Univerzita Karlova

Filozofická fakulta

Ústav obecné lingvistiky

BAKALÁŘSKÁ PRÁCE

Obecná lingvistika

Ivan Kartáč

Prenominal possessive genitive in Czech

Prenominální posesivní genitiv v češtině

Acknowledgements

Prohlášení autora

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně a výhradně s využitím uvedených a řádně citovaných pramenů, literatury a dalších odborných zdrojů a že práce nebyla předložena jako splnění studijní povinnosti v rámci jiného studia nebo předložena k obhajobě v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne 5. srpna 2019   Ivan Kartáč

Abstrakt

Tato bakalářská práce zkoumá prenominální posesivní genitiv z hlediska funkční a typologicky orientované lingivistiky. Vlivy životnosti, váhy fráze, rekurze a alomorfie jsou zkoumány pomocí statistických metod aplikovaných na materiál z korpusů mluvené češtiny. Výsledky ukazují, že životnost je nejvýznamnějsí faktor ovlivňující užití prenominálního genitivu, a tato vlastnost je dále probírána v mezijazykovém kontextu. Data dále naznačují, že prenominální pozice může mít jisté výhody při užívání rekurzivních posesivních frází. Zároveň byly pozorovány vlivy morfofonologických alternací.

Klíčová slova: posesivita – atributivní konstrukce – prenominální genitiv – hierarchie životnosti – rekurze – korpusová analýza – logistická regrese

Abstract

This bachelor thesis explores prenominal possessive genitive from the perspective of functional and typologically oriented linguistics. Using statistical methods applied to material from corpora of spontaneous spoken Czech, effects of animacy, phrase weight, recursion and allomorphy on genitive position are examined. The results suggest that animacy is the most important factor predicting the use of prenominal genitive and this feature is further discussed in cross-linguistic context. The data also suggest that prenominal position may have certain advantages when using recursive possessive phrases. Effects of morphophonological alternations have been observed as well.

Key words: possession – attributive construction – prenominal genitive – animacy hierarchy – recursion – corpus analysis – logistic regression

Contents

Abbreviations

ACC – accusative

DAT – dative

DET – determiner

GEN – genitive

POSS.ADJ – possessive adjective

PRF – perfect

PST – past

REL.ADJ – relational adjective

1   Introduction

This bachelor thesis explores prenominal possessive genitive in Czech, a specific way of expressing attributive possession. Even though it appears to be quite common, at least in a specific semantic domain, this type of construction has not received much attention in literature, often being mentioned only briefly as a marginal phenomenon. The majority of works that comment on prenominal possessive genitive come from the field of dialectology, suggesting that it is relatively common, even dominant, in some regions of Czech republic (Nová 2016), gradually replacing possessive adjectives as a competing form (Uličný 2018). However, there are no empirical studies addressing prenominal possessive genitive in Czech from a functional perspective.

Therefore, this thesis aims to examine which factors related to morphology, syntax, semantics and language processing predict the use of prenominal possessive genitive. These questions are addressed empirically, using material from a corpus of spontaneous spoken Czech. Hypotheses are tested applying statistical methods, in particular binomial logistic regression, which allows to determine effect sizes and possible interactions between the independent variables.

The thesis is structured as follows. Chapter 2 presents the theoretical background and methodology. The concept of possession and various ways of expressing attributive possession in Czech are introduced in section 2.1. In section 2.2, I explain each factor that is explored as possible determinant for prenominal genitive: animacy, weight, recursion and allomorphy.

Chapter 3 is dedicated to the research method. After introducing my research questions in section 3.1, the process of data collection and annotation of the material is presented sections 3.2 and 3.3, where I describe the details of how each variable is operationalized. Finally, I formulate research hypotheses in section 3.4.

The results are presented in chapter 4. First, each variable is explored in isolation in section 4.2, then predictions of a binomial logistic regression model are discussed in section 4.3.

Chapter 5 is focused on discussion and interpretation of the results. In section 5.1, I discuss animacy and weight in cross-linguistic context and compare various

ways of expressing attributive possession in Czech to other Slavic languages. Next, I elaborate on the results for recursion and suggest some hypotheses regarding my observations. The last section of the chapter comments on the results related to allomorphy.

2 Theoretical background and methodology

2.1 Possession

2.1.1 Possession as a typological concept

In the context if linguistic typology, three main types of possession are distinguished: attributive (adnominal) (1a), predicative (1b) and external (1c). Possession is a concept covering a range of relations, the most central being ownership, whole-part and kinship relations, which are often expressed by the same type of adnominal construction in many languages (Dixon 2010).

(1) a. *jeho   kolo*

          his   bicycle

          'his bicycle'

     b. *on   má   kolo*

          he   has   bicycle

          'he has a bicycle'

     c. *ukradli   mu   kolo*

          steal.PST.PRF.PL   he-DAT   bicycle

          'they stole his bicycle'

This attributive construction may also express additional relations such as associations (2a), attributes (2b) or spatial relations (2c), which are thus often subsumed under the notion of possession as well. There are different ways to mark adnominal possession in the world's languages, with some of them simply apposing the elements, while others mark the relation morphologically on possessor, on possessum or on both.

(2) a. *její   zubař*

          her   dentist

          'her dentist'

     b. *barva   očí*

          color   eyes-GEN

          'eye color'

11

c.	*vrchol hory*

peak	mountain-GEN

'(a/the) peak of the mountain'

Many languages mark the distinction between alienable and inalienable possession, where a class of inalienable nouns (typically body parts and kinship terms) is marked differently than other nouns. In Yidiñ, an Australian language, relations between possessor and his body parts, illnesses, or his 'name' are marked by apposition (3a), while for other classes, dependent noun is marked by genitive suffix (Dixon 2010). In Czech, as in most other European languages, alienability is not grammaticalized.

(3)	a.	*wagu:ja*	*jina*

man	dog

'man's dog'(Dixon 2010: 284)

b.	*waguja-ni*	*guda:ga*

man-GEN	dog

'man's dog' (Dixon 2010: 284)

In this thesis, I work with a broader definition of possession and the exact criteria for including a particular example in the dataset are presented in 3.2.

2.1.2 Attributive possession in Czech

In Czech, possession is dependent-marked in all types of attributive constructions. There are three main ways of expressing attributive possession in Czech: possessive pronouns, possessive adjectives and possessive genitive. Possessive pronouns appear prenominally, agreeing with the head of a phrase (4a). Possessive adjectives are formed from declinable masculine and feminine nouns, mostly restricted to human referents (4b). Possessive adjectives cannot be modified, appear only prenominally and agree with the head noun. The exception is a special type of invariant possessive adjectives, where there is no agreement with the head noun (4c). Genitive can be used for all classes of nouns, allows modification and may appear in both positions, although postnominal position is

12

more common (4d). An example of a relational adjective, which is an additional type of construction used to express possession in a broader sense, is shown in (4e).

(4)   a.   *její   hlas*

               her   voice

               'her voice'

       b.   *záhradníkův   rok*

               gardener-POSS.ADJ   year

               '(a/the) gardener's year'

       c.   *bráchovo   žena*

               brother-POSS.ADJ   wife

               '(the) brother's wife'

       d.   *život   hmyzu*

               life   insect-GEN

               '(the) insects' life'

       e.   *psí   bouda*

               dog-REL.ADJ   shed

               'dogshed'

Prenominal possessive genitive, a special type of genitive construction, did not receive much attention in the literature and is mostly commented on in the context of dialectology. Voráč (1950) localized this variant to southeast Bohemia, where he considered it to be preferred to possessive adjectives. Using data of a dialectological field study from the 1960s and 1970s, Nová (2016) concludes that possessive genitive was quite common across the majority of regions, while in many of them prenominal variant was found to be dominant, even for unmodified phrases. Uličný (2018) states that in both spoken and written Czech, possessive adjectives are being replaced by genitive, including the prenominal variant, which is very common in the Bohemian region.

Outside dialectology, prenominal possessive genitive is usually mentioned only briefly and there are no empirical studies exploring possible factors that could predict the use of this variant. Koptjevskaja-Tamm (1993) notes that languages

with prenominal possessive forms such as Czech possessive adjectives often have genitive with two positional variants, while prenominal position "may be occupied by a very limited class of words". (Koptjevskaja-Tamm 1993: 201). Křivan (2015: 155) comments on an observed tendency to use prenominal genitive constructions with highly animate and accessible referents, approaching pragmatic properties of structurally similar possessive adjectives.

2.2 Predictors

2.2.1 Animacy

Animacy is a semantic concept that manifests itself in language structures in several different ways. There are many morphological splits controlled by this feature, among them case marking, verb agreement or split-ergativity (Comrie 1989). In Czech as well as in many other Slavic languages, case marking on direct object depends on animacy, which results in different declension paradigms for masculine animate nouns. Besides morphology, it has also been shown to be relevant for syntax, namely for word order. Studies on genitive variation in English indicate that animacy of possessor in interaction with other factors is a strong predictor in choice between the two types of possessive construction: flexive *s-genitive* and an analytical *of-genitive*[1] (Rosenbach 2003; Hinrichs & Szmrecsanyi 2007). Specifically, animate and accessible possessors tend to be expressed primarily by s-genitive, while of-genitive is preferred with inanimate and less accessible possessors. This distinction is not categorical, so in most contexts, (5a) would be considered much more likely than (5b), even though (5b) is still grammatical. As my informal observation and preliminary data from corpora of spoken Czech suggest, animacy may also be an important factor affecting the use of prenominal possessive genitive in Czech.

(5)   a. *Peter's car*

        b. *the car of Peter*

---

1   As an alternative name for s-genitive, 'Saxon genitive' is sometimes used, while of-genitive is also called 'Norman genitive' or 'periphrastic genitive'.

The notion of animacy hierarchy, first described by Silverstein (1976) in his paper on split ergativity, represents the idea that animacy as a semantic concept can have an impact on formal characteristics of linguistic units. Example (6) shows a basic animacy hierarchy, which is formally reflected in grammars of many world's languages.

(6)   Animacy hierarchy (Comrie 1989: 185)
        human > animate > inanimate

While this basic hierarchy is sufficient to describe effects of animacy in some languages, there are others that often make finer distinctions, such as the one between first and second versus third person pronouns, between pronouns versus other nouns, or between humans and non-humans (Comrie 1989). In Spanish possessive phrases, prenominal position is reserved only for pronouns (7a), whereas other nouns may appear only postnominally in a prepositional phrase (7b).

(7)   a. *su casa* 'his/her house'
        b. *la casa de Miguel* 'Miguel's house'
        c. **Miguel casa* 'Miguel's house'

Czech possessive pronouns have a categorical restriction for position and appear only prenominally as pronominal possessive adjectives (8a), while other human animate nouns (and in specific cases also non-human animates) may appear both prenominally as adjectives (8c), or postnominally as genitives (8d).

(8)   a. *můj dům* 'my house'
        b. **dům mě* 'my house'
        c. kamarádův dům 'a friend's house'
        d. dům kamaráda '(a/the) house of a friend'

Working with the hierarchy in (1), proper nouns are not more animate than kinship terms, as they are both human and thus would be placed on the same level. However, this may not capture important distinctions that are present in grammars and usage preferences of many languages, including Czech. Therefore, it will be extended to a more fine-grained hierarchy (9), allowing a distinction between proper names, kinship terms and other human nouns, as well as between potentially more accessible referents and those that are typically relatively distant. This extended hierarchy should not be understood in a strict literal sense, but as a combination of two different concepts: animacy and accessibility (Ariel 2001).

(9)   Extended animacy/accessibility hierarchy for nouns

proper nouns > kinship terms > close human > distant human > organization > non-human animate > inanimate concrete > inanimate abstract

The extended hierarchy and the coding process will be explained in detail in section 3.2.1.

2.2.2   Weight

The principle of end-weight, first stated by Behaghel (1909) is a notion that heavier or more complex phrases have a tendency to occur after shorter ones. It has been empirically demonstrated for English several times (e.g. Hinrichs & Szmrecsanyi 2007) and is hypothesized to contribute to an increased efficiency of parsing (Hawkins 2004). Constraints of weight are also present in East Slavic and South Slavic languages (O'Connor et al. 2013). In German, prosodically short genitive phrases in postnominal position are dispreferred (10) (Campe 2013).

(10)   a.   *der   Computer   Ulf-s*

                the   computer   Ulf-GEN

                'Ulf's computer' (Campe 2013: 276)

        b.   ?*der   Computer   Peter-s*

                the   computer   Peter-GEN

                'Peter's computer' (Campe 2013: 276)

c.   *der   Computer   Alexander's*

       the   computer   Alexander-GEN

       'Alexander's computer' (Campe 2013: 276)

Comparing Czech possessive adjectives with possessive genitives, there is a categorical constraint of weight, where adjective construction is used only in prenominal position (11a) and cannot be modified by adjectives or nested possessors (11c) (O'Connor et al. 2013). Only genitive form can be used when the possessor is modified, in which case postnominal position preferred (11d).

(11)   a.   *Jirkův   bratranec*

             Jirka-POSS.ADJ   cousin

             'Jirka's cousin'

    b.   *??bratranec   Jirkův*

             cousin   Jirka-POSS.ADJ

             '(the) cousin of Jirka'

    c.   **našeho   Jirkův   bratranec*

             our-GEN   Jirka-POSS.ADJ   cousin

             'our Jirka's cousin'

    d.   *bratranec   našeho   Jirky*

             cousin   our-GEN   Jirka-GEN

             '(the) cousin of our Jirka'

Moreover, Czech possessive adjective also forbids full names (12a), or even double names (12b), which suggest that it is indeed the weight of possessor phrase that determine this constraint. The only exception may be coordinated proper names, where there are two distinct possessive phrases (12c).

(12)   a.   **Karl-ova   Čapk-ova   hra*

             Karel-POSS.ADJ   Čapek-POSS.ADJ   play

             'Karel Čapek's play'

      b.   *Janův   Václavův   žák*

b.   *Janův*   *Václavův*   *žák*

Jan-POSS.ADJ   Václav-POSS.ADJ   student

'Jan Václav's student'

c.   *Karlův*   *a*   *Josefův*   *dům*

Karel-POSS.ADJ   and   Josef-POSS.ADJ   house

'Karel and Josef's house'

Prenominal genitive is similar to possessive adjectives, while allowing more freedom regarding length of possessor phrase, for which it has no categorical restriction (13). Therefore, we also examine possible effects of both possessor and possessum weight on the choice of prenominal genitive.

(13) a. *našeho*   *bratrancova*   *dcera*

our   cousin-POSS.ADJ   daughter

'our cousin's daughter'

      b. *našeho*   *bratrance*   *dcera*

our   cousin-GEN   daughter

'our cousin's daughter'

2.2.3 Recursion

In language, recursion is the ability of constructions to contain other constructions of the same type. In the generative tradition, it has been considered a fundamental property of grammar, while Hauser et al. (2002) claim that it is the only aspect of the hypothetical language faculty, originally postulated by Chomsky (1965), that is unique to humans. However, there is an evidence of languages that do not employ recursion at all (Evans & Levinson 2009), with Pirahã, an indigenous language spoken in Amazonia, being the most famous example (Everett 2005). Christiansen & MacDonald (2009) proposed a usage-based perspective on processing of recursive structures. They suggest, contrary to generative views on language processing, that more general and evolutionary older cognitive mechanisms might be responsible for processing of recursion in language. In English, complex recursive structures with nested center-embedded dependencies

have been shown to be difficult to process if not supported by an appropriate intonation (Christiansen 2015). In two English sentences shown in (9), dependencies are embedded within each other, leading to one level of recursion in (14a) and to two levels of recursion in (14b).

(14)   a. The dog that John saw chased the cat.

       b. The cat that the dog that John saw chased bit the mouse.

      (Christiansen & MacDonald 2009: 128)

Example (15) shows a center-embedded sentence in Dutch. Here, the first noun has to be linked to the last verb, the second noun to the penultimate verb, and so forth. Examining corpora of several European languages, Karlsson (2007) suggests that sentences with multiple center-embeddings such as this one are absent from spoken language.

(15)   a.   *(dat)*   *Jan*   *Piet*   *Marie zag*   *laten*   *zwemmen*

          (that)   Jan   Piet   Marie saw   make   swim

         '(that) John saw Piet make Marie swim' (De Vries et al. 2011: 12)

Nested prenominal possessor is an example of a simpler, noun phrase level recursive structure, where elements are ordered linearly without nested dependencies. Possessive noun phrases containing nested genitives could be illustrated by (16).

(16)   *jeho*   *bráchy*   *přítelkyně*   *bratranec*

      his   brother-GEN   girlfriend-GEN   cousin

      'his brother's girlfriend's cousin'

In section 5.2, I suggest some hypotheses regarding the use of this kind of construction.

2.2.4 Allomorphy

Possessive adjectives are the most common option for unmodified, animate and highly accessible possessors in Czech. However, their declension is rather complex, with different forms depending on case of the head noun.

In addition, a number of stems exhibit phonological allomorphy with different types of morphophonological alternations. Example (17a) shows zero alternation in the last syllable of the stem, where the vowel is deleted if a vowel-initial suffix follows. In (17b), there is an alternation between *k* and *č* which occurs in all forms of the possessive adjective derived from this noun. While the type shown in (17a) applies to both genitive and possessive adjectives, alternations exemplified by (17b) occur only in the latter.

(17) a. *otec > otc-ův*
      b. *matka > matč-in*

In the west Bohemian border region, invariable possessive adjective (*bratrovo*) is a dominant type of adnominal construction used with human possessors (Nová 2016). This variant is formally very similar to genitive, as it is also used in modified phrases and its declension is completely reduced to a single form. However, the feminine variant is quite rare. Nová attributes this to stem alternation, which might be the reason why it is being replaced by genitive.

Therefore, following Hawkins (2004: 35) and his Minimize Forms principle (18), I assume that using prenominal genitive may be a more economical alternative to possessive adjectives.

(18) Minimize Forms (MiF)

      The human processor prefers to minimize the formal complexity of each linguistic form F (its phoneme, morpheme, word, or phrasal units) and the number of forms with unique conventionalized property assignments, thereby assigning more properties to fewer forms. These minimizations apply in proportion to the ease with which a given property P can be assigned in processing to a given F.

Hawkins summarized MiF, together with another principle (Minimize Domains), by the slogan 'Express the most with the least'. It is a principle of economy, which states that the effort in processing of linguistic forms can be minimized by reducing them in terms of their units, such as phonemes, morphemes, words or phrases. Zipf (1949) observed that more frequent words tend to be phonologically and morphologically more reduced, the most frequent words often having very short forms.

It is also less economical to have a distinct formal unit for each property. In expressing number, most languages do not mark singular explicitly, neither mark each unique value with different form. Even though there are languages with complex grammatical distinction for number, most express only the distinction between one and more entities (Corbett 2000).

In addition to a lack of complex declension when compared to adjectives, prenominal genitive has the advantage of avoiding forms with morphological alternations that require greater articulatory effort, therefore being potentially used as a more economical variant.

3   Method and analysis

3.1   Research questions

This thesis aims to identify possible determinants in choice of prenominal possessive genitive in Czech. Based on informal observations of corpus data and existing studies in both Czech and other languages, I identified some factors that may have an effect on the use of this variant.

First, I examine if there is a relationship between prenominal position of genitive and animacy of both possessor and possessum. Using the observations from both Czech and other European languages as a basis, I expect that this position will favor highly animate and accessible possessors (2.2.1). However, previous observations of corpus data indicate that animacy of possessum might also be relevant.

Regarding phrase weight, I assume that prenominal genitives will be used with syntactically light phrases more often than with heavy phrases (2.2.2). Moreover, I expect that weight constraints apply in both word orders, thus heads of postnominal genitive phrases will not tend to be significantly longer than heads of prenominal genitive phrases.

Next, recursive possessive phrases and their effect on genitive position are examined (2.2.3). Even though this variable correlates with weight, there are specific cases where it is expected to override possible weight effects. It is expected that speakers will use recursion in a spontaneous conversation mostly when talking about relationships between human referents, typically kinship relationships. In such cases, nested prenominal phrases may have an advantage of allowing to proceed from the most accessible to the least accessible referent. This is further discussed in 5.2.

The last factor I explore is presence of morphophonological alternations in possessive adjective forms. As pointed out in 2.2.4, prenominal genitive could be considered a more economical form and thus is expected to be used more often if it could avoid potential stem alternation.

Hypotheses based on these assumptions are formulated in 3.4.

3.2   Data collection

3.2.1   Corpus

Prenominal genitive is considered a highly colloquial form, and for that reason I decided to focus on corpora of spontaneous spoken language. As the source of the material for the analysis, I use corpus ORALv1 (Kopřivová et al. 2017) created by the Institute of Czech National Corpus. It consists of transcriptions of recordings representing spontaneous colloquial Czech and is based on 582 hours of recordings from years 2002–2011, containing 1,546 recorded conversations with 1,297 unique speakers. The corpus is lemmatized and morphologically tagged.

3.2.2   Material

Working with broadly defined possession following Dixon (2010), we consider the following relations as possessive: ownership[2], part-whole relations, kinship relations, attributes, associations and spatial relations/locations. Nominalizations are not included, even in cases where they could be reanalyzed as possessive relations. I also excluded collocations and idioms, temporal relations and non-possessive uses of genitive, such as quantity, collectivity, material or product. All examples of genitives and possessive adjectives satisfying these constraints were extracted using KonText interface of Czech National Corpus.

When dealing with corpora of spoken language, a trade-off between precision and recall[3] will often be an issue. This could be usually translated to a trade-off between amount of relevant data obtained and time dedicated to manual inspection of the results. Even though I assume that prenominal genitive is not as rare as it is often perceived, the number of occurrences in ORALv1 is still expected to be relatively low. While working with results obtained using a query with high precision requires much less manual effort, the recall is low and much valuable data may be lost. The query in (19a) is more restricted and returns only 5,213 hits. On the other hand, using a query with high recall would result in a large amount of results and low precision, and manual data inspection would

---

2   Temporary ownership is also included as a kind of possessive relation.

3   In the context of information retrieval, these measures were first defined in Perry, Kent & Berry (1955). Precision is the percentage of retrieved results that are relevant, i.e. the number of relevant results returned by the query divided by the number of all returned results. When all retrieved results are relevant, precision equals 1. Recall is the percentage of all successfully retrieved relevant results, i.e. number of relevant results returned by the query divided by the number of all relevant results. When all relevant results are retrieved, recall equals 1.

become infeasible. As an example, very simple query returning all nouns in genitive case could be used (19b). This has the advantage of returning many relevant examples where possessum is erroneously tagged or where there is a larger amount of tokens between the two nouns of interest, which would be lost using the more restricted query in (19a). Due to the nature of spoken language and its transcription, a number of tokens, such as false starts, pauses or filler words, may occur between both referents in a possessive construction, and these examples cannot be automatically filtered out from others, in which there is no possessive relationship at all. However, (19b) returns 94,208 hits, which is not feasible to examine manually.

(19) a. `[tag="N...2.*"][tag="N.*"]`

      b. `[tag="N...2.*"]`

As a trade-off between the two options, queries in (20) were created to extract all occurrences of prenominal genitives (20a), postnominal genitives (20b) and possessive adjectives (20c). They allow up to three tokens to occur between possessor and possessum nouns, tagged as an adjective, adverb, pronoun, numeral, punctuation or word fragment. Prepositions *od* and *vod* are added to (20b) retrieve prepositional genitive phrases for postnominal genitive as well.

(20) a. `[tag="N...2.*"][tag="[APCDZF].*"]{0,3}[tag="N.*"]`

      b. `[tag="N.*"][tag="[APCDZF].*"|word="vod"|word="od"]`
          `{0,3}[tag="N...2.*"]`

      c. `[tag="AU.*"][tag="[APCDZF].*"]{0,3}[tag="N.*"]`

3.3   Annotation

Based on the assumptions presented in 3.2, the extracted results were annotated for the following characteristics: animacy of possessor, animacy of possessum, length of possessor, length of possessum and recursion depth of the phrase. Coding schemes and explanations for each variable are presented in the following sections.

3.3.1 Animacy

In this section, I explain the extended animacy hierarchy (21) outlined in 2.2.1. and describe the coding scheme for the variable.

(21)   Extended animacy/accessibility hierarchy for nouns

proper nouns > kinship terms > close human > distant human > organization > non-human animate > inanimate concrete > inanimate abstract

The category of proper nouns includes first and last names, while full names are treated separately and are included in the *distant human* category discussed below. In addition to blood relationships and affinal relationship, the category of kinship terms includes terms for partners (relationships not involving marriage) as well, e.g. *přítelkyně* 'girlfriend', *snoubenec* 'fiancé' or *holka* 'girlfriend'[4]. For other nouns referring to humans, I decided to distinguish expressions typically used for people that are closer to the speaker or hearer from those expressions usually reserved for more distant referents. The reason is that there may be a significant difference in accessibility between these two categories and consequently a different preferences in language use. Therefore, I make a distinction between nouns for close human referents, i.e. people with whom speaker or hearer is assumed to be in a close relationship (*kamarád* 'friend') or at least relatively regular contact (*spolužák* 'classmate') and expressions used for more distant referents (*prezident* 'president', *zaměstnanec* 'employee').

A special class is reserved for collectives or human organizations, which is an interesting category on the boundary between human and inanimate, nevertheless with a strong tendency for its referents to be used similarly to human referents. This includes all cases where a noun is used with a collective meaning (cf. Zaenen 2004).

The category of non-human referents includes both domestic (*pes* 'dog') and wild animals (*straka* 'magpie'), as well as human-like entities (*čert* 'devil'). Inanimates

---

4   This is one several different polysemous nouns present in our dataset. In cases such as this one, where there were several possible meanings (among them 'girl', 'girlfriend' or 'daughter') spanning more than one category on the hierarchy, the correct level was assigned depending on the context.

are usually not differentiated in languages[5], and thus the inanimate macro-category is divided into only two subcategories. The first subcategory represents concrete tangible entities, including body-parts (*ucho* 'ear'), material objects (*hrnek* 'mug') as well as places and cities. The second represents a broad category of abstract entities, such as events (*oslava* 'celebration'), natural phenomena (*bouřka* 'thunderstorm') or attributes (*barva* 'color').

All examples were annotated for animacy of both possessor and possessum, which were assigned numeric values from 1 to 8. The coding scheme for animacy is shown in Table 1.

value	category	macro-category
8	Proper noun	Human
7	Kinship term	
6	Close	
5	Distant	
4	Organization	Animate
3	Non-human	
2	Concrete	Inanimate
1	Abstract	

Table 1. Coding scheme for animacy hierarchy

Pronouns, representing the most animate and accessible type of referent, are not included in this hierarchy, as they are possible only in adjective forms. Consequently, examples of phrases with pronominal possessors were not included in the dataset.

3.3.2 Weight

Weight of both possessor and possessum will be included as another independent variable, allowing us to examine if its effects also apply in the variation between two genitives, or if interaction of weight and other factors affects the variation. As I will discuss in section 5.3, in the specific case of nested genitive, cognitive processing factors may outweigh the effect of constituent length.

---

5 Navaho is one of the few exceptions, where inanimates with an ability of spontaneous motion are treated as higher on the hierarchy than other inanimates (Comrie 1989).

There is a number of possible ways how to operationalize complexity of sentences and phrases. One option, popular in experimental research and considered psychologically most real, is to count phrasal nodes dominated by a unit. As an alternative, simpler measures can be used, such as number of stress units or number of words. Wasow (1997) suggests that word counts are a good approximation for weight in English, and Szmrecsanyi (2004) has shown statistically that using simple word counts as a proxy for syntactic complexity works almost as well as using node counts. One of the major advantages of this approach for corpus-based studies is that it does not require manual coding, which may save a lot of time if there would be a large amount of data to annotate, especially when the corpus is not syntactically tagged. Consequently, this measure has been used in many corpus-based studies (e.g. Arnold et al. 2000; Hinrichs & Szmrecsanyi 2007).

Even though I assume nested genitives to be more complex for processing compared to modified noun phrases with the same word count, using node counts as a proxy for noun phrase weight of simpler constituents should not be necessary in this case and I also use an approach based on simple word counts. However, it is slightly modified, so that prepositions are not counted as separate words. The reason is that these are grammatical markers which usually do not form a stress group by themselves, but instead attach to a stress group of another word.

Both possessor and possessum phrase lengths were coded as discrete numeric variables, starting from 1.

3.3.3 Recursion

Recursion was operationalized as a number of additional possessors in a nested possessive phrase and coded as a discrete numeric variable. Simple genitive phrases were assigned value of zero (22a) and each additional level of recursion increased the value by one (22b, 22c).

(22)   a. *bratrance syn* 'a counsin's son' (recursion = 0)

        b. *táty bratrance syn* 'dad's cousin's son' (recursion = 1)

        c. *našeho táty bratrance syn* 'our dad's cousin's son' (recursion = 2)

Mixed phrases are also included, i.e. nested possessive phrases in which the initial possessor may be expressed by possessive adjective or possessive pronoun (22c). There is some evidence that speakers of English tend to avoid identical genitive phrases within a noun phrase, which is typically resolved by a combination of s-genitive and of-genitive (Hinrichs & Szmrecsanyi 2007). An equivalent strategy for nested possessors in Czech would be to combine an adjective with one or more genitives in the phrase.

3.3.4 Allomorphy

Presence of potential allomorphy in adjective form was coded as a binary variable, with value 1 for an alternation and 0 for no alternation. Alternations that occur in both adjective and genitive forms are not considered, and only those where there is a difference between stems of adjective and genitive forms are taken into account.

3.4 Hypotheses

Following the assumptions introduced in 3.1, I formulate hypotheses which will be tested quantitatively, using the corpus data collected and processed as described in 3.2 and 3.3.

First, the effect of animacy on the choice of prenominal genitive will be explored. There are two hypotheses related to animacy:

(i)   Possessors of prenominal genitives will tend to be higher on the animacy hierarchy than possessors of postnominal genitives

(ii)   Possessa of prenominal genitives will be tend to be higher on the animacy hierarchy than possessa of postnominal genitives

Second, using the metric introduced in 3.3.2, the effect of both possessum and possessor weight will be tested:

(iii)   Possessor phrases of prenominal genitives will tend to be shorter than possessor phrases of postnominal genitives

28

(iv)   Possessum phrases heading postnominal genitives will not tend to
be longer than possessum phrases heading prenominal genitives

Next, I will test the following hypothesis regarding recursion:

(v)   The relative frequency of nested possessor phrases will tend to be
higher for prenominal genitive than for postnominal genitive if the
referents are high on the animacy hierarchy

The last hypothesis concerns the effect of morphological alternations on the
choice between prenominal genitive and possessive adjectives:

(vi)   The relative frequency of occurrences with potential stem
alternation will be higher for prenominal genitives than for
possessive adjectives

4   Results

4.1   Frequencies

The final dataset consists of 2107 observations, with 1627 examples of postnominal genitive and 103 examples of prenominal genitive (Table 2). Possessive adjectives are represented by 334 examples and a separate category of invariant possessive adjectives includes 43 examples. Except where noted, prepositional genitive constructions (using preposition *od/vod*) are not included when comparing prenominal with postnominal genitives, as well as in the analysis of allomorphy effects, where possessive adjectives are compared with prenominal genitives. The final dataset consists of 1954 examples.

category	total	prepositional	final
adjectives	334	-	334
invariant	43	-	43
prenominal	103	37	66
postnominal	1627	116	1511
total	2107	153	1954

Table 2. Frequencies of different types of constructions in the dataset

Note that after excluding all prepositional genitives, only two thirds of original prenominal genitive examples remained in the dataset. The distribution of the two classes is highly unbalanced, which suggests that prenominal variant is indeed relatively rare, although its sparseness should most probably be attributed to specific restrictions that will be discussed in following sections. However, when compared to possessive adjectives, it does not appear so uncommon, representing 13.8% of all prenominal possessive constructions in the sample.

4.2   Univariate analysis

Before applying binomial logistic regression to the data, I will first discuss the effect of each predictor in isolation. All statistical tests were performed in R programming language (R Core Team 2013).

4.2.1 Animacy

Mean scores and standard deviations for animacy of both possessor and possessum noun phrases are shown in Table 3. These results suggest a strong tendency for prenominal genitive to encode highly animate possessors. Moreover, its standard deviation (1.35) suggests greater specialization for specific semantic categories, while the value for postnominal position (2.05) implies more variation. To compare the mean scores, an independent sample Wilcoxon test was performed, showing a significant difference in animacy scores between the two categories ($W = 89556$, $p < 0.001$).

	possessor		possessum	
	mean	std	mean	std
prenominal	6.86	1.45	6.29	1.62
postnominal	3.36	2.03	2.67	2.05

Table 3. Means and standard deviations for animacy scores of possessor and possessum

The results for possessum phrase are similar, again with a significant difference of mean animacy scores between the two groups ($W = 89417$, $p < 0.001$). In addition to a tendency to occur lower on the animacy hierarchy, standard deviation for prenominal genitive is now slightly larger (1.69). Even though possessum appears to be less restricted in the range of preferred semantic categories, there is still a strong preference for highly animate referents in prenominal genitives.

4.2.2 Weight

As Table 4 indicates, lengths of both possessor phrases are very similar on average in both genitive variants. An independent sample Wilcoxon test shows no significant difference between the two groups, thus the null hypothesis cannot be rejected ($W = 46677$, $p = 0.32$). However, a significant difference between the two groups is observed for mean length of possessum phrases ($W = 43852$, $p < 0.05$).

	possessor		possessum	
	mean	std	mean	std
prenominal	1.64	0.89	1.14	0.35
postnominal	1.66	0.82	1.30	0.56

Table 4. Means and standard deviations for weight of possessor and possessum phrases

Looking at the data more closely, 39% of all prenominal genitive phrases are longer than one word (Table 5). Even though the number of examples is relatively low, this suggests that unlike adjectives, prenominal genitives are not restricted to one-word phrases.

	1 word	2 words	3 words	4 words	> 4 words
prenominal	40	12	12	2	0
postnominal	704	671	107	17	12

Table 5. Frequencies for possessor phrase length for prenominal and postnominal genitive

4.2.3 Recursion

As was only one example with recursion level higher than one in the sample, the results in Table 6 were converted from a scale to a binary distinction denoting presence of recursion.

	postnominal		prenominal	
	count	%	count	%
yes	67	4.4	10	15.2
no	1444	95.6	56	84.8

Table 6. Counts and proportions of presence or absence of recursion in possessor phrases

	postnominal		prenominal	
	count	%	count	%
no	350	87.1	55	84.6
yes	52	12.9	10	15.4

Table 7. Counts and proportions of presence or absence of recursion in possessor phrases when animacy level is larger 4

First, the proportions indicate that nested possessive phrases are overall not very common: only 4.9% of all phrases are nested. These results may also imply that recursion is relatively more common in prenominal genitive phrases, compared to postnominal phrases. However, the number of examples with nested possessive phrases where possessor is non-human (lower than *distant* in the hierarchy) is very low, only 0.8% of all examples. As pointed out in section 3.1, recursion in possessive phrases is expected to occur mainly when referring to relationships

between relatively accessible human referents. Table 7 shows the proportions when the sample is limited only to examples with nouns referring to humans (*distant* and higher on the hierarchy).

Therefore, relatively high proportion of nested phrases in prenominal position is better explained by a tendency of prenominal genitive to be used with highly animate referents (see 4.2.1). Pearson's chi-squared goodness-of-fit test indicates no significant difference when the two groups are compared in the relevant domain only, with $\chi^2(1) = 0.118$, p = 0.73. Therefore, the null hypothesis of no difference between the frequency of recursion in the two variants cannot be rejected. Nevertheless, the result is still interesting, since it shows that potential recursion does not prevent speakers to use prenominal genitive. I will further discuss these results in section 5.3.

4.2.4 Allomorphy

To examine possible effects of allomorphic alternation, only interchangeable examples of adjectives and prenominal genitives were used (see Table 2), i.e. all examples with possessors expressed by prepositional or modified genitive noun phrases were excluded. Cross-tabulating the two types of construction with potential morphophonological alternation results in Table 8.

	adjective		genitive	
	count	%	count	%
no	265	79.8	14	37.8
yes	67	20.2	23	62.2

Table 8. Counts and proportions of examples with possible alternation for adjective form

It is evident that the proportion of prenominal genitives is much greater in cases where there is an alternation in adjective forms. They are used in 25.6% of all potentially alternating cases, while only in 5% cases of non-alternating lexemes. Pearson's chi-squared goodness-of-fit test shows this difference to be highly significant, with $\chi^2(1) = 29.58$, p < 0.001.

4.3   Logistic regression

To examine effect sizes of independent variables, binomial logistic regression will be used. Logistic regression is a method that allows quantifying relative

33

importance of different explanatory variables, as well as interactions between them. The method finds coefficients *w* to compute odds or probabilities of different outcomes (23).

(23)   Logistic regression equation

$$g(x) = b + w_1x_1 + w_2x_2 + ... + w_nx_n$$

In (23), *n* is the number of explanatory variables, $x_1, x_2, ..., x_n$ are the values of these variables and $w_1, w_2, ..., w_n$ are the regression coefficients. The intercept *b*, also called *bias*, is the value where the graph of the function crosses y-axis. It is the value that *g(x)* takes in case *w* is equal to zero[6], or when it is at its reference value in case of categorical variables.

Table 9 reports various statistics related to predictive ability of the model as a whole. Model χ^2 is an indication of an overall significance of the model. The value of the index of concordance C is 0.93 which means 'outstanding discrimination'[7]. I also report Nagelkerke R^2, which represents the explanatory power of the model.[8] To estimate potential multicollinearity between the predictors, VIF-scores (Variance Inflation Factors) were used. Multicollinearity can be observed when two or more independent variables of the model are linearly associated, so there is a high positive or negative correlation between them. The obtained VIF-scores are relatively low and do not indicate multicollinearity between predictors in the model. In addition, validation with bootstrapping was performed to check that the model is not overfitting.

---

6   If there are multiple explanatory variables, *w* is a vector, thus each value has to be equal to zero.

7   This evaluation is based on the following scale, proposed by Hosmer & Lemeshow (2000: 162):

     C = 0.5   no discrimination
     $0.7 \leq C < 0.8$ acceptable discrimination
     $0.8 \leq C < 0.9$ excellent discrimination
     C > 0.9   outstanding discrimination

8   Reporting R^2 for logistic regression is often not recommended, as it is not easy to interpret and its value tends to be much lower than when using linear regression (e.g. Levshina 2015).

	value
Model χ^2	209.77 (df = 5)*
C	0.928
Nagelkerke R^2	0.424

Table 9. Statistics representing predictive ability of the model[9]

Table 10 shows the estimates of coefficients obtained after fitting the model, represented as log odds ratios. Using the formula in (23), log odds ratios can be transformed into odds ratios. Only the effect of animacy is significant, estimating that the odds of using prenominal instead of postnominal genitive is 1.91 higher with each increase on the animacy hierarchy for possessor and 1.59 higher for possessum. Consistent with the results from section 4.2, the effects of possessor weight and recursion are not significant. Although the test in 4.2.2 indicates a significant difference in possessum length between the two genitives, the effect is not significant when taking all variables into account and quantifying their relative importance. Gradually removing these variables from the model does not result in an increase of deviance, which suggests that animacy is the most important predictor of the variation.

	coefficients
Intercept	-7.7303*
Possessor weight	0.0139
Possessum weight	-0.5783
Recursion	-0.1673
Possessor animacy	0.5733*
Possessum animacy	0.4886*

Table 10. Coefficients of a logistic regression model

Equation (24) can be used compute logit, indicating the chances of prenominal genitive compared with postnominal genitive for a particular combination of values of independent variables. Logit can then be used to compute probabilities of a given outcome using the logistic function in, where *t* is logit.

---

9   Asterisk indicates that the effect size is significant at p < 0.001.

(24)   Logistic function

$$\sigma(t) = \frac{e^t}{1 + e^t}$$

To illustrate how to compute logits and probabilities of prenominal genitive for constructions with different parameters, I will use a simpler model where only animacy variables are used as predictors (Table 11).

	coefficients
Intercept	-8.4374*
Possessor animacy	0.5724*
Possessum animacy	0.4925*

Table 11. Coefficients of a simpler model using only animacy as a predictor

Example (25a) shows the probability of prenominal position for genitive phrase where *Jirky* (proper name, score = 8) is possessor and *brácha* 'brother' (kinship term, score = 7) is possessum. Logit is computed using the coefficients from Table 11, and is then used to compute the probability.

(25)   a.   Logit and probability of *Jirky brácha* 'Jirka's brother'

$t = -7.7303 + 0.5733 * 8 + 0.4886 * 7 = -0.4107$

$P(t) = \sigma(t) = 0.40$

        b.   Logit and probability of *kamaráda bratranec* 'a friend's cousin'

$t = -8.4374 + 0.5724 * 6 + 0.4925 * 7 = -1.5555$

$P(t) = \sigma(t) = 0.17$

This means that when possessor is expressed by a proper name and possessum is a kinship term, the model predicts 40% probability of observing prenominal genitive. This indicates that even though the probability of this construction is generally very low, it increases significantly as we go up on the animacy hierarchy. Based on the given data, the model estimates that more than one third of constructions with the given configuration (where possessor is a proper name and possessum a kinship term) will be expressed using prenominal genitive.

Example (25b) shows the probability for a phrase with referents slightly lower on the animacy hierarchy (*kamaráda* 'friend's' with score = 6 as possessor, and *bratranec* 'cousin' with score = 7 as possessum). Proceeding analogously for different animacy levels, probability of 0.1% is obtained for a prenominal phrase with two inanimate concrete nouns (both having score = 2) and a probability of 0.06% for a phrase with two abstract nouns (score = 1).

It should be noted that when taking possessive adjectives into account, the values would end up being different, as the probability of using adjectives in prenominal position instead of genitives is much higher (see Table 2).

The model predictions are consistent with the results from 4.2.1, estimating that animacy is the most important factor determining the choice of prenominal genitive, with higher effect size for animacy of possessor, while animacy of possessum still being relatively important. Thus, even though the overall probability of prenominal genitive is low, it should be expected much more often when both referents are relatively high on the animacy hierarchy.

5   Discussion

5.1   Animacy, accessibility and weight in a cross-linguistic perspective

As both the univariate analysis and predictions of logistic regression indicate, animacy appears to be the most important determinant for using prenominal possessive genitive, while no significant effects of possessor weight have been observed. In this section, the results are discussed in a broader context, where I first compare prenominal genitive to other possessive constructions in Czech and then explore similar constructions across other Slavic languages.

5.1.1   Attributive possession in Czech

The results of my corpus analysis indicate that prenominal genitive shares some features with possessive adjectives, while differing from them in other aspects. Similar to adjectives, this kind of construction encodes highly animate and accessible referents. Based on the available material, collectives of humans seem to be the cutoff point on the hierarchy, but given the small amount of data, no conclusions can be made regarding possible categoricity of the constraint.

As discussed in 2.2.4, invariant possessive adjective forms of type *-ovo/-ino* are even more similar to prenominal genitives, being structurally simpler compared to regular adjectives, encoding animate referents and occurring prenominally. Like genitives, they could be considered more economical, allowing to avoid forms with morphological alternations and complex declension.

Rather than a variation between genitive positions, this could be viewed from a functional perspective as a choice between two positional variants (cf. Křivan 2015: 155). The first variant precedes possessum, is specialized to human referents and could be expressed either by morphologically more complex and semantically more restricted possessive adjectives and possessive pronouns, or by a simpler genitive (which can be modified, and therefore could be viewed as expanding the possibilities of prenominal possessor in the dimension of weight). Second variant follows possessum, is semantically less restricted, but generally tends to be used with less animate and/or accessible referents, genitive being the only formal option.

5.1.2 Attributive possession in other Slavic languages

In this section, I compare attributive possession in Czech with typological data from other languages, focusing on Slavic language group. I present some examples of expressing attributive possession in South Slavic (Bosnian-Croatian-Serbian, Bulgarian, Macedonian and Slovene), East Slavic (Russian) and West Slavic (Polish, Slovak and Upper Sorbian) languages.

In Bulgarian, possessive adjectives are mostly restricted to kinship terms and proper names (26a) and are only rarely formed from common nouns for humans and animals (Corbett 1987). The same applies for Macedonian, a very closely related language. Bulgarian does not have genitive, using a prepositional phrase in postnominal position (26b). As an alternative to possessive pronoun, it is also possible to use a dative clitic (26c), with a further restriction in Macedonian, allowing only kinship terms as possessum.

(26) a. *Pred nas e mamin-ija-t apartament.*
          before us is mother-POSS.ADJ-DET flat
          'Before us is mother's flat.' (Corbett 1987: 310)

      b. *apartamenta-t na mama*
          flat-DET of mother
          'mother's flat' (Corbett 1987: 310)

      c. *stara-ta mu kašta*
          old-DET him-DAT house
          'his old house' (Corbett 1987: 310)

Bosnian-Croatian-Serbian typically use possessive adjectives for specific human referents, usually with some association with the speaker (27). Common nouns for animals are also acceptable.

(27) *zubar-eva kuća*
      dentist-POSS.ADJ house
      '(my) dentist's house' (O'Connor et al. 2013: 110)

In Slovene, attributive possession is commonly expressed by possessive pronouns if the possessor is animate, or by genitive in other cases. Thus, both options in (28) are possible, while (28a) is more likely than (28b).

(28)   a.   *māter-ina*   *híša*

               mother-POSS.ADJ   house

               'mother's house' (Priestly 1993: 440)

       b.   *híša*   *máter-e*

               house   mother-GEN

               'mother's house' (Priestly 1993: 440)

Like Czech, Russian allows non-modified possessive adjectives in prenominal position. These can be derived from proper names or kinship terms (29b), while genitive can be used for modified phrases (29a).

(29)   a.   *kniga*   *mojej*   *mam-y*

               book   my   mom-GEN

               '(a/the) book of my mom' (O'Connor et al. 2013: 106)

       b.   *mam-ina*   *kniga*

               mom-POSS.ADJ   book

               'mom's book' (O'Connor et al. 2013: 106)

The domain of possessive adjectives in Polish is restricted mostly to pronouns (30a). Postnominal genitive is an option for other cases (30b), but similar to Czech, it is sometimes used in prenominal position as well (30c), in particular when referents are human (Rothstein 1993).

(30)   a.   *mój*   *dom*

               my   house

               'my house' (Rothstein 1993: 747)

       b.   *dom*   *naszego*   *koleg-i*

               house our   colleague-GEN

        '(the) house of our colleague' (Rothstein 1993: 747)

c.   *naszego*   *koleg-i*   *siostra*

    our   colleague-GEN   sister

    'our colleague's sister' (Rothstein 1993: 747)

In Slovak, possessive adjectives are used in prenominal position, having similar attributes as in Czech, referring only to humans and occasionally to animals. While typically not modified, in the literature, there are examples of full names (31a) and possessive adjectives modified by possessive pronouns (31b) or attributive modifiers (31c), where the constraint of weight does not apply. In these constructions, modifiers are in genitive form and are controlled by the underlying noun of the possessive adjective. Short (1993) deems this kind of construction obsolete, while Corbett (1987) describes it as acceptable but colloquial in contrast with stylistically neutral postnominal genitive. Some examples of prenominal genitive were also found in the main subcorpus of Slovak National Corpus (31d) (Slovenský národný korpus – prim-6.0-public-all 2013).

(31)   a.   *Františk-a*   *Mik-ove*   *diela*

        František-GEN   Mika-POSS.ADJ   works

        'the works of František Mika' (Short 1993: 581)

     b.   *môjho*   *otc-ova*   *knižnica*

        my   father-POSS.ADJ   library

        'my father's library' (Corbett 1987: 316)

     c.   *nášho dobr-ého*   *sused-ova*   *záhrada*

        our   good-GEN   neighbor-POSS.ADJ   garden

        'our good neighbor's garden' (Dvonč et al. 1966: 220)

     d.   *Môjho otc-a*   *brat*   *má*   *manželk-u*

        my   father's   brother   has   wife-ACC

        Slovenk-u   *z*   *Doln-ej*   *zem-e.*

        Slovak-ACC   from   Dolná-ACC   zem-ACC

        'My father's brother has a Slovak wife from Dolná zem.'

        (MY Banskobystrické noviny 2007)

The same type of construction appears in Upper Sorbian, where possessive adjectives also lack weight constraint and can be modified in the same way (32). As in other West Slavic languages, prenominal genitive is sometimes used instead (Faßke 1996: 67).

(32) *To   je   našeho   wučerj-owa   zahrodka.*
      that is   our   teacher-POSS.ADJ   garden
      'That is our teacher's garden.' (Stone 1993)

In addition to allowing modification, adjectives in examples from Slovak and Upper Sorbian can also serve as antecedents for anaphora. Because regular adjectives lack these properties, it may be an argument for approaching these weight constraints as properties of possessive constructions across languages, not attributable to their adjectival status only (O'Connor et al. 2013; Corbett 1987).

5.1.3 Categorical constraints and statistical tendencies

There are many languages where the same constraints we observed for Slavic languages are present only as statistical patterns. In English, there is a tendency to use s-genitive for animate, accessible and syntactically light possessors (e.g. Hinrichs & Szmrecsanyi 2007; O'Connor et al. 2013; Rosenbach 2003). Afrikaans also prefers prenomional position for animate nouns, but compared to English, inanimate nouns are used more freely in this position (Rosenbach 2017). On the other hand, West Flemish, a closely related language, has two prenominal phrasal possessive constructions, both allowing only humans or personified animal names as prenominal possessors (Haegeman 2013). These correspondences are a case of *stochastic generalization* (e.g. Manning 2003), which states that statistical tendencies in some languages are very often present as rules in grammars of other languages. Similarly, Hawkins (2004) argues that grammatical rules are conventionalizations of preferences in language use.

In a corpus study of genitive variation in English, O'Connor et al. (2013) explored correspondences between probabilistic patterns in English and categorical restrictions of several European languages. They identified a number of languages

containing what they call a *Monolexemic Possessor Construction* (MLP), characteristic for being animate, highly accessible, one-word and prenominal.

In Slavic languages, splits induced by animacy or accessibility are slightly different in each language, with Russian prenominal possessor being more restrictive, while Bosnian-Croatian-Serbian allowing a broader range of semantic categories. However, the patterns are very similar. In all of these languages, highly animate and accessible referents are preferred in prenominal position. Possessor weight appears to be a less important factor across Slavic languages generally, with some of them allowing only one-word, whereas others being more generous. In Upper Sorbian and to some extent also Slovak, only animacy and accessibility appear to be relevant for possessive adjectives. In addition, Slovak, Polish and Czech all have prenominal genitive, violating the one-word constraint, but sharing other features with MLP.

5.2 Recursion

Results from section 4.2.3 indicate that potential recursion does not contribute significantly to avoidance of prenominal position. In this section, I will discuss these results and propose some possible explanations and hypotheses.

When choosing the coding scheme for this variable, discrete numeric scale starting from zero was used. However, there was only one example with more than one nested possessor, which could be explained by certain cognitive limitations on recursive processing. Using a connectionist model supported by a series of experiments employing grammatical acceptability judgment tests, Christiansen & MacDonald (2009) examined processing of complex recursive structures in English, such as center-embeddings, as well as simpler left- and right-branching structures, including prenominal possessive genitive. Their results suggest that increased recursion depth has a negative effect on processing that should not be attributed only to effects of length. If I assume that similar restrictions apply in other languages, this may be one of the explanations for the absence of nested possessive phrases with more than one level of recursion in the dataset. Examples from ORALv1 corpus (33) show some strategies that speakers used trying to avoid more than one nested possessor.

(33) a.   *Michalový*   *mámy*   *manžel*   *tak*

Michal-POSS.ADJ   mother-GEN   husband   so

jeho   *táta*

his   dad

'Michael's mother's husband, so his son.'

b.   *Takže* *bratrance*   *vlastně*   *našeho*   *táty*

So   cousin-GEN   actually   our-GEN   dad-GEN

bratrance   *to*   *je*   *syn*

cousin-GEN   it   is   son

'So, cousin's, actually our dad's cousin's [it is] son'

Nevertheless, as the results in 4.2.3 suggest, singly nested possessive phrases are quite common in the prenominal position where it could simply be avoided using the postnominal genitive. Following the discussion in section 5.1, I suggest some hypotheses regarding this tendency.

Based on cross-linguistic observations and results of psycholinguistic experiments, Hawkins (2004; 2014) proposed three principles of economy and efficiency in language use and grammars. He summarized the first two of them, including Minimize Forms discussed in 3.2.4 with the slogan 'Express the most with the least'. The third principle, summarized as 'Express it earliest' is defined in (34).

(34)   Maximize Online Processing (MaOP)

The human processor prefers to maximize the set of properties that are assignable to each item X as X is processed, thereby increasing O(nline) P(roperty) to U(ltimate) P(roperty) ratios. The maximization difference between competing orders and structures will be a function of the number of properties that are unassigned or misassigned to X in a structure/sequence S, compared with the number in an alternative.

(Hawkins 2004: 51)

This means that in processing sequences of formal units and their syntactic and semantic properties, early property assignments are more efficient than delayed property assignments. The example in (35) illustrates this principle. Even though the processing domain[10] is the same in both (35a) and (35b), there is a delay in assignment of a semantic property (*goal* in this case) where this property could be assigned earlier.

(35)   a. John VP[**went [to London] [in the late afternoon] [after** a long siesta]]
        b. John VP[**went [in the late afternoon] [to London] [after** a long siesta]]
        (Hawkins 2004: 49)

There are other types of constructions where instead a 'look ahead' may be required to assign properties to a certain category, postnominal possessive genitive being an example of such construction. Using example (36), I will illustrate the difference in timing of property assignments for two nested possessive phrases. These phrases express three direct relations: speaker and his brother (expressed implicitly), (speaker's) brother and his girlfriend and (speaker's brother's) girlfriend and her cousin.

(36)   a.   *bratranec*   *bráchovy*   *přítelkyně*
              cousin   brother-POSS.ADJ   girlfriend-GEN
              '(a/the) cousin of (my) brother's girlfriend'

        b.   *bráchovy*   *přítelkyně*   *bratranec*
              brother- POSS.ADJ   *girlfriend-*GEN   *cousin*
              '(my) brother's girlfriend's cousin'

As the first unit is processed in (36a), the relation between its referent and referents of subsequent units is not yet assigned. Processing the second unit, the implicitly expressed relation between the speaker and the referent of the second unit is assigned, and only after processing the third unit, the whole relationship

---

10 Length of a structure that allows to process all relevant syntactic properties of a given construction. See Hawkins (2004: 31) for a definition of his Minimize Domains (MiD) principle.

45

could be correctly interpreted. On the other hand, (36b) assigns the relationship between the speaker and the referent of *bráchovy* earlier, right after processing the first unit, and there is an additional property assignment with processing of each subsequent unit.

Another explanation consistent with the previous discussion follows the idea of an activation scale of concepts in speaker's mind (Chafe 1987). Chafe proposed three activation states, based on the assumption that only a small amount of information could be active at a particular moment:

> "An active concept is one that is currently lit up, a concept in a person's focus of consciousness. A semi-active concept is one that is in a person's peripheral consciousness, a concept of which a person has a background awareness, but which is not being directly focused on. An inactive concept is one that is currently in a person's long-term memory, neither focally nor peripherally active." (Chafe 1987: 25)

The terms *active* and *inactive* are alternatives for the more traditional terms *given* and *new.* Chafe proposes what he calls *light starting point constraint*, which states that most often, speakers use a given referent as a starting point, while occasionally semi-active referent may also be a starting point. Efficient packaging of information, both for the speaker and the hearer, allowing to start with the most active and proceeding to less active referents, may be another reason why nested prenominal possessive constructions are not uncommon in the sample. Almost all nested phrases start with pronouns, which are considered as having a very high degree of accessibility (Ariel 2001).

When the examples of prenominal genitives from ORALv1 are examined more closely, considering both prepositional genitives and prepositional constructions this time (see Table 2), the majority of the phrases contain various kinds of disfluencies such as pauses, hesitations, false starts and filler words. This might be an indication of an increased demand on retrieval from speaker's memory in this types of constructions. Therefore, it may be more efficient to start with the most active referent, then accessing the previously semi-active and finally previously inactive referents. The advantage is that speaker may proceed linearly from the most active referent to the least active one.

Returning to example (36) and representing possessive relations as trees, (36b) should be more efficient, as the path from the most active element to the least active one is linear. In (36a), the speaker starts with the least active node and only after that proceeds to the most active node, while the activation of the last unit is somewhere between the first two.

Another factor, although applicable only to some cases and probably not as important as those discussed above, is avoidance of possible ambiguity. Comparing the postnominal alternative (37a) to other options, it is not clear if the relationship between referents should be parsed as 'cousin of (my) brother's girlfriend' or if the referent of the phrase is 'cousin of (my) girlfriend's brother'. This kind of construction is absent in the dataset and a vast majority of nested postnominal phrases denoting human referents are combinations of possessive pronoun and genitive. On the other hand, there are some examples of constructions of type (37c), but their number in the sample is too small to make any conclusions.

(37)   a.   *bratranec*   *bráchy*   *přítelkyně*

               cousin   brother-GEN   girlfriend-GEN

               'the cousin of (my) brother's girlfriend'

     b.   *bratranec*   *bráchovy*   *přítelkyně*

               cousin   brother-POSS.ADJ   girlfriend-GEN

               'the cousin of (my) brother's girlfriend'

     c.   *bráchy*   *přítelkyně*   *bratranec*

               brother-GEN   girlfriend-GEN   cousin

               '(my) brother's girlfriend's cousin'

     d.   *bráchovy*   *přítelkyně*   *bratranec*

               brother-POSS.ADJ   girlfriend-GEN   cousin

               '(my) brother's girlfriend's cousin'

Considering the previous discussion related to the results from 4.2.1 and 4.2.4, (37c) and (37d) may have an advantage of expressing the most accessible referent

first, while (37c) also allows using a more economical variant, avoiding adjective with complex declination, and in some cases morphological alternation as well.

In sum, examining the corpus data more closely revealed interesting patterns in using recursive possessive phrases. While I proposed some possible explanations, the amount of material is rather small to test these assumptions. Because this kind of construction is relatively rare and relying on corpora of spoken language would probably not allow to test these assumptions quantitatively, a psycholinguistic experiment should be designed to further examine processing of recursive possessive phrases in Czech.

5.3   Prenominal genitive as a more economical variant

In 2.2.4 I discussed some properties of prenominal genitive as an alternative to other options of expressing attributive possession in terms of economy. Even though the amount of available data is small, results in 4.2.4 suggest that speakers may use this alternative as a more economical variant, avoiding alternations that require increased articulatory effort. As suggested by Nová (2016) and discussed in 2.2.4, this may also be the reason for very low frequency of invariant possessive adjectives formed from feminine nouns. Frequencies of this type of adjective in the dataset used in the present thesis are consistent with this argument. Among 43 examples of this invariant adjective form, only one was formed from a feminine noun.

6   Conclusion

In this thesis, I have examined prenominal possessive genitive in Czech and several factors assumed to have an effect on the use of this variant. The results suggest that animacy is an important predictor and prenominal genitive tends to be used with highly animate and accessible referents, being functionally similar to possessive adjectives, while allowing modification. Interesting results have been found regarding nested possessive phrases, indicating that even though recursion increases the weight of a phrase, it does not prevent speakers from using prenominal position. I have suggested some hypotheses regarding possible advantages of using prenominal nested possessive phrases in certain contexts. These hypotheses should be further tested through psycholinguistic experiments. The results also suggests that prenominal possessive genitive may often substitute possessive adjectives as a more economical variant, allowing to avoid morphophonological alternations.

References

Ariel, Mira. 2001. Accessibility theory: An overview. In Sanders, Ted & Schilperoord, Joost & Spooren, Wilbert (eds.), *Text Representation: Linguistic and Psycholinguistic Aspects [Human Cognitive Processing 8]*, 29–87. Amsterdam: John Benjamins.

Behaghel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25. 110–142.

Chafe, Wallace. 1987. Cognitive Constraints on Information Flow. In Tomlin, Russel (ed.), *Coherence and Grounding in Discourse,* 21–51. Amsterdam: Benjamins.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Christiansen, Morten & Charter, Nick. 2015. The language faculty that wasn't: a usage-based account of natural language recursion. *Frontiers in Psychology* 6(1182). 1-18.

Christiansen, Morten H. & MacDonald, Maryellen. 2011. A Usage-Based Approach to Recursion in Sentence Processing. *Language Learning* 51(1). 126-161.

Comrie, Bernard. 1989. *Language Universals and Linguistic Typology.* 2nd edn. Oxford: Blackwell.

Corbett, Greville. 1987. The Morphology/Syntax Interface: Evidence from Possessive Adjectives in Slavonic. *Language* 63(2). 299-345.

Corbett, Greville. 2000. *Number*. Cambridge: CUP.

De Vries, Meinhou & Christiansen, Morten, & Petersson, Karl. 2011. Learning recursion: Multiple nested and crossed dependencies. *Biolinguistics* 5(1/2). 10-35.

Dixon, Richard M. W. 2010. Possession. In *Basic linguistic theory, Volume 2, Grammatical topics,* 262–312. Oxford: Oxford University Press.

Dvonč, Ladislav & Horák, Gejza & Miko, František & Mistrík, Jozef & Oravec, Ján & Ružička, Jozef & Milan Urbančok. 1966. *Morfológia slovenského jazyka*. Bratislava: Vydavateľstvo Slovenskej akadémie vied.

Evans, Nicholas & Levinson, Stephen. 2009. The myth of language universals: language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32(5): 429-48.

Everett, Daniel L. 2005. Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology* 46(4). 621-634.

Faßke, Helmut. 1996. *Sorbischer Sprachatlas 15: Syntax*. Bautzen: Domowina.

Haegeman, Liliane. 2013. Two prenominal possessors in West Flemish. In Börjars, Kersti & Denison, David & Scott, Alan (eds.), *Morphosyntactic Categories and the Expression of Possession*, 219-251. Amsterdam: John Benjamins.

Hauser, Marc & Chomsky, Noam & Fitch, Tecumseh. 2002. The faculty of language: what is it, who has it, and how did it evolve? *Science* 298. 1569–1579.

Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. New York: Oxford University Press.

Hawkins, John A. 2014. *Cross-linguistic Variation and Efficiency*. New York: Oxford University Press.

Hinrichs, Lars & Szmrecsanyi, Benedikt. 2007. Recent changes in the function and frequency of Standard English genitive constructions: A multivariate analysis of tagged corpora. *English Language and Linguistics* 11(3): 437–474.

Hosmer, David & Lemeshow, Stanely. 2000. *Applied Logistic Regression*. New York: Wiley.

Karlsson, Fred. 2007. Constraints on multiple center embedding of clauses. *Journal of Linguistics* 43. 365–392.

Kopřivová, Marie & Lukeš, David & Komrsková, Zuzana & Poukarová, Petra & Waclawičová, Martina & Benešová, Lucie & Křen, Michal. 2017. *ORAL: korpus neformální mluvené češtiny, verze 1 z 2. 6. 2017*. Ústav Českého národního korpusu FF UK, Praha. Retrieved from http://www.korpus.cz

Koptjevskaja-Tamm, Maria. 1993. *Nominalizations*. London: Routledge.

Křivan, Jan. 2015. *Teoretické nástroje pro kategorizaci a funkční explanace vztahů v jazyce.* Doctoral Dissertation. Retrieved from *https://dspace.cuni.cz/handle/20.500.11956/82584*

Levshina, Natalia. 2015. *How to Do Linguistics with R. Data Exploration and Statistical Analysis*. Amsterdam/Philadelphia: John Benjamins.

Manning, Christopher. 2003. Probabilistic Syntax. In Bod, Rens & Hay, Jennifer & Jannedy, Stefanie (eds.), *Probabilistic Linguistics,* 289–341. Cambridge MA: The MIT Press.

MY Banskobystrické noviny. 2007. Týždenník. Bratislava: Petit Press 18.09.2007. In Slovenský národný korpus – prim-6.0-public-all. Bratislava: Jazykovedný ústav Ľ. Štúra SAV 2013. Retrieved from https://korpus.juls.savba.sk.

Nová, Jana. 2015. K vyjadřování posesivity v dialektech češtiny a v běžném úzu mládeže z oblasti západočeského pohraničí. *Naše řeč* 3. 132-146.

O'Connor, Catherine & Maling, Joan & Skarabela, Barbora. 2013. Nominal categories and the expression of possession: A cross-linguistic study of probabilistic tendencies and categorical constraints. In Börjars, Kersti & Denison, David & Scott, Alan (eds.), *Morphosyntactic Categories and the Expression of Possession*, 219-251. Amsterdam: John Benjamins.

Perry, James W. & Kent, Allen & Berry, Madeline M. 1955. Machine literature searching X. Machine language; factors underlying its design and development. *American Documentation* 6(4). 242-254.

Priestly, Tom M. S. 1993. Slovene. In Comrie, Bernard & Corbett, Greville (eds.), *The Slavonic Languages,* 388-451. London and New York: Routledge.

R Core Team. 2018. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna.

Rosenbach, Anette. 2003. Aspects of iconicity and economy in the choice between the s-genitive and the of-genitive in English. In Rohdenburg, Günter & Mondorf, Britta (eds.), *Determinants Of Grammatical Variation in English,* 379-412. Berlin, New York: Mouton de Gruyter.

Rosenbach, Anette. 2017. Constraints in contact: Animacy in English and Afrikaans genitive variation – a cross-linguistic perspective. *Glossa: a journal of general linguistics* 2(1): 72.1–22.

Rothstein, Robert A. 1993. Polish. In Comrie, Bernard & Corbett, Greville (eds.), *The Slavonic Languages,* 686-758. London and New York: Routledge.

Short, David. 1993. Slovak. In Comrie, Bernard & Corbett, Greville (eds.), *The Slavonic Languages,* 533-592. London and New York: Routledge.

Silverstein, Michael. 1976. Hierarchy of Features and Ergativity. In Dixon, Robert. M. W. (ed.), *Grammatical Categories in Australian Languages*, 112-171. Canberra: Australian National University.

Slovenský národný korpus – prim-6.0-public-all. 2013. Bratislava: Jazykovedný ústav Ľ. Štúra SAV. Retrieved from https://korpus.juls.savba.sk.

Stone, Gerald. 1993. Sorbian. In Comrie, Bernard & Corbett, Greville (eds.), *The Slavonic Languages,* 593-685. London and New York: Routledge.

Szmrecsanyi, Benedikt 2004. On Operationalizing Syntactic Complexity. In Purnelle, Gerald & Fairon, Cedric & Dister, Anne (eds.), *Le poids des mots. Proceedings of the 7th International Conference on Textual Data Statistical Analysis. Louvain-la-Neuve, March 10-12, 2004,* 1032-9. Louvain-la-Neuve: Presses universitaires de Louvain.

Uličný, Oldřich. 2018. *Posesivní Genitiv Redivivus*. Jazykovedný časopis 69(3): 497-501.

Voráč, Jaroslav. 1950. K jazykovému zeměpisu Čech. Sousedovo brambory, Novákojc a p. v českých nářečích jihozápadních. *Naše řeč* 34: 92–97.

Wasow, Thomas (1997). Remarks on grammatical weight. *Language Variation and Change* 9. 81-105.

Zipf, George. 1949. *Human Behavior and the Principle of Least Effort.* New York: Hafner.

Zaenen, Annie, Carlette, Jean, Garettson, Gregory, Bresnan, Joan, Koontz-Garboden, Andrew, Nikitina, Tatiana, O'Connor, Mary Catherine & Wasow, Thomas. 2004. Animacy encoding in English: Why and how. In Byron, Donna & Webber, Bonnie (eds.), *Proceedings of the 2004 ACL Workshop on Discourse Annotation, Barcelona, July 2004*, 118–125. Stroudsburg PA: Association for Computational Linguistics.