Variation in case marking in Heritage Polish

Paulina Lyskawa
MA Forum Paper
Advisor: Naomi Nagy
Second Reader: Aaron Dinkin

September 2nd, 2015

Acknowledgements

Table of contents

List of abbreviations:
NOM – nominative, ACC – accusative, GEN – genitive, DAT – dative, LOC – locative, INSTR –
instrumental, VOC – vocative, REFL – reflexive, NEG – negation, COP – copula, masc – masculine, fem
– feminine, neut – neuter, pl – plural, sg – singular, FW – factor weight

Abstract

The paper looks at the variation in case marking in the naturalistic speech of Toronto Heritage Polish speakers and compares it to the case marking used by Homeland Polish speakers in the same conversational setting.

Experimental studies show that the heritage case marking diverges from its standard homeland baseline (Polinsky 1995) while some naturalistic investigations report the opposite pattern where the case marking is largely unaffected (Preston 1986). I argue that one of the reasons for this difference in results is that there are two distinct types of variation in case marking: one is found both in homeland and heritage while the other is unique to a small group of heritage speakers.

150 nominal tokens were extracted from the speech of each of 12 Homeland and 13 Heritage speakers. The dependent variable of matching is determined by comparing the observed produced form to the prescribed one. The immediate picture is that there is very little mismatch in Heritage data (under 5%) and even less in the Homeland sample (~0.5%). The match is categorical for prescribed nominative case in both groups. The two groups exhibit the exact same type of variation with the case assigning lexical verbs and copula objects. However, some speakers of Heritage Polish show an additional type of variation not found in Homeland Polish – replacing any case with default nominative form. None of the social factors in the Heritage data were found to be a significant predictor of the overall rate of mismatch. On the other hand, rate of mismatch is found to correlate with the incorrect use of agreement. Finally, an analysis of a subset of data, genitive-selecting verbs, suggests that 1st generation Heritage speakers are more like Homeland speakers than like 2nd generation Heritage speakers.

I argue that distinguishing the environments within which different rules operate (as in Experimental studies) within naturalistic data (as in exploratory Variationist studies) produces the most accurate results reflecting the variable grammar of a heritage language.

1 Introduction

This paper investigates variation in case markings in Heritage Polish spoken in Toronto, Canada, comparing it to Homeland Polish. The goal is to contribute to the ongoing discussion regarding language contact effects, particularly in heritage languages. Languages within a bilingual setting have attracted a lot of research in different frameworks (speech perception and production – Au, Knightly, Jun, and Oh 2002; morphosyntax – Polinsky 1995; language processing – Hulsen 2000). The field of heritage languages is sometimes associated with attrition or incomplete acquisition (Laskowski 2009, Montrul 2008). However, it does not mean that that heritage languages are just impoverished versions of their respective full homeland languages – they are argued to be varieties on their own (Pires 2011). Viewed this way, they provide unique insight into grammar as they exhibit structures not easily explained by either simplification or borrowing. Parallel structures across heritage languages, regardless of the dominant language, suggest that they provide an insight into a kind of language blueprint. To strengthen this argument, more heritage language phenomena need to be researched. While there have been several studies of Heritage Polish case marking, to my best knowledge this will be the first one employing the Variationist method. It will allow an integrated examination of case marking in casual conversation from both linguistic and social angles.

The rich case-marking paradigm of Polish allows for a lot of logically possible changes. With seven cases and six major declension patterns, the system consists of over 42 suffixes. The abundance of case suffixes is further complicated by the fact that many of them take the same phonetic realization but their function remains distinct. Despite this apparent cognitive load, the paradigm seems to remain fairly stable in Homeland Polish. There are very few reported variation patterns in Homeland Polish and they are circumscribed to very narrow contexts (Miodek 2000: 33 on verbs selecting genitive, Doboszyńska-Markiewicz 2011 on homophonous prepositions).

Toronto constitutes a great field site for the investigation of heritage languages due to its multicultural character. The first documented large wave of migration from Poland took place in 1870 when people were fleeing occupied lands in search for a better life (Groniowski and Skowronek 1987: 410). The second wave of migration started after World War II and lasted until the end of the communist regime in Poland in 1989. This time, on top of economic motivation, there were political reasons as well. Present immigration is not as numerous but Canada remains an attractive country for immigration – about 800 Polish immigrants become permanent residents annually (Statistics Canada 2014). Toronto's Polish community is centered around different Polish organizations, most prominently the Catholic ones. In 1911 the first Polish church was built in Toronto (http://www.stanislaw.ca/historia%20parafii.htm). In the Greater Toronto Area there are 23 Saturday schools where children take classes in Polish (http://www.fahrenheitcenter.org/info-toronto---mississauga-ontario.html). In 2011 according to the National Household Survey, there were 214,455 people in the Toronto Metropolitan Area who identified their ethnic origin as Polish (Statistics Canada 2011). They same census reports that 87,795 people declared speaking Polish at home. The fairly tight-knit character of the Polish community in Toronto as well as a constant influx of Polish immigrants allows for Heritage Polish to be sustained.

The previous research on Heritage Polish in Toronto shows that it undergoes change compared to the Homeland variety. Chociej (2011) reports that the use of null subjects decreases with each generation and its linguistic constraints change. Her analysis shows that social factors such as ethnic orientation or degree of language mixing does not linearly predict the frequency of null subject in Heritage Polish. She argues that "external factors other than contact with English play a greater role in the degree and direction of change" (Chociej 2011: 4). Similarly, Lyskawa, Maddeaux, Melara and Michaud (2015) show that the phonological word-final devoicing found in Homeland Polish is not simply replaced by an English

pattern in Heritage Polish. Rather, the two processes combine and yield a unique system taking features from both from Homeland Polish and English.

The goal of this paper is twofold. Firstly, I establish the pattern of variation in case marking in Homeland Polish. As mentioned before, such variation has been observed but to my knowledge never analyzed quantitatively. Speech of 12 native speakers is analyzed, comparing actual use vs. prescribed use of case marking. I hypothesize that several linguistic factors favor the use of non-standard forms. Accusative suffixes are hypothesized to show the same form as genitive as an extension of syncretism found in the masculine animate paradigm (Miodek 2000: 33). Furthermore, the instrumental case selected by the copula is expected to be replaced with nominative. This is a tendency found in other Slavic languages as well (Masojć 2001: 121). As far as social factors are concerned, the typical prestige pattern is expected – females will use the prescribed forms more frequently than males. Finally, as mentioned before, the variation is considered stable thus age is not expected to be a linear predictor.

The results of this investigation serve as a baseline for the second and primary objective of this study – analysis of variation in case marking in Heritage Polish spoken in Toronto. The hypothesis is that the same variation is observed here as in Homeland Polish since this type of variation is linguistically motivated. Moreover, the less Polish-oriented speakers are likely to exhibit another type of variation typical for heritage languages where any prescribed case may variably be replaced by a citation form in the nominative (Polinsky and Kagan 2007: 381). Finally, a change in case marking system is expected to correlate with simplification in other morphosyntactic variables (Polinsky 1995: 110) such as null subject use and agreement as these variables may be functionally related.

Finding out where the variation occurs is just as important as finding out where it does not. Thus, I begin by examining an expansive set of environments in order to establish the envelope of variation. I look at all nominal phrases as they

hypothetically could take a different marking than prescribed. Eventually, I discard the environments where variation does not appear, e.g., nominative subjects. I argue why these contexts are special and heritage speakers never fail to use a prescribed form.

To recap, the two goals and associated hypotheses are summarized below:

1. determine the variation in case marking in Homeland Polish:
    a. speakers show variation in the use of genitive (standard) and accusative (non-standard) forms as a result of syncretism spread
    b. speakers show variation in the use of instrumental (standard) or nominative (non-standard) case following a copula
    c. females use the standard forms more often than males
    d. age is not a significant factor as the variation is stable
2. determine the variation in case marking in Heritage Polish and contrast it with Homeland Polish:
    a. same variation as in Homeland Polish (as in 1a-c)
    b. additional variation where the nominative form will take the place of any other case for the less Polish-oriented speakers
    c. the overall rate of mismatch between prescribed and observed case is higher for speakers of less exposure to Polish
    d. the rate of case mismatch is correlated with the ones of agreement and null subject use

The paper is organized as follows: Section 2 provides information about heritage language research. It also describes prescribed Polish and English case systems and discusses literature on Heritage Polish case system. Then, Section 3 lays out the methodology used in the current study. Section 4 presents the distributional and multivariate results of Heritage Polish case system analysis. It also looks at a subset of the results pertaining to the variation found both in Heritage and Homeland Polish, which allows for a quantitative comparison of the two varieties. Section 5 discusses the results and places them within the current stage of heritage language

research. Section 6 concludes that Heritage and Homeland show variation in case marking in very narrow contexts – both have a tendency of replacing genitive with accusative when selected by a lexical verb and instrumental with nominative when selected by a copula verb with Heritage Polish showing higher frequency of such substitutions than Homeland. Furthermore, I show that a number of Heritage Polish speakers also variably replace any inflectional form with nominative – a pattern found elsewhere in heritage languages, but not in Homeland Polish.

2   Background

This section first introduces the concept of heritage languages and shows the tendencies and unique patterns found there. It contrasts definitions used in the literature drawing attention to the differences that result in different foci, and sometimes even results, of research. Secondly, I give a brief overview of the Polish case system, pointing to its complexity, and compare it to English. Finally, I discuss the literature on Heritage Polish case marking and situate my research within it.

2.1   Previous heritage language research

Heritage languages are a particularly interesting topic as they occur within a bilingual environment. The interface of language contact and language variation is an understudied area of linguistics despite reflecting the experiences of the majority of world population (Tucker 1999). Recent experimental research in different heritage languages has shown that they have a lot in common and systematically simplify some of their respective homeland structures (Polinsky 2008). For example, heritage languages tend to simplify verbal agreement, e.g., Heritage Arabic (Albirini et al. 2011), Heritage Spanish (Anderson 2001) and Heritage Hungarian (Fenyvesi 2005). In Arabic and Spanish, verbal agreement indicates person and number of the subject. Heritage speakers of these languages show a significant increase in the incorrect agreement use with a tendency to use 3rd person singular inflection as a default (Anderson 2001: 389). In Hungarian, verbal conjugation agrees with person/number of the subject as well as definiteness of the object. In

Heritage Hungarian, standard subject agreement is maintained; however, the object agreement is affected with definite and indefinite conjugations being mixed (Fenyvesi 2005: 96).

Furthermore, it does not seem to be the case that a heritage language always adopts the pattern of the dominant language it is in contact with, e.g., Heritage Russian spoken in Finland has a simplified case system despite Finnish having an even more elaborate marking paradigm than standard Russian (Leisio 2006). Therefore, it appears that the heritage structures are the most unmarked and thus, studying heritage languages gives us an insight into the underlying structure of Universal Grammar - the core interest of linguistics.

The definition of a *heritage speaker* takes different forms depending on the focus of the research. Polinsky (2008: 149) and Montrul (2009: 241) describe *heritage language* as a language that is acquired first but incompletely as a switch to another, dominant language interrupted the acquisition. Thus, heritage speakers are defined in opposition to *competent speakers* who, in turn, fully mastered their language's grammatical and stylistic patterns. Similarly, Benmamoun et al. (2013: 285) define heritage languages as the ones acquired early, yet with insufficient input. The common denominator for these definitions of a heritage speaker is worse performance compared to a monolingual native speaker. Alternatively, Putnam and Sánchez (2013: 478) do not refer to distinction in proficiency between heritage and monolingual speakers but rather define a heritage speaker solely by sociogeographic circumstances – "an individual who acquired an L1 grammar (to some degree of success) of a language that is not the socially-dominant language in a given geographical area" (Putnam and Sánchez 2013: 478). Likewise, the current paper will define heritage speakers in line with the Heritage Language Variation and Change project's guidelines (more about the project in Section 3.1 Corpus), as well as the government of Canada census where a heritage language is defined as any of the non-official languages in a given country that is spoken as a mother tongue

(Harrison 2000). In HLVC heritage speakers can be distinguished based on generations as follows (Nagy 2015: 313):

- GENERATION 1 – speakers who were born and raised in one country (e.g., Poland) and immigrated to another country with a different official language (e.g., for HLVC - Canada) after the age of 18 and who have been living in the new country for at least 20 years
- GENERATION 2 – children of Generation 1; speakers who were either born in one country (e.g., Poland) and immigrated to another country with a different official language (e.g., Canada) before the age of six or people already who were born and raised in that country
- GENERATION 3 – children of Generation 2 (not included in this study)

A further stipulation required by the variationist research methodology is that a heritage speaker feels comfortable participating in an hour-long interview conducted in their heritage language. There is no direct reference to the proficiency in a heritage language in the selection of participants. The two approaches bear different implications for the research. The first approach, heritage language as a deficient variety of an otherwise full language, views the produced constructions as evidence of lack of competence or performance skills resulting from incomplete acquisition or attrition (Polinsky 1995). It excludes participants whose proficiency in their heritage language come across as native-like. Conversely, the second approach, heritage language as an equal variety to the homeland one, includes speakers of all proficiencies but for methodological reasons excludes those whose poor performance would not allow for participation in the study. The latter inclusive approach is more appropriate for the Variationist framework as it allows for drawing a big picture of what is actually going on in the speech community, conveying its range of variation. Furthermore, it allows testing whether a given pattern, or lack thereof, correlates with any social factors. Therefore, the current paper employs the inclusive approach of including any heritage speakers who agree to participate in an hour-long conversation in their heritage language.

The difference in focus and outcomes between Experimental and Comparative Variationist approaches is seen in the study on Heritage Polish null subject in Polinsky (1995) and the previously mentioned Chociej (2011). In the experimental setting, heritage speakers (but not homeland speakers) reject sentences with null subjects (Polinsky 1995: 120) while in the naturalistic situation they (and homeland speakers) are shown to produce them quite often (Chociej 2011: 50). Even though the two studies took place in different communities (the US and Canada respectively) it is unlikely that this factor alone could contribute to such large differences and in the direction where competence is affected but production is not. The current morphosyntactic variable of Polish case is also discussed in the same experimental study. It claims that the system is significantly reduced (Polinsky 1995: 95). However, as mentioned earlier, my observational approach shows that the distinction between Heritage and Homeland Polish is fairly small.

2.2   Polish case system (prescriptive)

The Polish case system distinguishes seven cases:

- nominative
- genitive
- dative
- accusative
- instrumental
- locative
- vocative

Each case can be described in terms of the function it serves in Polish. The list provided below is based on Gruszczyński and Bralczyk (2002: 203-206) and includes only the functions attested in the data:

- nominative – default form; usually a subject (1) or a complement to a conjunction/copula *to* (2):

(1) **mam-a**   tam   trochę trudniej   ma
    mom-NOM there little   more.difficult has
    'Mom has it a bit more difficult.' (P2M23A)[1]

---

[1] The speakercode has the following format: P = language; 0/1/2 = generation, M/F = sex, ## = age, A/B = additional identifiers if doubled speakercodes

(2) ojciec- Ø   Stalin-Ø   **to**   naprawdę fajny **facet- Ø**
     father-NOM Stalin-NOM   *COP* really   nice   **guy-NOM**
     'Father Stalin is a really nice guy.' (P2M44A)

- genitive – complement to a N, as a possessive (3) or a unit of measure; complement to a numeral over five (4); negation of a verb that selects an accusative (5); subjects of a verb indicating lacking (6), object of some prepositions (7):

(3) to   była **imprez-a   koleg-i**
     it   was **party-NOM friend-GEN**
     'It was a friend's party' (P2F23A)

(4) ja miałam **osiem- Ø   lat- Ø**
     I   had   **eight-ACC   years-GEN**
     'I was eight years old.' (P2F85A)

(5) **nie**   miałam   **czas-u**
     NEG   I.had   **time-GEN**
     'I didn't have time.' (P2F22A)

(6) czasami   **brakuje słów- Ø**
     sometimes   **lack   words-GEN**
     'I lack words sometimes.' (P2M21A)

(7) wysłać **do Polski**
     to.send **to Poland-GEN**
     'to send to Poland' (P2M23A)

- accusative – object of most verbs (8) and some prepositions (9) (including directional):

(8) ja **biorę krymynologi-ę**
     I   **take   criminology-ACC**
     'I'm taking criminology.' (P2M21A)

(9) jak   poprosi **o   pomoc-Ø**
     when   asks   **for help-ACC**
     'when he/she asks for help' (P2M29A)

- dative – semantic goal (10), object of some prepositions (11) and lexical verbs (12):

(10) rodzice-Ø   **kupowali**   **dzieci-om**
     parents-NOM   **were.buying** **kids-DAT**
     'Parents were buying for the kids.' (P2F26A)

(11) dyskryminowali   **przeciwko mnie**   czy **ludzi-om**
     they.discriminated **against**   **me-DAT** or **people-DAT**
     'They discriminated against me or the people.' (P2M23A)

(12) dziadek- Ø   trochę **pomaga ojc-u**
     grandpa–NOM little   **helps**   **father-DAT**
     'Grandpa helps father a little.' (P0M25A)

- instrumental – object of copula 'to be' (13), semantic instrument (14) and object of some prepositions (15):

(13) jej ojciec-Ø   **był brat-em**   mojej mam-y
     her father-NOM **was brother-INSTR** my   mother-GEN
     'Her father was my mom's brother.' (P1F67A)

(14) więc się   **posługuje** dosyć dobrze tylko brytyjskim **accent-em**
     SO   REFL **uses**   quite well   only British   **accent-INSTR**
     'So he/she speaks well only with British accent' (P2M23A)

(15) siedzieć **z**   **koleg-ami**
     to.sit   **with friends-INSTR**
     'to sit with friends' (P2M22A)

- locative – object of prepositions (16) (including non-directional):

(16) nawet **w** **Montreal-u**
     even **in** **Montreal-LOC**
     'even in Montreal' (P2M47A)

- vocative – mark the addressee of the utterance, the only case that does not form a syntactic relation with other arguments in the clause:

(17) jeżeli, **Al-u**,   ty   teraz   posprzątasz
       if   **Allie-VOC** you now   tidy
       'if you tidy now, Allie' (modified from P0F16A)

Note that it is either syntactic heads (e.g., verbs or prepositions) selecting specific cases for arguments that they are in relation to (called *rekcja* in Polish; Gruszczyński and Bralczyk 2002: 337), as in (18) – (21), or the case can be determined based on the structure it is found in. One important example that will be considered further in the analysis is that negation overrides the accusative selection of lexical verbs seen above and yields genitive (Gruszczyński and Bralczyk 2002: 204) as in (22).

(18) brać torb-ę
       take   bag-ACC
       'to take a bag'

(19) przyglądać się   torb-ie
       look.at   REFL   bag-DAT
       'to look at a bag'

(20) na   torb-ie
       on   bag-LOC
       'on a bag'

(21) pod   torb-ą
       under bag-INSTR
       'under a bag'nie

(22) brać mam-y   torb-y
       NEG   take   mom-GEN bag-GEN
       'to not take mom's bag'

Each case takes a different form depending on the declension pattern, largely (but not only) dependent on the grammatical gender and number, and in some cases animacy of the marked constituent, e.g.:

(23) masculine singular inanimate:
       brać dom-ø
       take   house-ACC
       'to take a house'

(24) masculine singular animate:
    brać   męż-a
    take   husband-ACC
    'to take a husband'

(25) feminine singular:
    brać   torb-ę
    take   bag-ACC
    'to take a bag'

(26) neuter singular (in-/)animate
    brać   dzieck-o
    take   child-ACC
    'to take a child'

(27) masculine plural:
    brać   męż-ów
    take   husband-ACC
    'to take husbands'

(28) non-masculine plural:
    brać   torb-y
    take   bag-ACC
    'to take bags'

Table 1 presents the declension of nouns in all seven cases intersecting with the major declension patterns (which are largely based on grammatical gender).

	singular				plural	
	masculine		**feminine**	**neuter**	**masculine**	**non-masc[2]**
	inanimate	**animate**				
NOM	dom-**Ø**, baton-**Ø**	pan-**Ø**	nog-**a**	dzieck-**o**	pan-**owie**	nog-**i**
GEN	dom-**u** baton-**a**	pan-**a**	nog-**i**	dzieck-**a**	pan-**ów**	nóg-**Ø**
DAT	dom-**u**, baton-**owi**	pan-**u**	nodz-**e**	dzieck-**u**	pan-**om**	nog-**om**
ACC	dom-**Ø**, baton-**Ø**	pan-**a**	nog-**ę**	dzieck-**o**	pan-**ów**	nog-**i**
INSTR	dom-**em**, baton-**em**	pan-**em**	nog-**ą**	dzieck-**iem**	pan-**ami**	nog-**ami**
LOC	dom-**u**, baton-**ie**	pan-**u**, pan-**ie**	nodz-**e**	dzieck-**u**	pan-**ach**	nog-**ach**
VOC	dom-**ie**! baton-**ie**!	pan-**ie**!	nog-**o**!	dzieck-**o**!	pan-**owie**!	nog-**i**!
	house, bar	*sir*	*leg*	*kid*	*sirs*	*legs*

Table 1. Prescribed Polish noun inflection in all seven cases, by grammatical gender, animacy and number

---

[2] In plural, the grammatical gender distinction is often labeled as *virile* vs. *non-virile* in Slavic literature.

To add to the complexity of the case system, some nouns follow a different declension than their formal features may suggest, e.g., *mężczyzna* 'man' is a singular animate masculine noun (i.e. its modifiers take the singular animate masculine form) but it declines according to the feminine pattern. For the most frequent nouns (*mężczyzna* 'man'), these may be memorized more easily than those that are rare and follow a mixed declension pattern (*państwo* polite plural 'you' has a mostly feminine declension despite plural masculine agreement).

Case inflection occurs not only on nouns but also on adjectives, numerals and pronouns. The current paper will look only at the most frequent arguments of verbs and prepositions, i.e. nouns, and personal and question pronouns. The declension of relevant pronouns is provided in Tables 2 and 3.[3]

Cross-linguistically pronouns retain more case forms than nouns (Blake 1994) including English (accusative and genitive marking survived in pronouns). Synchronic variation in Heritage Polish is expected to pattern the same way – pronouns favoring a match of prescribed case and nouns favoring a mismatch.

English case morphology seems more straightforward. Nouns are not inflected for case with the exception of possessives taking genitive *'s*. Pronouns exhibit a bit more complexity. On top of the nominative form, e.g. *I*, there are also accusative forms used for objects, e.g., *me* and genitive form used for possessives, e.g., *my* or *mine*.

---

[3] For more detailed patterns of declension refer to Sadowska (2012).

		1st	2nd	3rd		
				masc	**fem**	**neut**
singular	**NOM**	ja	ty	on	ona	ono
	GEN	mnie	ciebie/cię	jego/go/niego/-ń	jej/niej	jego/go/niego/-ń
	DAT	mnie/mi	tobie/ci	jemu/mu/niemu	jej/niej	jemu/mu/niemu
	ACC	mnie	ciebie/cię	jego/go/niego/-ń	ją/nią	je/nie
	INSTR	mną	tobą	nim	nią	nim
	LOC	mnie	tobie	nim	niej	nim
	VOC	ja	ty	on	ona	ono

		1st	2nd	3rd	
				masc	**non-masc**
plural	**NOM**	my	wy	oni	one
	GEN	nas	was	ich/nich	
	DAT	nam	wam	im/nim	
	ACC	nas	was	ich/nich	je/nie
	INSTR	nami	wami	nimi	
	LOC	nas	was	nich	
	VOC	my	wy	oni	one

Table 2. Prescribed Polish declension of person pronouns in all seven cases by grammatical gender and number. Multiple forms of some pronouns are allomorphs associated with sentence position in standard Polish and highly variable in vernacular.

	animate	**inanimate**
NOM	kto	co
GEN	kogo	czego
DAT	komu	czemu
ACC	kogo	co
INSTR	kim	czym
LOC	kim	czym
VOC	kto	co

Table 3. Prescribed Polish declension of question pronouns in all seven cases by animacy

2.3   Heritage Polish case system (descriptive)

The previous section sketched the Polish case system. The richness of the paradigm has attracted many researchers to look at this phenomenon in the heritage variety of Polish. Preston (1986) looked at the frequency of occurrence of each case in both homeland and heritage varieties of Polish. His homeland data compared speech from two Polish cities and his heritage data came from Polish Americans in Western

New York. He reported that all seven cases occurred in the speech of both groups but acknowledges that he did not investigate whether the forms occurred in the expected environment (e.g., with the correct verb). The surprising fact that heritage case marking is more similar to either dialect of homeland Polish than these dialects are to each other may be an artifact of the methodology. The need for a more thorough dependent variable coding justifies the current research.

Despite some issues with methodology, some of his findings are replicated in the present study. He points to the decrease of instrumental under copula and an increased frequency of accusative over genitive in heritage data. Preston argues that it is a result of a decline of genitive under negation. Finally, his implicational scale shows that pronouns are less affected than nouns.

The main point of contrast between Preston's findings and the current ones, beside the already mentioned genitive under negation, is the behavior of datives. Dative is the most affected case in New York Polish. Preston reports that it is lost for 13 out of 17 speakers. If by 'lost' Preston meant that it is not found in the naturalistic data, then my data is similar – dative is a rarely occurring case. On the other hand, if the appropriateness of the use of dative case is intended then Toronto Polish speakers match the prescribed form categorically.

Laskowski (1993: 131) discusses the very same issue and argues that dative is largely lost in Swedish Polish. Firstly, he shows that accusative forms are found in place of dative in his data. This is an example of actual loss and it is not found in my data. Furthermore, he suggests that the loss of dative is manifested in the use of propositional phrases. While this may be a strategy of avoidance through calquing a structure from the dominant language, using prepositional phrase in place of dative is found in some dialects of homeland Polish as well (Dzięgiel 2009):

(29) Laskowski (1993: 131):
Dają   coś   **dla**   **ptaszk-ów**
they.give   something   **for**   **birds-GEN**
'They give something to birds.'

(30) the only acceptable homeland variant according to Laskowski:
Dają   coś   **ptaszk-om**
they.give   something   **bird-DAT**
'They give something to birds.'

In my data, there is only one example of a prepositional phrase that could be alternatively expressed with a dative and it was coded as mismatch to replicate Laskowski's methodology. He views the use of prepositional structures as avoidance of dative environments, thus implication of the loss of dative. I argue that loss and avoidance are two distinct phenomena and further, avoidance in heritage data cannot be claimed without comparing to the rates in homeland data.

Laskowski (1993: 140) found that the genitive selected by a preposition is "represented entirely faultlessly" unlike the genitive selected by a verb. However, he does not refer to the fact that verbal genitive variation is also present in Homeland Polish. Finally, he reports that genitive is "decaying" when it is under negation or non-possessive N.

Laskowski (1993: 124-125) also provides an implicational scale of case loss. His general hierarchy is as follows, with the cases most likely to be retained at the left:

nominative > **a**ccusative > **g**enitive > **i**nstrumental > **l**ocative > **d**ative (NAGILD)

He also acknowledges that this pattern does not hold for each individual and there are some speakers for whom the order of accusative and genitive is reversed. He hypothesizes that it has to do with the additional attrition of the animate vs. non-animate distinction. It is an interesting explanation of the observed data; however, it is not the only logical one. Perhaps, his first proposed implicational scale is too powerful and thus inaccurate for a larger sample of people. It is not unthinkable that

accusative and genitive are overall not ordered in terms of attrition or there may be non-linguistic factors contributing to the surface difference.

Whatever the right explanation for the two observed patterns is, it is interesting to see that the pattern is similar but not entirely identical with my data. As mentioned before, I hesitate to place dative as the lowest case in the scale since my data shows the opposite. Furthermore, there is nothing universal about dative being the first candidate to be removed from the system. In some modern varies of German (Blake 1994: 174), it has been observed that dative absorbed genitive, not vice versa.

Laskowski (1993: 125 & 154) explains that the order of attrition must be a result of the order of acquisition in the heritage community with nominative being acquired first and dative last. He does not provide any data for this but instead finds acquisition data from a monolingual Polish population matching this order (Smoczyńska 1985: 626). However, such a rigid acquisitional pattern is not uncontroversial either. Łuczyński (2004: 16) and Zarębina (1965: 37-38) suggest it is locative that is acquired last. Kaczmarek (1953: 40) points to vocative being acquired earlier than most oblique cases. Overall, there is not an exact order that the majority of literature agrees on. Nevertheless, a big picture of nominative > other core cases (genitive, accusative) > oblique cases is apparent. Thus, it is this general pattern that I argue to be present in the current results as well, when rate of mismatch vs. match to homeland patterns is considered.

Polinsky's (1995) experimental approach is another one that considered case marking in Heritage Polish. It does not go into as much detail as Laskowski (1993) but nevertheless, notes that nominative overwhelmingly replaces oblique forms in American Polish. However, there are no details regarding conditioning by other linguistic factors and the phenomenon is by no means categorical.

Thepboriruk (2009) has argued that some heritage languages show linguistic conservatism. It is motivated by the fact that the homeland baseline is not a

fossilized variety – if a change is initiated after immigration, the heritage community lacks the innovative forms. Karaś (2009b) provides a study of Polish spoken in Eastern Borderlands (current Lithuania, Belarus and Ukraine). Beside the changes caused by the transfer of structures from local dominant languages (e.g.: the verb *kierować* 'direct' takes accusative instead of instrumental and the verb *czekać* 'wait' is replaced by genitive instead of prepositional phrase *na* + accusative), she also (Karaś 2009b: 38-29) notices the nominative forms used instead of instrumental with a copula, like Preston (1986). However, she explains it as an archaic form preserved from the old variety of Polish spoken before the Polish community immigrated to Eastern Borderlands. Linguistic conservatism is not expected to show in the Toronto Polish community as it is fairly young – the oldest first generation speaker left Poland 60 years ago and there is a constant influx of new immigrants. Thus, a change in case marking would need to be massively rapid in order for the conservative effect to show.

On the other hand, two other patterns of change between heritage and homeland varieties are attested in my data. First, based on research in other heritage language and their case marking, Polinsky and Kagan (2007: 381) show that heritage languages exhibit variation already present in homeland varieties where it is limited to non-standard dialects or child language. For example, in Heritage Russian prepositional case replaces any other oblique case. While, in my data, there is not one oblique case generalizing to other ones, there is a pattern of accusative replacing genitive in the lexical verb context that is also observed in the homeland data where it is associated with non-standardness. Second, Polinsky and Kagan (2007: 381) also report that many heritage languages use the unmarked case (nominative or absolutive) as a default. I show that this is indeed true in Heritage Polish; however, it is limited to only some speakers.

3 Methodology

The following section will present the methodology used in the study. Firstly, the subsection on the corpus talks about the Heritage Language Variation and Change project of which this study is a part. Secondly, the part on participants provides details on the distribution of speakers in terms of sex, age, homeland or heritage status and generation. Then, the subsection on data collection includes the description of the sociolinguistic interview that the speakers participated in as well the Ethnic Orientation Questionnaire that they answered. It also provides information regarding finding participants for the present study. The subsection on the transcription and coding protocol includes a detailed description of the dependent variable as well as conditioning linguistic and social factors. Finally, a subsection on the statistical tools and methods of analysis is provided.

3.1 Corpus

The present study is a part of the larger project Heritage Language Variation and Change (Nagy 2009). It investigates several heritage languages spoken in Toronto – Cantonese, Italian, Faetar, Hungarian, Korean, Polish, Russian and Ukrainian. Using a consistent methodology across the languages allows for a direct comparison of variation in any variable across these languages. For example, the variables of Voice Onset Time and null subject have been analyzed for five heritage languages and compared with Toronto English. The change in the VOT variable proceeds differently in, e.g., Heritage Italian and Heritage Russian (Nagy and Kochetov 2013), suggesting that it is not only the effects of contact with English that must be considered. Morphological variables are more problematic as some appear in one language but not the others. Nevertheless, I suggest that even though case marking works differently in these languages it can be studied for all of them except for Cantonese where there is no case inflection. I believe that the present study on Heritage Polish is a good start for this variable to be analyzed further across other heritage languages.

3.2   Participants

A total of 25 participants took part in the study. They are divided into two major groups based on their exposure to Polish – Homeland and Heritage.

3.2.1   Homeland

The Homeland group consists of 12 monolingual Polish speakers born and raised in Poland. All Homeland participants under the age of 60 were born and raised in the same north-west region of Poland. As the region became a part of Poland only post WWII, there are very few speakers over 60 who are born there. The present study includes speakers over 60 who were born in various different regions of Poland but came to the north-west in their early childhood. It means that the have spent a large majority of their lives in this region. Furthermore, the history of the settlement of this region, involving the migration of people from different dialect regions of Poland, yielded dialect leveling which now is the closest to standard Polish (Karaś 2009a). Table 4 presents the distribution of Homeland speakers by sex and age:

	male	female
youngest (12-18)		2
young (19-38)	2	2
middle (39-60)	1	1
oldest (60+)	1	3
		TOTAL: 12

Table 4. Distribution of Homeland speakers by sex and age.

3.2.2   Heritage

All 13 Heritage speakers reside in Toronto and can be divided further according to generation. Generation 1 includes speakers who were born and raised in Poland but immigrated to Toronto, Canada in their adulthood, after the age of 18, and have lived in Toronto for at least 20 years. Generation 2 are speakers born to Polish-background parents who meet the definition for Generation 1. Thus, they were raised in Toronto and were either born there or immigrated before the age of six. Further generations follow the same criteria but were not found in the Heritage

Polish community and thus will not be present in the study. The distribution of Heritage speakers by age, sex and generation is presented in the Table 5 below:

	Generation 1		Generation 2	
	male	female	male	female
youngest (12-18)	absent by			
young (19-38)	definition		4	4
middle (39-60)	1		1	
oldest (60+)	1	1		1
	TOTAL:3		*TOTAL:10*	

Table 5. Distribution of Heritage speakers by generation, sex and age.

3.3 Data collection

Data collection consisted of three main parts – interview, First Words elicitation and Ethnic Orientation Questionnaire. Firstly, the speakers volunteered to participate in about an hour-long sociolinguistic interview. The interviews were fully recorded with a Zoom H4 digital recorder and an Audio-Technica model AT8531 lapel microphone. The quality of the recording allows for a reliable analysis of some phonetic variables as well as lexical, morphological, syntactic, and pragmatic variation. The interviews were conducted by a native Polish speaker speaking Polish. However, heritage speakers sometimes code-switched to English. The code-switching was neither discouraged nor suggested by the interviewer, thus it was a sole choice of the participant. Code-switches were excluded from the data analyzed in this paper. The topic of the interview was also at the discretion of the interviewee – the interviewer suggested topics and pursued asking further questions only if the speaker expressed some interest in them. This way the interview elicited a participant's vernacular in the Labovian sense - "that variety which is used by speakers in their most unself-conscious states" (Labov 1972). After the interview was completed, the speakers were asked to participate in a First Words elicitation session. It required naming people and objects from a children's book (Amery and Cartwright 1987) and then describing pictures that included these people and objects. There is a variety of uses for eliciting this type of data, including but not limited to the comparison of styles or dialects. Nevertheless, the data from this part is not used in this study. Finally, the participants were asked a list of Ethnic

Orientation Questions adapted from Keefe and Padilla (1987). The list includes questions regarding language learning and use, participation in cultural events and attitude towards one's ethnic background. The full list is included in the Appendix. The goal of EOQ is to gauge the orientation of the speaker toward the heritage or dominant language and culture.

Nine of the interviews were recorded by Joanna Chociej between 2009-2010 and the remaining 14 were recorded by me between 2014-2015. The heritage speakers were recruited through the Polish community network. The interviewer was always a member of the community and was introduced to the speaker via a mutual friend, so that the Observer's Paradox (Labov 1972) could be at least partially minimized.

3.4 Data extraction

The next step involved verbatim orthographic transcription of each interview done by a native Polish speaker in ELAN, a time-aligned transcription program (Brugman and Russel 2004). Then, starting with the 10th minute into each interview, the first 150 nouns and pronouns, including covert subject pronouns were marked.[4]

3.4.1 Linguistic factors

The dependent variable, *case matching*, was coded as an interaction of two variables: prescribed and observed case. If the prescribed case was the same as observed, for example prescribed genitive and observed genitive, then the variant was coded as *match*. Conversely, if the prescribed case differed from the observed one, for instance, prescribed genitive and observed accusative, then the variant was coded as *mismatch*.

---

[4] The covert subject pronouns are not included in the dependent variable; rather, they are used to compare rates of null subjects across speakers (see Section 4.4).

Tokens were coded according the following linguistic variables:

1) prescribed case:

   a. nominative,
   b. genitive,
   c. dative,
   d. accusative,

   e. instrumental,
   f. locative,
   g. vocative.

In case of any doubt, the prescribed case was verified with the on-line versions of two major Polish dictionaries *Słownik Języka Polskiego* (www.sjp.pwn.pl) and *Wielki Słownik Języka Polskiego* (www.wsjp.pl)

2) observed case:
   a. nominative,
   b. genitive,
   c. dative,
   d. accusative,
   e. instrumental,

   f. locative,
   g. vocative
   h. other (forms not ascribed to any case or calque prepositional phrases)

3) declension pattern:
   a. masculine animate singular,
   b. masculine inanimate singular,
   c. feminine singular,
   d. neuter singular,

   e. masculine plural,
   f. non-masculine plural,
   g. pronominal
   h. other (mixed declension, e.g., with borrowings)

Declension pattern coding was adapted from a traditional approach to Polish nouns (Mańczak 1956: 116 – 121). Pronouns constitute a separate factor here because their paradigms are for the most part idiosyncratic.

Polish case is either determined structurally (e.g., nominative for most subjects) or selected lexically by a verb or a preposition (e.g., verb *pomagać* 'to help' takes a dative argument). Although it has been proposed that all non-oblique cases (nominative selected by D, genitive by N and accusative by V) are determined structurally and all oblique cases (dative, instrumental, locative) are lexical

(Pesetsky 2013: 7), there are reasons not to follow such a radical theory here. Firstly, there are prepositions that clearly take genitive or accusative arguments. Committing to the structural case selection theory would require arguing that such prepositions are more noun-like or verb-like, or that there is a null N or null V head. Secondly, it will be shown later that accusative inflection behaves differently if it is selected by a verb or by a preposition. Thus, for the sake of this study, it is necessary to distinguish between these situations:

4) case selection:
   a. subject,
   b. object of a verb,
   c. object of a preposition
   d. other

The "other" variant includes structural case selection such as negation (genitive), complement to numeral (nominative for paucals and genitive for numerals higher than four), complement to an adverb of quantity (genitive), or adjunction (nominative).

The subject type was further coded according to the following factors:

5) subject type:
   a. noun,
   b. overt pronominal,
   c. null,
   d. other

The subject type "other" was chosen for example when the subject was an adjective (31), adverb, numeral or a whole TP (32):

(31) P2M29A:
dużo przybyło w roku- latach osiemdziesiątych
many came   in year- years eighties
'Many came in the eighties.'

(32) P2M47A:
tutaj jest łatwiej **spotkać ludzie**
here is   easier **meet   people**
'It is easier to meet people here'

The "other" subjects are non-nominal tokens and even though some of them carry

case inflections, they were not included for the dependent variable of matching; they are only looked at in terms of the independent variable of subject type. The rate of null subjects out of all subjects[5] is checked for correlation with the rate of mismatch.

Finally, agreement was coded on all verbs whose arguments were coded for the dependent variable. The rate of incorrect agreement (out of all potential sites for agreement) is later used to check for correlation with the frequency of mismatch:

6) verbal agreement:
      a. correct,
      b. incorrect

Polish verbal morphology is fairly synthetic, i.e. verbal agreement forms are different in past and present tense. For past tense Embick (1995) proposes the following template:

czytałaś (read.PAST.2sg.FEM)
czyt   + a   + ł   + a   + ś
Stem + Thematic Vowel + Past + Gender-Number + Person-Number
'you were reading'

The present tense template is as follows:

czytasz (read.PRES.2sg)
czyt   + a   + sz
Stem + Thematic Vowel + Person-Number
'you are reading'

Here, incorrect agreement means that the subject and the inflection on a verb did not match in at least one φ feature (gender, person or number) as in (33):

---

[5] Chociej (2011) calculated the rate of null pronoun out of all pronouns (null + overt) thus, excluding all other subject types from the envelope of variation. In my view, the envelope should be more inclusive as null pronoun is an alternative variant to any type of subject not only overt pronoun.

(33) P2M47A:

I   kiedyś   czas-**Ø**   był-**Ø-o**
and   once   time-_{sg.MASC} was-_{sg-NEUT}
'And once there was a time....'

    (34) standard:

I   kiedyś   czas-**Ø**   był- **Ø-Ø**
and   once   time-_{sg.MASC} was-_{sg-MASC}
'And once there was a time....'

The prescribed agreement sometimes takes a null form. It is the case for past 3rd person masculine singular as seen above. Again, as with the syncretic inflection on a noun, the agreement is coded as correct if it has a null form and such form is prescribed.

It has to be noted here that the speakers were given the benefit of the doubt regarding the production of cases whose forms are syncretic with another case. For example, if a noun required a genitive suffix –*a* and such was observed, then it was coded as observed genitive even though the suffix is homonymous with accusative form and theoretically a speaker might have intended to produce a (mismatched) accusative ending. The decision to include or exclude such tokens was not trivial and either alternative has both pros and cons. The overwhelming syncretism in the paradigm (both horizontal and vertical) would mean that after excluding any forms that could be ambiguous, very few forms are left.[6] On the other hand, including such tokens means that the rate of match may be falsely high. Furthermore, even if there is a clear case of mismatch, coding for the exact observed case is problematic. Hypothetically, if a construction calls for instrumental marking *dom-em* 'house' but the observed form is *dom-u*, it can be a genitive, dative, or a locative. However, there are very few cases when this is really ambiguous. Mostly, mismatch occurs in a systematic direction verified by other fully unambiguous markings (e.g. accusative – genitive pair). Nevertheless, it is remains an obvious shortcoming of the use of

---

[6] To be precise, only three forms are non-syncretic in the entire paradigm: masculine singular inanimate dative –*owi,* feminine singular dative –*e* and instrumental –*q.*

naturally-occurring speech.

Last but not least, the tokens were flagged for several other properties. These are the properties that were not expected to play a role in the variation but could be referred to at any point. For example, all copulas and verbal objects under negation were marked so that a separate analysis on these could be done in the future.

3.4.2 Excluded tokens

Even though all nominal tokens were marked, not all of them were included in the analysis. All tokens where the dependent variable of matching was undetermined, i.e., when the argument was covert, were not analyzed. Furthermore, preliminary investigation showed that the prescribed nominative nominals do not mismatch and they were omitted in the main analysis..

3.4.3 Social factors

The analysis of social factors allows us to determine which speakers are more likely to produce a given variant. Firstly, standard social factors of age and sex are analyzed for both heritage and homeland speakers. If age is a significant factor it usually means that the variable is undergoing change or it is age-graded. Significance of the sex factor in addition to the age factor is expected in the case of change in progress, especially a change from above. Sex is expected to have an effect for variables whose two variants differ in terms of prestige, in the case of stable variation.

Furthermore, for heritage speakers, contact with English is expected to have an effect on case mismatch. Therefore, predictors quantify the degree of contact in various ways. Firstly, generation was tested. It is hypothesized that since 1st generation speakers fully acquired Polish before emigration (and did not have much exposure to English prior to emigration), their rate of mismatch will be lower than 2nd generation speakers for whom English and Polish competed from early on. Secondly, the rate of mismatch in heritage speakers was also analyzed in terms of

EOQ – the overall average and several subparts. It allows for a finer distinction of speakers regardless of their generation status. It is hypothesized that the more Polish-oriented speakers will show fewer mismatches than those who are more English-oriented.

Some of the factors are collinear, i.e., usually 1st generation speakers have a higher EOQ average than 2nd generation speakers. There are gaps in the sample, e.g., by definition there is no <38 age group in the 1st generation. There are few 2nd generation speakers older than 30. Thus, these factors cannot be fed into the multivariate model at the same time and were added individually until the best model was found based on the Akaike Information Criterion (AIC).

3.5   Data analysis

The coded data was exported into an Excel spreadsheet to create a tab-delimited data file and then loaded into R (R Development Core Team 2008). The statistical package Rbrul (Johnson 2009) was used for logistic regression. Firstly, I used the step-up, step-down procedure to build various models and determine which predictors are statistically significant. Each factor group is separately added to/removed from the model and statistical tests determine whether it improves the overall model. By Occam's Razor, only the factor groups that significantly improve the model are retained. Secondly, the output of these logistic regression models shows the strength of a particular factor group in relation to others, i.e., which of the factor groups is the strongest predictor. The relative strength is gauged by subtracting the value of the lowest Factor Weight of that group from the highest one. The bigger the obtained value, *Range*, the stronger the group. Finally, the results show the effect of individual factors within each group – which ones favor a given variant and which ones disfavor it. This is measured based on Factor Weights. Factor Weights range from 0 (most disfavouring) to 1 (most favouring) with 0.5 considered a neutral cut-off point.

Furthermore, the advantage of multivariate analysis over distributional analysis is

that it takes into account the differences in the number of tokens for each factor within a group. For example, if one speaker happens to produce unusually more of type X tokens than type Y, multivariate analysis can handle such data produce results regardless of these imbalances. It is unusual to assume that all speakers behave the same thus including speaker as a random factor is desired.

Last but not least, Rbrul allows continuous factors. It means that age and EOQ scores do not have to be divided into discrete groups based on arbitrary boundaries but rather they can be tested as linear factors.

3.6 Comparing Heritage and Homeland

The first argument showing the difference between Homeland and Heritage variable inflection is Fisher's Exact Test to determine if there is a statistically significant difference between mismatching rates of the two communities. If the p value is less than 0.05, then the null hypothesis, i.e., that there is no difference between the communities, may be rejected.

The second argument is shown by qualitative comparison of variation. The differences are found in the areas in which one variety exhibits categoricty and the other variation. Finally, each area of variation found in both groups is described and analyzed separately.

A subsequent quantitative analysis of contexts where there is variation in both Homeland and Heritage Polish is the final argument of difference between the two grammars. Here, a multivariate analysis with generation as a factor (including homeland as a separate "generation") shows that it is a significant predictor with every subsequent generation producing more mismatch.

4 Results

The following section reports the distributional results of mismatch in Heritage and Homeland Polish. It shows that for Heritage Polish some cases are almost categorically realized as prescribed while others show variation. In the varying ones, there are two clear strategies – one of a default replacement and one of a specific direction of replacement also observed in Homeland. Next, removing the invariant cases allows a multivariate analysis and comparison of its results to the previously suggested implicational hierarchies (see Laskowski 1993 and Preston 1986 discussed in Background section). Testing social factors shows that there is no one predictor that is driving the entire case system reanalysis. Furthermore, correlation of the rates of case matching, correctness of agreement and null pronoun use show that only the first two are interrelated within an individual. Finally, a subanalysis of one particular pattern of variation observed in both Homeland and Heritage Polish, genitive – accusative selected by lexical verb, shows increased variation with subsequent generations and clusters 1st generation with Homeland speakers.

4.1 Pre-data analysis

The mismatch in Homeland Polish is very infrequent. Only 9 out of 1487 tokens show mismatch (~0.5%). It is only slightly more frequent in Heritage Polish with 74 mismatch tokens out of 1586 (almost 5%). An abundance of the tokens are prescribed nominative and, removing them from the analysis is justified – nominative, or citation form, matches categorically except for a few isolated tokens whose mismatching inflection is unrelated to the case selection. In the Heritage Polish corpus, there are very few tokens of *prescribed nominative* that exhibit mismatch (N=4) and these may occur as a result of unfamiliarity with the Polish word as a whole rather than a mismatch in form. Examples (35) and (36) below show that the same speaker uses the word *połowa* 'half' consistently in a wrong form. Although the nominative form in (35) is not found in the dictionary, an

analogous alternation (as in 35 & 36) is listed in the Polish dictionary for *par-a* and *par-ę*[7] for a quantifier 'couple' where the former means strictly 'two, a pair' and the latter refers to 'a few, some'. Intuitively 'half' can have a similar double meaning with where form *połow-a* refers to strictly 50% of something and *połow-ę* to roughly a big part of something.

(35) P2M29A:
    połow-ę pracowników znika
    half-ACC of-employees disappear
    prescribed:
    połow-a pracowników znika
    half-NOM of-employees disappear
    'Half of the employees disappear.'

(36) P2M29A:
    połow-ę   ludzi   w Toronto pochodzi z   gdzieś   indziej
    half-ACC   of-people in Toronto comes   from somewhere else
    prescribed:
    połow-a ludzi   w Toronto pochodzi z   gdzieś   indziej
    half-NOM of-people in Toronto comes   from somewhere else
    'Half of the people in Toronto come from somewhere else.'

Another example in (37) shows a mismatch for the word *niania* 'nanny' that could be an influence of English pronunciation of this word. The word overall is produced with Polish phonology but the presumed inflection does not come from any form in the paradigm for this word:

(37) P2F23A:
    pracowała   jako **nianie**
    she.worked as   nanny-?
    standard:
    pracowała   jako **niania**
    she.worked as   nanny-NOM
    'She worked as a nanny.'

The last nominative mismatch shows on a very rare and archaic word *ordynans* 'orderly' (from French 'ordonnance'). Again, the mismatching form is likely to be a

---

[7] http://sjp.pwn.pl/doroszewski/pare;5469878.html

result of unfamiliarity with the Polish word. The speaker code-switches, pauses and produces false-starts until arriving at this form:

(38) P2F85A:
     [FR: ordonnance]- ordynanz-- **ordynans-a,** który się zmieniał cały czas
     French CS   false-start **orderly-ACC** who REFL changes all   time
     'orderly, who was changing all the time, ... '

Considering nominative to be a default (Polinsky and Kagan 2007: 381) is supported by the fact that it never mismatches the same way other case markings do. Thus, prescribed nominative tokens are outside the envelope of variation and will not be included in the analysis.

This step results in reducing the data to 907 Homeland tokens and 945 Heritage.

4.2   Distributional results

After trimming the data, the overall frequency of mismatch for Homeland Polish is around 1% (9/907) and for Heritage Polish over 7% (69/945). Fisher's exact test shows that the difference between two varieties is extremely statistically significant (p<0.0001).

4.2.1   Heritage Polish - linguistic factors

4.2.1.1   *Type of declension*

The cross-linguistic pattern of case being preserved more on pronouns than on nouns is found here. In Heritage data, there is a significant difference in mismatches depending on whether the inflected nominal is a pronoun or a noun (Fisher's Exact Test p=0.0351). Only 3% of pronouns (6/186) but over 8% of nouns (63/759) are mismatching suggesting that inflected forms of pronouns are easier for Heritage speakers to acquire and retain than nouns. The breakdown of nouns into their prescribed declension patterns reveals some mixed patterns, see Table 6. Plural masculine inflection shows the most mismatch (16% 8/49) closely followed by neuter (15% 13/85). The least mismatch is found with feminine nouns (6% 17/269). The frequency of mismatch does not align with the degree of syncretism

within a declension type. It does not confirm Laskowski's (1993) hypothesis either that the animacy distinction is lost. If it were, one (or both) of the singular masculine nouns would show most mismatch. A possible explanation is that heritage speakers have not mastered the grammatical gender of a noun they are declining. This hypothesis is only partially testable within my data – the agreement is coded and correctness of agreement with the subject's grammatical gender shows only in past tense. The correlation of rate of case mismatch and rate of agreement is shown later in the section. Since the source of the mismatch within each declension pattern is ambiguous, it will not be considered for a multivariate analysis.

	%	N			
pronoun	3%	186			
noun	8%	759	(detailed noun)	%	N
			plural masculine	16%	49
			neuter	15%	85
			masculine animate	11%	46
			masculine inanimate	7%	175
			plural non-masculine	6%	135
			feminine	6%	269
TOTAL		945			

Table 6. Distribution of case mismatch in Heritage Polish by nominal type with detailed breakdown of declension type for nouns.

4.2.1.2 Case selector

Table 7 shows that the mismatch in Heritage data is also more frequent when a verb selects the case, e.g., nominal object of a verb, (10%; 35/348) rather than a preposition (under 6%; 28/470). It is even more rare when the case is structurally determined (under 5%; 6/127), i.e. in a given structure the prescribed case is always the same (e.g., N modifiers are always GEN, negated objects are always GEN). Recall that Laskowski (1993) also reported relative correctness of the marking produced when the noun is selected by a preposition. As for the appropriateness of the structurally selected forms, the low frequency of mismatch is not in line with the previous findings. Laskowski (1993)'s Swedish Polish data showed a substantial amount of mismatching in case marking selected by negation and non-possessive

adnominals. In Toronto Polish, none of the mismatching cases are under negation. The six mismatching tokens are instances of four non-possessive adnominal selection, one numeral selection and one by a subject expressing insufficiency.

	%	N
verb	10%	348
preposition	6%	470
other (structural)	5%	127
TOTAL		945

Table 7. Distribution of case mismatch in Heritage Polish by case selector.

4.2.1.3   Prescribed case vs. Observed case

Table 8 below shows a cross-tabulation of prescribed cases with their observed forms in Heritage Polish. The number of matches (grey-shaded cells) is consistently much larger than their corresponding mismatches, suggesting that there is no case that is missing from the Heritage paradigm.[8] Some of the most frequent mismatches (non-shaded cells) are in the pairs of *prescribed GEN-observed ACC* (N=19) and *prescribed INSTR-observed NOM* (N=7). These two types of mismatches are limited to a narrow context – GEN-ACC always occurs when case is selected by a lexical verb and six out of seven INSTR-NOM when it is selected by the copula 'to be'.

There is also a fair number of mismatches where the NOM is used instead of any prescribed case, except for *prescribed DAT*. In fact, *prescribed DAT* almost categorically matched the observed case (98% 47/48). The only case of DAT mismatch is the replacement with a prepositional construction.[9] It counters Laskowski's (1993) implicational hierarchy by which dative is the most affected case. Recall the methodology, where observed cases that are ambiguous due to syncretism are coded as correct as long as the form matched the prescribed one. It is

---

[8] There is no instance of *prescribed vocative* in the Heritage data, thus I will not make any claims about its status. It is the nature of a sociolinguistic interview that this form is unlikely to be produced.

[9] Labeling the use of prepositional constructions in place of a simple dative as mismatch is not unquestionable. For Laskowski (1993) such tokens indicate dative loss but there seems to be no prescriptive rules for such context.

not the case that dative has any more syncretism than other cases and thus has a falsely inflated match frequency. The possible confound is that only 5 of the prescribed cases are nouns and the rest are pronouns – already shown to preserve case well. This is nevertheless characteristic of natural speech and a similar ratio is attested for Homeland Polish as well (44 dative pronouns and 4 nouns).

		prescribed case					
		ACC	**DAT**	**GEN**	**INSTR**	**LOC**	Total
	ACC	257	0	19	0	1	277
	DAT	0	47	2	0	0	49
observed case	**GEN**	0	0	256	0	1	257
	INSTR	0	0	0	113	1	114
	LOC	0	0	1	2	203	206
	NOM	9	0	9	7	12	37[10]
	Other[11]	2	1	1	1	0	5
	Total	268	48	288	123	218	945

Table 8. Cross-tabulation of prescribed and observed case forms in Heritage Polish

To summarize, there are two distinct types of mismatch. One displays a particular direction of variation within a narrow context (GEN-ACC and INSTR-NOM) and the second shows a default mismatch where nominative replaces any other case regardless of the context[12]. The distinction into two types is supported by the fact that only a subset of speakers implements the default NOM-substitution. These are P2M47A (most prominently, N=26) as well as P2F26A (N=2), P2M22A (N=1) and P2M29A (N=2). They all fall in the lower range of EOQ score but other speakers of similar EOQ averages do not exhibit such rates of mismatch. There may be a more

---

[10] Recall that the additional 641 matching nominative tokens were excluded from the analysis.

[11] This category includes a rare instance of inflection that does not match any case – it may be a misproduction or some hybrid form or prepositional constructions, possibly calques.

[12] INSTR-NOM under copula is considered to be a special type of mismatch rather than a part of default nominative mismatch for three reasons: (1) its narrow context, (2) a parallel tendency observed in other Slavic languages and (3) the fact that the INSTR-NOM under copula variation is found for some speakers that do not show the nominative default variation.

fine-grained explanation based on their background or they did not bother to avoid using an inflection that they find problematic.

The remaining mismatches found in other pairs are rare and do not form a pattern.

4.2.2 Homeland Polish

Nine[13] mismatches in Homeland Polish come from the following pairs (prescribed – observed):

- GEN-ACC 4 tokens
- INSTR-NOM 2 tokens
- VOC-NOM 3 tokens

This is very similar to the Heritage data. The first two pairs of mismatches, accounting for 67% of the mismatches, are exactly the same as in Heritage data. The VOC-NOM pair was unattested in the Heritage data because speakers did not produce any of the *prescribed vocative* forms.

One important difference is that in the Homeland data there is no general pattern of a NOM default form taking the place of another prescribed case. I argue that NOM in the INSTR-NOM pair is not always the default form because this mismatch in Homeland Polish is restricted to one particular construction, object of a copula. In contrast, in the Heritage data[14] there are examples of INSTR-NOM mismatch in other environments as well.

---

[13] The fossilized archaic forms that are mismatching according to general modern rules of grammar but are nevertheless the only acceptable forms within a particular phrase are not counted as a mismatch. For example, the noun *mąż* in the phrase *wyjść za mąż* 'get married' (lit. 'go after a husband') looks like a NOM in place of a prescribed ACC selected by a preposition but using an ACC form would yield a difference in meaning,

[14] There is only one non-copula INSTR-NOM token in the coded subset of data but there are a few more such instances beyond this subset. The are few tokens of non-copula INSTR overall (compared to other prescribed cases), thus a mismatch is difficult to find in a small sample like this one.

4.2.3 Social factors

Table 9 shows that there are three heritage speakers who do not show any mismatch in their sample[15], while the extreme mismatcher reaches over 50%.

speaker	% of mismatch	N	EOQ average	EOQ identity	EOQ language	EOQ lg choice	EOQ heritage	EOQ parents	EOQ partner	EOQ culture	EOQ discriminat.
P1F67A	2	81	1.28	1.20	1.40	1.25	2.00	2.00	1.33	0.67	0.25
P1M44A	0	67	1.44	1.50	1.67	NA[16]	2.00	2.00	0.00	2.00	2.00
P1M88A	4	85	1.12	1.00	2.00	0.67	1.33	NA	0.00	NA	NA
P2F18A	0	69	1.19	1.20	1.43	1.50	0.50	1.75	NA	1.33	0.60
P2F22A	10	62	1.15	1.40	1.29	1.00	0.75	2.00	0.50	1.67	0.40
P2F23A	0	79	1.06	0.80	1.00	1.25	1.50	1.50	2.00	1.00	0.00
P2F26A	7	74	0.64	0.40	0.57	0.75	0.75	1.75	0.00	0.67	0.00
P2F85A	1	67	0.65	0.40	0.43	0.50	1.25	2.00	0.00	0.00	NA
P2M21A	6	67	1.61	0.80	1.86	1.50	0.50	1.67	1.00	1.33	0.25
P2M22A	16	63	1	0.40	1.14	1.25	1.00	2.00	1.00	1.67	0.20
P2M23A	2	92	0.88	0.60	1.14	1.50	0.50	1.50	0.50	1.00	0.20
P2M29A	6	79	1.07	0.40	1.29	1.50	0.50	2.00	NA	1.33	0.80
P2M47A	52	60	0.86	1.00	0.67	1.00	0.25	1.00	NA	2.00	2.00
correlation with % mismatch		R	-0.31	-0.27	-0.22	-0.22	-0.47	-0.03	0.04	0.36	0.03
		p	0.3	0.37	0.47	0.48	0.1	0.93	0.88	0.23	0.93

Table 9. Rate of mismatch and EOQ scores per heritage speaker

As predicted, female speakers (3% 14/432) mismatch less than male speakers (11% 55/513) and the difference is statistically significant (p<0.0001). 1st generation speakers mismatch only 2% of the time (5/233), while the 2nd generation mismatches at almost 9% (64/712). While the Fisher's exact test result suggests that the generation difference is significant (p=0.0002), there are many possible interactions. The multivariate analysis provided later in this section addresses this

---

[15] The subanalysis of GEN-selecting verbs also shows that these 3 speakers do not have any mismatches; however, for P1M44A none of these verbs were present in his speech and for the two 2nd generation speakers these verbs were infrequent as well (2 and 3 tokens).

[16] The empty cells means that the speaker did not answer any/enough of the questions from this subset.

issue. The Spearman correlation tests show that overall average Ethnic Orientation Score only moderately correlates with the frequency of mismatch and is insignificant (r=-.31, p=0.3). The subsets of EOQ score pertaining to ethnic identity (who you'd say you are, who are your friends, neighbors, coworkers) and language choice (which language do you speak with your family, friends, when you're angry) correlate a bit more strongly (r=-.26 and -.41 respectively) but they are not significant (p=.41 and .84 respectively). It suggests that the social predictors are much more complex. The Discussion section examines potential reasons for not finding an effect of social factors and suggests other approaches.

4.3   Multivariate analysis

The results of distributional analysis reported so far show that case matching may in fact be several distinct variables, i.e. verbal GEN, copula INSTR and default NOM. It means that each of the variables may be subject to different conditioning thus each would require a separate multivariate analysis. One such narrow analysis is presented later in Section 4.5 for GEN with lexical verbs. However, much of the previous literature treats case matching holistically and based on overall frequencies to produce some sort of implicational scales (*cf.* Laskowski 1993 and Preston 1986). The following multivariate analysis takes a similar approach so that the results can be compared across these studies.

Recall from the Methodology section that there were six linguistic factor groups coded. Prescribed case, nominal type and case selector are expected to have an effect on the dependent variable of case matching. Some of the factors had to be honed down in a way justified by linguistic theory in order to remove categorical effects. Prescribed obliques contain nearly categorical and relatively infrequent datives as well as locatives and instrumentals and these three cases are at the bottom of the both the implicational scale for Laskowski (1993) and frequency ranking for Preston (1986). Therefore datives, instrumentals and locatives were recoded together as obliques. Nominal type was recoded into pronoun vs. noun from the previous Declension type, which turned out to be too detailed for the

current amount of tokens (too many gaps while cross-tabulating with, e.g., prescribed case). The remaining three coded factors groups are not expected to have a direct influence on the variable. Observed case was coded to determine either match or mismatch variant of the dependent variable and further, the direction of the mismatch. Verbal agreement and subject type were coded to test their correlation with the rate of mismatch (presented in the next subsection) but their role in conditioning the variation in case marking is not the same as, e.g., prescribed case.

All tested linguistic factors were significant (see results of the logistic regression analysis in Table 10). Prescribed case was the strongest predictor, followed by the Case selector and Nominal type. Within the Prescribed case factor group, the ordering of the factors show that genitive prefers mismatch even more than oblique cases. This is surprising given previous implicational hierarchies where oblique cases followed the grammatical cases of genitive and accusative. I argue that is an artifact of a heavy interaction of prescribed case and context. Genitive under negation[17] is not affected the same way as genitive modifying a noun thus any type of case-based implicational scale is missing the point.

Within other factor groups, the results are consistent with the predictions. When case is selected lexically by a verb, the mismatch is favoured. A similar tendency is found when the case is assigned to a non-nominative subject. Genitive case is selected by a particular type of verb indicating insufficiency (as in example 6). Thus, depending on the theory, it may be viewed as lexical or structural. This is, however, based on a single mismatching token thus the trend would need to be verified in a larger sample. Finally, case selected by a preposition or determined purely structurally disfavors mismatch. Finally, there is more mismatch on nouns than on pronouns. Both of the results are in line with previous observations of changing case systems.

---

[17] Recall from the Background section that negation "changes" every accusative argument into genitive.

The results of logistic regression for linguistic factors, presented in Table 10, confirm much of the distributional analysis patterns in Heritage.

Variable (case match) in Heritage Polish				
probability of *mismatch* for linguistic factors				
factor group	logodds	N	% mismatch	FW
Prescribed case				
GEN	1.400	287	11%	0.80
OBL	0.625	387	6%	0.65
ACC	-2.025	266	3%	0.12
			Range	*68*
Case selector				
verb	1.307	345	9%	0.79
subject	0.759	9	11%	0.68
preposition	-0.523	469	6%	0.37
other	-1.543	117	3%	0.18
			Range	*61*
Nominal type				
noun	0.7	754	8%	0.67
pronoun	-0.7	186	3%	0.33
			Range	*34*

Table 10. Results of logistic regression for linguistic factors in Heritage Polish. Speaker was run as a random factor. N = 940 due to honing of the model. Centered input probability = .01. AIC = 337.161

Interestingly, none of the tested social factors in a multivariate analysis was found significant. The factors were added to the honed linguistic model (as in Table 10 above, with Speaker as a random factor) individually (categorical: sex, generation; continuous: age, EOQ, EOQ language choice, EOQ ethnic identity) as well as an interaction (generation*EOQ) but none of them improved the model significantly.

4.4 Correlation of mismatch and other morphosyntactic variables

The results of correlation between rate of mismatch and rates of other reported heritage variants are given in Table 11. A Spearman's Rho test shows that case matching correlates with agreement – the more mismatch an individual exhibits, the more incorrect agreement on a verb (R=.77 p=.002). It is too early to hypothesize whether one is causing another or the two just happen to be affected at a similar rate. On the other hand, the frequency of overt subject does not correlate with the

rate of mismatch (R=-.077 p=.8). This finding is not surprising considering the fact that the use of null/overt subject is variable in Polish and conditioned by multiple linguistic factors. Chociej (2011) shows that the change in rate of overt subject use in Heritage Polish, from one generation to the next, is not large but there seem to be two distinct variable grammars within this group based on the difference in significance and relative ordering of the factor groups .

speaker	mismatch	incorrect agreement	overt subject
P1M44A	0.00	0.00	0.80
P2F18A	0.00	0.00	0.44
P2F23A	0.00	0.00	0.41
P2F85A	0.01	0.01	0.75
P2M23A	0.02	0.00	0.40
P1F67A	0.02	0.03	0.62
P1M88A	0.04	0.00	0.57
P2M21A	0.06	0.01	0.38
P2M29A	0.06	0.03	0.50
P2F26A	0.07	0.24	0.66
P2F22A	0.10	0.01	0.55
P2M22A	0.16	0.08	0.30
P2M47A	0.52	0.10	0.79

Table 11. Rates of mismatch, incorrect agreement and overt subject for each heritage speaker.

4.5   Comparing Heritage and Homeland within GEN-ACC context

The number of factors and contexts for which Heritage Polish shows variation while Homeland Polish is categorical shows that the two systems are distinct. The most obvious area is the heritage substitution of any case with a default nominative form regardless of the context – such variation is not present in Homeland Polish. However, I argue that, in some aspects, the Heritage system is varying just like the Homeland system. Two robust examples are genitive-accusative mismatch with lexical verbs and instrumental-nominative mismatch with copula 'to be'. The mismatch of genitive-accusative is a handy example to compare whether the rates of mismatch are the same for the two varieties of Polish. In order to perform this analysis, it was necessary to gather more tokens of his type. Therefore, a number of lexical verb stems requiring a genitive argument extracted from the entire corpus.

These were:

- *uczyć* 'teach/learn',
- *używać* 'use'
- *szukać* 'search'
- *potrzebować* 'need'
- *udzielać* 'to grant'

The nominals for which the above verbs selected genitive were coded for the dependent variable of match as in the rest of the study. However, ambiguous marking was not included - there are no singular animate masculine and plural masculine nouns as these two declension types show GEN-ACC syncretism.

The results of the subanalysis show that Homeland speakers mismatch 30% of the time (tokens N=23) and Heritage speakers much more often: 67% of the time (tokens N=73). A multivariate analysis of this variation (Table 12 below) shows that there is an effect of verb as well as generation.

The verb 'need' and 'use' favour mismatch and 'learn/teach' and 'search' disfavor it. It is expected that the less frequent verbs favour mismatch but the present sample shows that while 'need' has the lowest N, 'use' is the second most frequent verb of all those selected. It is too small of a sample to argue for the frequency effects but it is one of the directions that a future study could take. The ordering of generations with the 2nd generation favouring mismatch followed by the 1st generation disfavouring mismatch similarly to generation 0 (homeland) suggests that in this context, 1st generation heritage speakers clusters with homeland variety while 2nd generation diverges.

Similarly, future work could also include an expanded analysis of the INSTR-NOM under copula mismatch pattern, with the addition of more tokens.

Variable (case match) in Polish				
probability of mismatch for Verb and Generation				
factor group	logodds	N	% of mismatch	FW
Verb[18]				
potrzebować	1.223	6	67%	0.77
używać	0.42	32	72%	0.60
uczyć	-0.409	38	55%	0.40
szukać	-1.234	18	33%	0.23
			Range	*54*
Generation				
2	1.082	61	69%	0.75
1	-0.343	10	50%	0.42
0	-0.739	23	30%	0.32
			Range	*43*

Table 12. Multivariate results of mismatch in the genitive-accusative context. N=94, centered input= .51, AIC= 121.214. Not significant: sex.

5 Discussion

The main picture that emerges from the results is that the case system in Heritage Polish is fairly similar to Homeland Polish. The areas that show the largest divergence from the prescribed norms are precisely the same ones as in Homeland Polish. It stresses the need to situate findings from any heritage data study in its respective homeland situation and for careful consideration of methodology for homeland data analysis. While some studies provide a reference point with the homeland data (Polinsky 1995, Preston 1986, Laskowski 1993), this is often a very limited, non-systematic comparison with one or two speakers involved.

Divergence from Homeland Polish is found only in four of the 13 heritage speakers They show variable nominative substitution for any other case seemingly not constrained by context. This is not attested in Homeland Polish; thus, I argue that it constitutes a unique heritage type of mismatching. This replacement with a default has been shown for other heritage languages as well (Polinsky and Kagan 2007).

---

[18] The verb *udzielać* is categorically mismatched but its low N=2 makes the result unreliable so it was not included in the multivariate analysis.

Nevertheless, the scale of the phenomenon is not as large as it may appear. Focusing on the entire case system (rather than its most vulnerable areas), all heritage speakers (rather than only on those with most severe attrition) and naturalistic speech (rather than experiments), suggests than as a whole, the Polish Heritage community in Toronto used Polish case inflection similarly to Homeland speakers. There may be advantages of alternative methodological approaches. Narrowing down the envelope of variation to non-nominative cases would save a lot of time. Testing only the less proficient speakers would yield more mismatching tokens therefore, more detailed question could be asked, e.g., regarding frequency. Finally, experiments would target problems with speakers' competence than otherwise can be cleverly hidden by strategies that avoid producing some structures.

On the other hand, the inclusive approach allows positing a question of how the heritage community varies. Heritage Polish individuals show a wide range of mismatch, from 0 to up to 50%. They also show a difference in the mismatching strategies used. While pretty much all speakers show genitive-accusative variation, only some speakers show the default nominative substitution. Correlating the frequency of mismatch or the use of particular mismatch to any social factors is not a trivial task. Responses to the Ethnic Orientation Questionnaire, generation, age, and sex are not by themselves strong predictors. Most of the social factors are highly collinear (1st generation speakers usually have a higher EOQ score than 2nd and by definition includes speakers of older age groups only). None of them is a significant factor in this variation.

I argued that the results show two different types of mismatch – homeland-like and default nominative. If these are, indeed, at least two separate variables, then each of them needs a separate analysis of social predictors. Generation was found to be a significant factor in the subanalysis for prescribed genitive with lexical verbs but it may not be for the default nominative type of mismatch, hence the apparent lack of effect in the general mismatching analysis. Since only a subset of 2nd generation heritage speakers show this variation, perhaps the threshold can be found based on

some parts of EOQ. Similarly, EOQ may still be a significant predictor for the homeland-like variation, whether prescribed genitive with lexical verbs or prescribed instrumental with a copula, but since this variation seems to be socially conditioned in a very different way, the EOQ effect is likely to be different.

If the mismatch comprises two distinct variables then their envelopes of variation may be, but are not necessarily, overlapping. For example, for the variation of GEN-ACC selected by a lexical verb the variable context is precisely only the prescribed genitive selected by lexical verbs. However, as the default nominative mismatch is supposed to be context-free (regardless of case selector or prescribed case) the envelope of variation is as inclusive as in the present study. Yet there is no robust EOQ effect – even after trying so many ways of slicing the pie, the correlations are not strong. The premise of the EOQ score is to show the speakers' involvement in the local heritage community and by extension their exposure to the heritage language. A possible issue with EOQ may be that the questions being asked do not target the trigger of variation. As every community centralizes around different values, additional ethnographic study would be desired. For the Polish community in Toronto, the emergent topics are traditions, religion, entertainment and politics. Additional, potentially valuable questions to ask would involve:

1. Do you celebrate holidays like Christmas and Easter in traditional Polish ways?
2. Do you cook/eat Polish food and shop in Polish stores?
3. Are you religious? Do you attend Polish church?
4. Do you attend Polish festivals/events?
5. Do you listen to Polish bands/watch Polish movies?
6. Do you follow Polish politics or vote in national election?
7. Have you taken Polish classes at university?

I expect that especially questions 3-7 target the exposure to Heritage Polish. Positive answers to these questions require at least a very solid passive command of Polish. The last question also addresses proficiency in producing Polish utterances.

A few of the questions in the existing EOQ questionnaire also aim at a similar effect of exposure to the heritage language (e.g., question B3 *Do you prefer to speaker Polish or English?*) and yet no effect is found. The issue may be that the self-reports are not reliable or that quantifying the orientation is not accurate. Better results might emerge with weighted average or a larger range of scores than 0-1-2. It may be connected to an observation from heritage language pedagogy that it is persistently difficult to determine what qualities a successful heritage language learner possesses (Parra 2015).

5.1   Universal or language-specific source of variation?

The goal of researching the same variable within different heritage communities is to reveal a larger, perhaps universal, linguistic pattern. Such patterns are not immediately apparent if language-specific mechanisms obscure them. Laskowski (1993: 136) argues that in Swedish Polish case marking functional issues override universal tendencies. His implicational hierarchy of attrition NAGILD (recall the description in Section 2.3) corresponds to the stated order of acquisition in both heritage and monolingual communities. Based on this parallel, he infers that the order of acquisition indicated the markedness of case inflection. The more complex the function of a particular case, the more marked it is. Since for each language the order of case acquisition is different,[19] the markedness of each case must be different across languages. His conclusion is that the disappearance of cases in heritage languages is functionally motivated by their language-specific markedness. The Swedish Polish hierarchy of case inflection is surely not universal as there is nothing inherently weak about, e.g., dative inflection (cf. modern German dialects mentioned in Section 2.3, where dative replaces genitive). Due to lack of acquisition data from Toronto Polish, I cannot make the same claims.

---

[19] *cf.* Gvozdev (1961) for a very careful study of the closely related language Russian which shows the following order of acquisition: 1. NOM, ACC, GEN; 2. DAT LOC 3. INSTR.

Frequency effects may be shaping the variation in case marking as well. There is a cross-linguistically common tendency of case retention on pronouns as opposed to nouns (e.g., in English as well as Swedish Polish). There are overall fewer pronouns than nouns in a language thus the rules and forms of declension for the latter may be costly in terms of the acquisition and memory load. This is further supported by the fact that even within nouns, the declension types that are more frequently used also show less mismatch; plural masculine inflection shows the most mismatch (16% 8/49) closely followed by neuter (15% 13/85) and the least mismatch is found with feminine nouns (6% 17/269).

Finally, paradigm-internal motivation for case system change is a possibility to consider as well. Miodek (2000) suggests that the genitive-accusative mismatch is the result of syncretism spreading from masculine animate singular declension onto all others. Natural language allows a large degree of ambiguity and thus variation apparently producing ambiguity is not uncommon (Labov 1987). On the other hand, Laskowski (1993: 156) points out that the syncretism of dative-accusative in the feminine declension does not spread – there is no mismatch of dative-accusative in other declension types. Therefore, arguing for syncretism spread would have to explain why it happens in one case but not the other. It would require postulating that there is something special about the masculine animate declension type (or just masculine and animate features) and all other declension types copy its syncretic pattern. Alternatively, It could be that the GEN-ACC syncretism is necessary but not sufficient. It may be the case that the GEN-ACC mismatch is favored because using the genitive as the complement to a verb is an unusual and marked construction; thus, there is motivation to replace it with a less marked case such as the accusative. The syncretism comes handy in this case. On the other hand, dative complements of verbs are perfectly fine and there is no need to take advantage of syncretism to replace the dative with the accusative.[20]

---

[20] This clever suggestion is thanks to Aaron Dinkin, the Second Reader.

Solving the puzzle of an internally-driven change does not preclude asking a question about broader design. I argue the detailed hierarchies like NAGILD are irrelevant for universal patterns. Stepping away from precise ordering immediately allows us to notice that, cross-linguistically, the citation form (nominative or absolutive) is very stable, followed by other grammatical cases, with oblique being far behind (Polinsky and Kagan 2007: 381). This supports the distinction in the theory of case between the grammatical case (syntactically determined) and the obliques (lexically determined). A strong version of this hypothesis would suggest that heritage speakers' syntax is stronger than their lexicon.

5.2 Qualitative and quantitative difference across heritage generations

Heritage speakers are not a homogenous group. The greatest difference is seen in terms of generation. From the perspective of the critical period hypothesis (Penfield and Roberts 1959), speakers of the 1st generation, born and raised in Poland, should be qualitatively different than speakers of the 2nd generation, raised in the English-dominant environment. Since the very basics of Polish noun inflection are not acquired until the age of 3 (Smoczyński 1955: 207) and the full development of syntactic competence is not complete until much later (age 7 according to Kaczmarek 1966: 68 or even age 10 according to Aichtison 1991: 105), then the fact that 2nd generation speakers are inevitably exposed to English early, and concomitantly, receive less heritage language input than a monolingual, may have some impact.[21] This can be evaluated based on the categoricity of the default nominative type of mismatching. Homeland speakers do not show this pattern and neither do 1st generation heritage speakers, contrary to 2nd generation speakers. Thus, in this area 1st generation speakers cluster with Homeland, in opposition to the 2nd generation. However, even the generational split is not enough to divide the heritage speakers according to linguistic behaviour. Many 2nd generation speakers

---

[21] An alternative outcome would be to maintain two separate grammars for English and Polish. A brief inquiry into the experimental bilingual literature suggests that there is always some shared area of syntax, though the degree varies (see: Golestani et al. 2006, Hartsuiker et al. 2004, Marian et al. 2003).

also do not exhibit this default type of mismatching. At this point, there is no clear factor that is causing the distinction within the 2nd generation. It may be qualitative again like the type of bilingual acquisition (simultaneous or sequential) or more fine-grained quantifiable factors like the amount and intensity of exposure mixed with ethnic orientation discussed earlier.

Finally, a different method of comparing across generations may be applied where there is already variation in homeland speech. For Homeland Polish, genitive-accusative replacement with lexical verbs as well as instrumental-nominative replacement with copula verbs variation is present. The subanalysis of the genitive-accusative variation showed that the Heritage speakers outdo Homeland speakers and produce twice the amount of mismatch. A closer look at the distinction between generations in the distributional analysis shows almost equal steps of increasing mismatch in subsequent generations (Homeland 30%, 1st generation Heritage 50% and 2nd generation 69%) but the differences in Factor Weights in a multivariate analysis (see Table 10) suggest that 1st generation clusters with Homeland, not with 2nd generation. (Homeland FW=0.32, 1st gen FW=0.42, 2nd gen FW=0.75).

5.3   Separability of the variables

Even though the current study looked at nominal inflection, I do not claim that the case system operates in a void. In fact, it is an even more interesting question to research whether different variables change in parallel or perhaps are intertwined. Laskowski suggests that a part of the inflection change that he observed in Swedish Polish is related to the lack of animacy distinction. I have shown that the rate of mismatch correlates to the incorrect use of agreement on verbs. Even though agreement analysis in Heritage Polish was not the primary focus of the analysis, the immediate post-hoc observation was that out of the three features that are manifested via agreement, gender, number and person, mostly gender was incorrect while the other two were unaffected. If case mismatch and gender agreement turn out to be the only two factors within the domain of morphosyntax that correlate with each other, then it may be argued that it is a result of incorrect gender *or* case

as the two are synthetically encoded by an inflectional suffix. Then, the competence in gender of noun would need to be tested for the same speaker. Only if the noun gender is not affected, the case mismatch assumption can be saved.

Furthermore, the cross-linguistic tendency to substitute rich case paradigms with fixed word order is uncontroversial (Blake 1994). For Heritage Polish a possible change in word order is potentially a missing link. However, finding a change in word order in Toronto Polish would not instantly suggest the internally-motivated causal link of the two variables – it may also be the influence of the dominant English and its fixed word order. Finally, a variable such as word order is itself is highly complex. Polish has the basic SVO pattern but all other permutations surface in the language due to scrambling. However, scrambling does not happen solely to put arguments in focus position - similarly to pronominal subject that cannot be said to be null by default and overt only for emphasis (Chociej 2011). The fact that the present study failed to find correlation between the rate of mismatch and null subject use may stem exactly from the complexity of the two variables. An interesting approach to investigate this issue would be to compare whether the speakers who use the default type of case mismatch are the same ones whose variable grammar for null subjects differs most from the Homeland population. It is an empirical question that could be asked as a follow-up to this study together with other ideas for extending the study that are described in the subsection below.

5.4   Further directions

Several potential ways of improving the analysis of case mismatch were mentioned throughout the paper – multivariate analysis of INSTR under copula, elaboration of the EOQ testing, testing the grammatical gender competence of Heritage speakers as well as in-depth analysis of the link between case matching and other morphosyntactic variables. Furthermore, more thorough investigation of case marking variation within particular contexts is needed. It is apparent from this as well as previous studies that the change in a particular case does not proceed uniformly across contexts. For example, Laskowski (1993) reports substantial

attrition of adnominal genitive. Laskowski and Preston (1985) further show a less severe attrition of genitive under negation. However, there is no indication of change in genitive governed by a preposition. Thus, genitive contexts are not all equal and a quantitative analysis of the interaction of case and context would provide a better insight that the hierarchies of case loss. My coding schema only flagged certain contexts. Some of them are relevant as they select genitive (negation, complement to a numeral) or instrumental (copula) that exhibit variation.

The dative's status in heritage Polishes around the world remains unresolved. On one hand, Preston and Laskowski talk about its vulnerability and replacement with other cases or structures. On the other hand, it is almost categorically matching in my data. Since the rate of dative use does not seem any lower in Toronto Polish than it is in Homeland Polish, the idea that Heritage speakers avoid producing such structures and choose to express their thoughts with the use of alternative simpler structures seems unjustified. The ultimate argument verifying the dative form's status in Heritage Polish grammar would be to conduct an experiment that targets especially this construction. It is not clear though which particular experimental test would work best in the heritage situation. Polinsky and Kagan (2007: 380) observe that heritage speakers have a 'yes' bias or perform at chance at grammaticality judgments tests that do not pose any problems for homeland speakers. It further confirms that no methodology is all-around superior when testing heritage speakers and the difference in results may be exactly the artifact of methodologies used.

6 Conclusions

The paper looked at the variation in case marking in the naturalistic speech of Toronto Heritage Polish speakers and compared it to the case marking used by Homeland Polish speakers in the same conversational setting. The dependent variable of matching was determined by comparing the observed produced form to the prescribed one. The immediate picture is that there is very little mismatch in Heritage data (under 5%) and even less in Homeland sample (~0.5%). The match is

categorical for prescribed nominative case in both groups. The two groups exhibit the exact same type of variation with lexical verbs and copula objects. Furthermore, some speakers of Heritage Polish show an additional type of variation not found in Homeland Polish – replacing any case with default nominative form. None of the social factors in the Heritage data were found to be a significant predictor of the overall rate of mismatch. On the other hand, rate of mismatch was found to correlate with the rate of incorrect use of agreement. Finally, an analysis of a subset of data, genitive-selecting verbs, suggests that 1st generation Heritage speakers are more like Homeland speakers than like 2nd generation Heritage speakers.

References

Aitchison, J. 1991. *Ssak, który mówi: Wstep do psycholingwistyki*. Wydawnictwo Naukowe.

Albirini, A., Benmamoun, E., and Saadah, E. 2011. Grammatical features of Egyptian and Palestinian Arabic heritage speakers' oral production. Studies in Second Language Acquisition 45.

Amery, H., Cartwright., S. 1987. First 100 Words. Usborne, London.

Anderson, R. 2001. Lexical morphology and verb use in child first language loss. A preliminary case study investigation. International Journal of Bilingualism 5, 377–401.

Au, T., Knightly, L., Jun, S. and Oh., J. 2002. Overhearing a language during childhood. Psychological Science 13, 238–243.

Benmamoun, E., Montrul, S., and Polinsky, M. 2013. Defining an "ideal" heritage speaker: Theoretical and methodological challenges Reply to peer commentaries. *Theoretical Linguistics*, *39*(3-4), 259-294.

Blake, B. J. 1994. Case. Cambridge textbooks in linguistics.

Brugman, H., Russel, A. 2004. Annotating Multimedia/ Multi-modal resources with ELAN. In: Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation. Retrieved from: http://tla.mpi.nl/tools/tla-tools/elan/

Chociej, J. 2011. Polish pro-drop: English influence on heritage Polish in Toronto. University of Toronto Generals Paper.

Doboszyńska-Markiewicz, K. 2011. Operator adnumeratywny ponad czyli nieco więcej niż 'więcej niż'. Linguistica Copernicana. Retrieved on August 18, 2015 from: http://www.linguistica.umk.pl/teksty/05_dobosz_mark.pdf

Dzięgiel, E. 2009. Zakres alternacji celownika i konstrukcji przyimkowej z dla we współczesnej polszczyźnie południowokresowej. In: Polszczyzna za granica jako jezyk mniejszosci i_jezyki mniejszosciowe w_Polsce, red. E. Dziegiel, A. Zielinska, Warszawa.

Embick, D. 1995. Mobile inflections in Polish. In *Proceedings-NELS* (Vol. 25, pp. 127-142). University of Massachusetts.

Fenyvesi, A. 2005. Hungarian in the United States. *Hungarian Language Contact Outside Hungary. Amsterdam: John Benjamins Publishing Company*, 265-318.

Golestani, N., Alario, F. X., Meriaux, S., Le Bihan, D., Dehaene, S., & Pallier, C. 2006. Syntax production in bilinguals. *Neuropsychologia*, *44*(7), 1029-1040.

Groniowski K., Skowronek, J. 1987. *Historia Polski 1795-1914*. Warszawa.

Gruszczyński, W., Bralczyk, J. (Eds.). 2002. *Słownik gramatyki języka polskiego*. Wydawnictwa Szkolne i Pedagogiczne Spółka Akcyjna.

Gvozdev, A. N. 1961. Voprosy izučenija detskoj reči.Moscow: Izd. Akad. ped. nauk.

Harrison, B. 2000. Passing on the Language: Heritage Language Diversity in Canada. http://www.statcan.gc.ca/pub/11-008-x/2000002/article/5165-eng.pdf Retrieved on August 10th 2015

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. 2004. Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science*, *15*(6), 409-414.

Hulsen, M. E. H. 2000. *Language loss and language processing: Three generations of Dutch migrants in New Zealand*. Dissertation

Johnson, D. E. 2009. Getting off the GoldVarb standard: Introducing Rbrul for mixed‐effects variable rule analysis. *Language and linguistics compass*, *3*(1), 359-383.

Kaczmarek, L. 1953. *Kształtowanie się mowy dziecka*. Nakładem Poznańskiego Towarzystwa Przyjaciół Nauk.

Kaczmarek, L. 1966. Nasze dziecko uczy się mowy. *Wydawnictwo Lubelskie, Lublin*.

Karaś, H. 2009a. *Nowe dialekty mieszane*. Retreived on August 18th, 2015 from http://www.gwarypolskie.uw.edu.pl/index.php?option=com_content&task=view&id=500&Itemid=111

Karaś, H. 2009b. Polszczyzna okolic Jeziorosów jako wariant terytorialny jezyka polskiego na Litwie. In: Polszczyzna za granica jako jezyk mniejszosci i_jezyki mniejszosciowe w_Polsce, red. E. Dziegiel, A. Zielinska, Warszawa, s. 25–46.

Keefe, S., Padilla, A. 1987. *Chicano Ethnicity*. Albuquerque, NM: University of New Mexico Press.

Labov, W. 1972. *Sociolingustic Patterns*. Philadelphia: University of Pennsylvania.

Labov, W. 1987. 12. The Overestimation of Functionalism. *Functionalism in linguistics*, *20*, 311.

Laskowski, R. 1993. The endangered language. The acquisition of the case system by Polish children in Sweden. W: G. Hentschel & R. Laskowski (red.). *Studies in Polish morphology and syntax*, 121-162.

Laskowski, R. 2009. *Język w zagrożeniu: przyswajanie języka polskiego w warunkach polsko-szwedzkiego bilingwizmu*. Towarzystwo Autorów i Wydawców Prac Naukowych Universitas.

Leisio, L. 2006. Genitive Subjects and Objects in the Speech of Finland Russians. Journal of Slavic linguistics: JSL., 14(2), 289-316.

Lyskawa, P., Maddeaux, R., Melara, E., Michaud, M. 2015. Word-final devoicing in Heritage Polish. Course paper for LIN1256, University of Toronto.

Łuczyński, E. 2004. *Kategoria przypadka w ontogenezie języka polskiego: czyli, O wchodzeniu dziecka w rzeczywistość gramatyczną*. Wydawnictwo Uniwersytetu Gdańskiego.

Mańczak, W. 1956. Ile rodzajów jest w polskim, „Język Polski" XXXVI, s. 116-–121.

Marian, V., Spivey, M., & Hirsch, J. 2003. Shared and separate systems in bilingual language processing: Converging evidence from eye tracking and brain imaging. *Brain and language*, *86*(1), 70-82.

Masojć, I. 2001. *Regionalne cechy systemu gramatycznego współczesnej polszczyzny kulturalnej na Wileńszczyźnie*. Dom Wydawniczy Elipsa.

Miodek, J. 2000. *ABC polszczyzny*. Wydawnictwo Dolnośląskie.

Montrul, S. 2008. Incomplete Acquisition in Bilingualism. Re-examining the Age Factor. Amsterdam: John Benjamins.

Montrul, S. 2009. Knowledge of tense-aspect and mood in Spanish heritage speakers. *International Journal of Bilingualism*, *13*(2), 239-269.

Nagy, N. 2009. Heritage Language Variation and Change. http://individual.utoronto. ca/ngn/research/heritage_lgs.htm

Nagy, N. 2015. A sociolinguistic view of null subjects and VOT in Toronto heritage languages. *Lingua* 164B: 309-327. Retrieved on August 30th, 2015 from: http://dx.doi.org/10.1016/j.lingua.2014.04.012.

Nagy, N., Kochetov, A. 2013. VOT across the generations: A cross-linguistic study of contact-induced change. In P. Siemund, I. Gogolin, M. Schulz & J. Davydova, (eds.). Multilingualism and language contact in urban areas: Acquisition - Development - Teaching - Communication. 19-38. John Benjamins.

Parra, M. L. 2015. *Tracking progress: A pre-post study of Spanish heritage narratives.* Presented on June 1st, 2015 at the Eighth Heritage Language Research Institute: New edges and broader domains. Boston, MA.

Penfield, W., Roberts, L. 1959. *Speech and Brain Mechanisms.* Princeton: Princeton University Press.

Pesetsky, D. 2013. *Russian case morphology and the syntactic categories.* MIT Press.

Pires, A. 2011. Linguistic competence, poverty of the stimulus and the scope of native language acquisition. In: Flores, C. (Ed.), Multiple Perspectives on Bilingualism. Humus, Braga, Portugal, pp. 115-143.

Polinsky, M. 1995. Cross-linguistic parallels in first language loss. Southwest Journal of Linguistics 14, 88-123.

Polinsky, M. 2008. Heritage Language Narratives. In: Brinton, D. and Kagan, O. (Eds.), Heritage Languages: A New Field Emerging. Erlbaum

Polinsky, M., Kagan, O. 2007. Heritage languages: In the 'wild' and in the classroom. *Language and Linguistics Compass*, *1*(5), 368-395.

Preston, D. R. 1986. The case of American Polish. *Linguistics across Historical and Geographical Boundaries. In Honour of Jacek Fisiak on the Occasion of his Fiftieth Birthday*, *2*, 1015-1023.

Putnam, M. T., Sánchez, L. 2013. What's so incomplete about incomplete acquisition?: A prolegomenon to modeling heritage language grammars. *Linguistic Approaches to Bilingualism*, *3*(4), 478-508.

R Development Core Team. 2008. R: *A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Sadowska, I. 2012. *Polish: a comprehensive grammar*. Routledge.

Smoczyńska, M. 1985. *The acquisition of Polish*. Lawrence Erlbaum Associates, Inc.

Smoczyński, P. 1955. *Przyswajanie przez dziecko podstaw systemu językowego* (Vol. 19). Ossolińskich we Wrocławiu.

Statistics Canada2011 http://www12.statcan.gc.ca/nhs-enm/2011/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CMA&Code1=535&Data=Count&SearchText=toronto&SearchType=Contains&SearchPR=01&A1=All&B1=All&Custom=&TABID=1

Statistics Canada 2014 http://www.cic.gc.ca/english/resources/statistics/facts2014/permanent/10.asp

Thepboriruk, K. 2009. Bangkok Thai tones revisited. *JSEALS*, *3*, 86-105.

Tucker, G.R. 1999. A Global Perspective on Bilingualism and Bilingual Education. Washington, D.C.: Center for Applied Linguistics.

Zarębina, M. 1965. *Kształtowanie się systemu językowego dziecka*. Ossolineum.

Appendix – Ethnic Orientation Questionnaire

A. Ethnic identification:
1. Do you think of yourself as Polish, Canadian or Polish-Canadian?
2. Are most of your friends Polish?
3. Are people in your neighbourhood Polish?
4. Are the people you work with Polish?
5. When you were growing up, were the kids in your school Polish? Were your friends? The kids in your neighbourhood?

B. Language:
1. Do you speak Polish? How well? How often?
If no: Can you understand Polish?
2. Where did you learn Polish? At home? In school?
3. Do you prefer to speak Polish or English?
4. Do you prefer to read and write in Polish or English?
Do you read Polish magazines and newspapers? — Which ones?
5. Do you prefer to listen to the radio or watch TV in Polish or English?

C. Language choice:
1. What language does your family speak when you get together?
2. What language do you speak with your friends?
3. What language do you speak when you're talking about something personal? When you're angry?
4. Did/do you speak to your parents in Polish? Your grandparents?
5. Do you speak to your children/grandchildren in Polish?

D. Cultural heritage:
1. Where were you born?
If in Poland: How old were you when you came here? How long have you lived here?
If in Canada: Have you ever been to Poland? When? For how long?
2. Where did you go to school?

E. Parents:
1. Do your parents think of themselves as Polish, Canadian or Polish-Canadian?
2. Do/did your parents speak Polish? English? And your grandparents?
3. How old were your parents when they came to Canada? Your grandparents?
F. Partner:
1. Is your husband/wife/boyfriend/girlfriend Polish?
2. Does she/he think of her/himself as Polish, Canadian or Polish-Canadian?
3. Does she/he speak Polish? Do you speak Polish to her/him?

F. Polish culture:
1. Should Polish-Canadian kids learn Polish? Polish culture?
2. Would you rather live in an Polish neighbourhood?
3. Should Poles only marry other Poles?
H. Discrimination:
1. Have you ever had a problem getting a job because you're Polish?
2. What about renting an apartment or buying a house?
3. Were you treated differently by your teachers in school?
4. Have you ever been treated badly because you're Polish?
5. Is there a lot of discrimination against Poles?