which we then compare to the grammars of the HLs.

<b>English</b>		N=400	
Fixed Effects:		<u>FW</u>	<u>n</u>
	Same Referent $\times$ Conjoined	.86	120
Subject Continuity $\times$ Conjunction	Same Referent $\times$ Main	.53	130
	Different Referent $\times$ Main	.34	123
	Different Referent $\times$ Conjoined	.21	27
	<i>Range</i>	<i>65</i>	
Random Effects:		standard deviation	
Individual		0	

*Non-significant factor groups: tense, grammatical person/number*

Table 3: Results of mixed-effects logistic regression analysis of English  $\emptyset$ -subjects.

For English, subject continuity and conjunction are selected as significant. However, there is a significant interaction between these two factors, and thus only the results of this interaction effect are reported.<sup>7</sup> Tokens in which the referent of the previous clause is the same as the referent of the token and that are the second element of a conjunction, such as the second token in (2), highly favour null subject realization. Tokens in main clauses with the same referent as the previous clause slightly favour null realizations. All tokens with different referents disfavour  $\emptyset$ -subjects. No other fixed effect is significant and there is no effect of individual speaker (as indicated by a standard deviation of 0 for the random intercept for individual).

Next, for each HL, we ran one analysis for speakers from all generations. To test for changes across generations, a fixed effect of generation was included in the model. This indicates whether the overall likelihood of null subject realization in the language has changed across generations. Additionally, we include interaction terms that cross the main effect of generation with each linguistic main effect. If these interaction terms significantly contribute to the model, then we can infer that some change has occurred across generations with respect to the linguistic effect in question. As these terms are not significant, we infer that no such change has taken place.

Table 4 presents the results of the mixed-effects logistic regression model of each of the HLs. As above, individual is included as a random intercept. For Cantonese, subject continuity is significant: same referents favour  $\emptyset$ -subjects and switch referents disfavour  $\emptyset$ -subjects. Grammatical

<sup>7</sup>Factor weight estimates for interaction factor groups are calculated based on Rbrul output that returns log-odds for main effects and interaction terms. The estimates for interaction terms are not immediately interpretable and calculations are made based on the regression model to determine interpretable factor weights.



person is also significant: 1st and 3rd person favour and 2nd person disfavours Ø-subjects.<sup>8</sup> Conjunction was significant: main clauses favour Ø-subjects while, unlike English, conjoined clauses disfavour Ø-subjects.<sup>9</sup>

<b>Cantonese</b>		Input = .199	N = 1,581
Fixed Effects:		<u>FW</u>	<u>n</u>
Subject Continuity	Same Referent	.64	966
	Different Referent	.36	615
	<i>Range</i>	28	
Grammatical Person	1 <sup>st</sup>	.64	1017
	3 <sup>rd</sup>	.58	434
	2 <sup>nd</sup>	.29	130
	<i>Range</i>	26	
Conjunction	Main	.60	1500
	Conjoined	.40	81
	<i>Range</i>	20	
Random Effects:	standard deviation		
Individual	0.627		

*Non-significant factor groups for Cantonese: generation, generation × subject continuity, generation × person, generation × conjunction, tense*

<b>Italian</b>		Input = .895	N = 1,047
Fixed Effects:		<u>FW</u>	<u>n</u>
Subject Continuity	Same Referent	.63	519
	Different Referent	.37	528
	<i>Range</i>	26	
Grammatical Number	Plural	.63	351
	Singular	.37	696
	<i>Range</i>	26	
Preverbal Direct Object	Direct Object	.65	43
	None	.35	1004
	<i>Range</i>	30	
Tense	Past perfect	.60	261
	Present	.46	547
	Past imperfect	.43	239
	<i>Range</i>	17	
Random Effects:	standard deviation		
Individual	0.728		

*Non-significant factor groups for Italian: conjunction, generation, generation × conjunction, generation × subject continuity, generation × number, generation × direct object, generation × tense*

(Table continued overleaf)

<sup>8</sup>In all the HLs, we checked for an effect of ambiguity in the verb paradigm in order to test the Functional Hypothesis. No effect was found, and this factor was then excluded from further analyses because it is non-orthogonal to the person, number and tense factor groups.

<sup>9</sup>We are not surprised by this result considering that conjunction in Cantonese is structurally very different from conjunction in English.

<b>Russian</b>		Input = .31	N = 2,507
Fixed Effects:		<u>FW</u>	<u>n</u>
Subject Continuity	Same Referent	.59	1,014
	Different Referent	.41	1,493
	<i>Range</i>	<i>18</i>	
Conjunction	Main	.63	2,223
	Conjoined	.37	284
	<i>Range</i>	<i>26</i>	
Generation × Negation	Gen. 2 × Negative	.77	220
	Gen. 2 × Affirmative	.45	1205
	Gen. 1 × Negative	.38	122
	Gen. 1 × Affirmative	.38	960
	<i>Range</i>	<i>39</i>	
Generation × Person	Gen. 2 × Third	.57	428
	Gen. 2 × Second	.67	166
	Gen. 2 × First	.64	831
	Gen. 1 × Third	.54	381
	Gen. 1 × Second	.47	109
	Gen. 1 × First	.38	592
	<i>Range</i>	<i>29</i>	
Random Effects:	standard deviation		
Individual	0.740		
<i>Non-significant factor groups for Russian: generation × subject continuity, generation × conjunction, tense, subject continuity × conjunction, generation × subject continuity × conjunction</i>			

Table 4: Results of mixed-effects logistic regression analyses of  $\emptyset$ -subjects in 3 HLs.

The critical result for us is that neither generation, nor any of the interaction terms we included that cross generation with a linguistic main effect, were selected by the model, suggesting that there is no difference between the first generation and second generation speakers with respect to the frequency of  $\emptyset$ -subjects and with respect to the constraint grammar. In other words, we see no indication that  $\emptyset$ -subject realization in Heritage Cantonese has moved toward an English model, or indeed has changed at all.

Table 4 shows a methodologically identical analysis for the Italian data. Broadly, the results resemble those of Cantonese: the effect of generation is not significant and neither are any of the interaction terms that cross generation with the linguistic main effects. Although, with the exception of the effect of subject continuity, the variable grammar is constrained differently in the two languages (Italian  $\emptyset$ -subjects are constrained by grammatical number, tense and the presence of preverbal direct objects), it is what has not been selected as significant by the model that is important. As for Cantonese, there is no indication that Heritage Italian has changed toward an English-like  $\emptyset$ -subject model, or indeed at all.

We reanalyze Hollett's (2011) data for consistency in method, but our results are consistent with her findings. There is a significant interaction between generation and negation and generation and grammatical person. Negated sentences favour  $\emptyset$ -subjects for second-generation speakers, while affirmative sentences disfavour them. In the first generation, no such effect is found. In the case of person, the hierarchy of constraints is reordered such that in the first generation, third person > second > first, while in the second generation, the relative order of favouring  $\emptyset$ -subjects is second person > first > third. As with the other languages, the subject continuity effect is significant such that tokens with the same referent as the previous subject favour  $\emptyset$ -subjects. Also, tokens that are the second subject of a conjunction favour  $\emptyset$ -subjects.

Although we observe some cross-generational changes in the Russian data, these changes are not in the direction of the English model reported in Table 3. However, Hollett (2011) suggests

that the favoring effect of  $\emptyset$ -subjects in negated sentences in the second generation could be an effect of English contact. Harvie (1998) hypothesized that English  $\emptyset$ -subjects are becoming more prevalent in negated constructions such as *don't know* and *can't say*. However, negation is not a significant factor in Harvie's data. Thus, the hypothesized change toward an English model in Heritage Russian, as with Heritage Cantonese and Heritage Italian, remains unconfirmed.

#### 4.2 Correlation with Ethnic Orientation Questionnaire (EOQ) Indices

Our investigation of null subjects thus far has revealed no effect of contact with English. We next examine the effect of contact in a more nuanced way by using the EOQ indices. In our analyses of all three HLs, a random intercept of individual speaker was significant, indicating that individuals vary with respect to their rate of  $\emptyset$ -subject usage, when linguistic factor effects are held constant. Do speakers' EOQ responses correlate with their probability of producing  $\emptyset$ -subjects?

We first determined whether the five categories of EOQ items (Language Choice, Family Language, Ethnic Identity, Reading/Writing, Discrimination) target different aspects of ethnic orientation. We calculated a Spearman rank correlation measure of scores for each category with each other category and found two significant correlations. Language choice correlates with family language ( $\rho = 0.6$ ) and with language preference for reading and writing ( $\rho = 0.45$ ). However, the measures of correlation are both intermediate effects at best. Aside from these, we can assume that each index targets different aspects of speakers' ethnic orientation.

However, when we measure the correlation between each individual's probability of using a  $\emptyset$ -subject with each individual's score for the five categories, there is no significant effect. Each measure of correlation is weak: language choice,  $\rho = 0.26$ ; family language,  $\rho = 0.28$ ; ethnic identity,  $\rho = 0.26$ ; reading/writing,  $\rho = 0.33$ ; and, discrimination,  $\rho = -0.17$ . None of the five categories correlate significantly with speaker's individual probability of producing  $\emptyset$ -subjects.

### 5 Conclusions

The analyses reported here indicate no effect of contact with English on any of the HLs examined: there is no correlation between either rate or constraint hierarchies and either generation or any ethnic orientation measure. This contradicts the popular belief that contact with English influences HLs, which would be supported if either the Canadian-born generations had HL grammars more like English than the homeland-born generation, or if speakers with more contact or lower EOQ scores had more English-like grammars. Although we see that the generations are virtually identical to each other, and different from English, we are not yet equipped to say *how* different from English they are. Once parallel analyses of homeland varieties are complete, we can see where the HLs fall between the homeland varieties and English. However, the results presented here show no effect of contact. It is also important to study this variable in the Heritage communities' English in order to see if it differs from the old-line monolingual English assumed here to be the contact variety. It is possible that English used in these communities has adapted toward the HL grammars and therefore its possible effect on the HLs would be reduced. We anticipate Marr's (in prep.) and Pustovalova's (in prep.) analyses of pro-drop in local English varieties and Homeland Russian, respectively, to take us to this next stage of the project.

### References

- Bailey, Robert, and Lucinda Pease-Alvarez. 1997. Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change* 9:349–371.
- Bentivoglio, Paola. 1987. *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela, Consejo de Desarrollo Científico y Humanístico.
- Biberauer, Theresa, Anders Holmberg, Ian Roberts, and Michelle Sheehan. 2010. *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge, NY: Cambridge University Press.
- Chocie, Joanna. 2010. Polish pro-drop: English influence on heritage Polish in Toronto. Paper presented at the 5th Annual Slavic Linguistic Society Conference.
- Cole, Melvyn Douglas. 2009. Null subjects: a reanalysis of the data. *Linguistics* 47:559–587.

- Haegeman, Liliane. 2000. Adult null subjects in non pro-drop languages. In *The acquisition of syntax: studies in comparative developmental linguistics*, ed. M.-A. Friedemann and L. Rizzi, 129–169. London: Longman.
- Harvie, Dawn. 1998. Null subject in English: Wonder if it exists? *Cahiers Linguistiques d'Ottawa* 26:15–25.
- Heap, David, and Naomi Nagy. 1998. Subject pronoun variation in Faetar and Francoprovençal. *Papers in Sociolinguistics: NWAVE-26 a l'Universite Laval*, ed. C. Paradis, D. Vincent, D. Deshaies, & M. Laforest. 291–300. Quebec: Nota bene.
- Hoffman, Michol, and James Walker. 2010. Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change* 22:37–67.
- Hollett, Meghan. 2011. *Null Subject Variation in Heritage Russian in Toronto*. Master's thesis, University of Toronto.
- Holmberg, Anders. 2005. Is there a little *pro*? Evidence from Finnish. *Linguistic Inquiry* 36:533–564.
- Johnson, Daniel Ezra. 2009. Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistic Compass* 3:359–383.
- Labov, William. 1984. Field methods of the Project on Linguistic Change and Variation. In *Language in Use*, ed. J. Baugh and J. Sherzer. 28–66. Prentice Hall: Englewood Cliffs.
- Labov, William. 1994. *Principles of Linguistic Change, v. 1*. Oxford: Blackwell.
- Marr, Isobel. *in prep.* Variation in Null-subject Pronoun Use in the English of Chinese and Italian Heritage communities in Toronto. MA thesis, University of Toronto.
- Nagy, Naomi. 2009. Heritage Language Variation and Change in Toronto. Accessed Nov. 3, 2010. URL [http://individual.utoronto.ca/ngn/research/heritage\\_lgs](http://individual.utoronto.ca/ngn/research/heritage_lgs).
- Nagy, Naomi. *In press.* A multilingual corpus to explore geographic variation. *Rassegna Italiana di Linguistica Applicata*.
- Otheguy, Ricardo, Ana Celia Zentella, and David Livert. 2007. Language and dialect contact in Spanish in New York: Toward the formation of a speech community. *Language* 83:770–802.
- Paredes Silva, Vera Lúcia. 1993. Subject omission and functional compensation: Evidence from written Brazilian Portuguese. *Language Variation and Change* 5:35–50.
- Perlmutter, David. 1971. *Deep and surface constraints in syntax*. New York: Holt, Rinehart, and Winston.
- Polinsky, Maria. 1995. Cross-linguistic parallels in language loss. *Southwest Journal of Linguistics* 14:87–123.
- Pustovalova, Alyona. *In prep.* Null subject variation in the Russian spoken language (based on materials of the Russian National Corpus). *ms.*, Moscow Higher School of Economics.
- Ranson, Diana L. 1991. Person marking in the wake of /s/ deletion in Andalusian Spanish. *LVC* 3:133–152.
- Rizzi, Luigi. 1982. *Issues in Italian Syntax*. Dordrecht: Foris.
- Roberts, Ian, and Anders Holmberg. 2010. Introduction: parameters in minimalist theory. In *Parametric Variation: Null Subjects in Minimalist Theory*, ed. T. Biberauer, A. Holmberg, I. Roberts, and M. Sheehan, 1–57. Cambridge; New York: Cambridge University Press.
- Samek-Lodovici, Vieri. 1996. Constraints on Subjects, an Optimality-theoretic Analysis. Doctoral dissertation, Rutgers University.
- Statistics Canada. 2007. Population by mother tongue, by census metropolitan area (2006 Census). [www40.statcan.gc.ca/101/cst01/demo12c-eng.htm](http://www40.statcan.gc.ca/101/cst01/demo12c-eng.htm), Updated Dec. 12, 2007.
- Statistics Canada. 2009. Visible minority population, by census metropolitan areas (2006 Census). [www40.statcan.gc.ca/101/cst01/demo53c-eng.htm](http://www40.statcan.gc.ca/101/cst01/demo53c-eng.htm), Updated Nov. 6, 2009.
- Tagliamonte, Sali, and Derek Denis. 2010. The stuff of change: General Extenders in Toronto, Canada. *Journal of English Linguistics* 38:355–367.
- Torres Cacoullous, Rena, and Cathleen Travis. 2010. Variable *yo* expression in New Mexico: English influence? *Spanish of the Southwest: A Language in Transition*, ed. S. Rivera-Mills and D. Villa. Frankfurt: Iberoamericana/Vervuert.
- Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. ELAN: a Professional Framework for Multimodality Research. *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*.

Linguistics Department  
 University of Toronto  
 Sidney Smith Hall 4073  
 100 St. George Street  
 Toronto, ON M5S 3H3  
[naomi.nagy@utoronto.ca](mailto:naomi.nagy@utoronto.ca)  
[nina.aghdasi@utoronto.ca](mailto:nina.aghdasi@utoronto.ca)  
[derek.denis@utoronto.ca](mailto:derek.denis@utoronto.ca)  
[alex.motut@utoronto.ca](mailto:alex.motut@utoronto.ca)