Charles University

Faculty of Arts

Autoreferát dizertační práce

Mgr. Michal Láznička

Discourse Production of Czech Speakers with Aphasia

A Usage-based Exploration

Produkce diskurzu českých mluvčích s afázií: Explorace s využitím usage-based lingvistiky

Department of Linguistics

Study program: Filologie (obecná lingvistika)

Dissertation supervisor: Mgr. Eva Lehečková, Ph.D.

2022

Table of contents

Introduction

The present work is an exploration of connected speech of Czech speakers with aphasia with the use of usage-based linguistics. Aphasia is a condition with significant negative impact on the quality of life of persons suffering from it as well as their significant others. Hundreds to lower thousands of people are affected by aphasia every year and, given the demographic trends, this number can be expected to increase in the coming years. The condition selectively impairs language and linguistics is thus an important part of aphasia research. Despite this fact a closer collaboration between linguists and aphasiologists has been rare in Czechia. The aim of the present work, motivated by this lack of interaction, was threefold. First, I assembled a corpus of connected speech of persons with aphasia that can provide other researchers with data that can be used to pursue and generate different research questions related to this topic. Second, I provide a first detailed characteristic of connected speech in aphasic Czech by using various measures of discourse production. A further, more generally framed aim was to show that usage-based construction grammar provides a theoretically motivated set of descriptive and analytical concepts and tools that can provide a better understanding of what happens with language in aphasia, grounded in a model of language representation and processing. Using the data collected in the corpus, I focus on the production of verbs, inflected nouns, and prepositional phrases to demonstrate the importance of various frequency measures for the analyses of aphasic language.

General introduction

This section provides a brief overview of the general context as well as the conceptual-analytical framework of usage-based construction grammar adopted in the present work.

Usage-based linguistics

The present work adopts usage-based construction grammar as a model of language representation and processing and uses this conceptual framework to describe some aspects of connected speech production of Czech speakers with aphasia.

The term usage-based refers to the idea that language structure is directly affected by usage and that structure and systematicity in fact emerge from usage.[1][2] Individual language users possess a representation of language and grammar which is based on their linguistic experience which modulates mental grammars throughout lifespan. Several important points follow from this basic assumption. First, performance factors/forces/mechanisms that have been viewed as secondary and unimportant in the study of language competence in generative grammar are an integral part of a model of language because these directly influence/shape structure of language. Second, the use of (big(ish)) authentic language data and robust empirical methods is an integral part of the approach, i.e. usage-based analyses make use of large corpora of written, spoken, and signed language but also experimental methods and, increasingly, the combination thereof (e.g. Divjak, Dąbrowska & Arppe 2016). Third, the idea that mental grammars remain dynamic and may be "updated" during each and every

1 Language is in this respect viewed as a complex adaptive system (Beckner et al. 2009), i.e. a system that emerges from localized interactions of agents who employ different strategies to achieve their (communicative) goals and adapt their strategies for future interactions based on the outcomes of past experience.

2 The following is based on the claims presented e.g. in Tummers and colleagues (2005), Bybee (2010), Taylor (2012), Diessel (2015), and Goldberg (2019).

usage event implies that language knowledge of members of a given language community need not, in fact, should not be identical (e.g. Dąbrowska 2012; Verhagen 2020; or Barking, Backus & Mos 2022). The basic unit of language representation in usage-based construction grammar is a construction. Constructions in the technical sense are defined as direct pairings of form and function (meaning) with varying degrees of schematicity and inner complexity. The function pole may include semantic, contextual, and pragmatic meanings as well as world knowledge (Goldberg 1995; Croft 2001; Fried & Östman 2004). Constructions are represented in language knowledge in a network (of networks) in a multidimensional conceptual space. The networks are organized based on overlaps of function and/or form (Fillmore 1988; Goldberg 2019; Diessel 2020). Language processing operates on this network of constructions. Frequency of use plays a crucial role in these processes and has become one of the central and most studied notions in usage-based linguistics (Bybee 2007; Divjak 2019). High frequency and, consequently, repeated processing leads to stronger memory traces, routinization, and easier access. Crucially, frequency effects are driven not only by absolute frequency of single words, but also by cooccurrence frequencies of two or more words and relative frequencies of inflectional variants and constructions.

Aphasia: definition, syndromes, and symptoms

A standard definition of acquired aphasia which is the topic of the present work states that it is an acquired neurogenic language specific disorder (Hallowell & Chapey 2008). This means that the language capacity is affected selectively as a cognitive domain and general intellect and other cognitive capacities are spared/intact. It is caused by lesions, most typically induced by stroke, to the brain regions that are involved in language processing, i.e., in particular, the so called perisylvian regions of the left hemisphere (cf. #figure), meaning that the consequent language impairment is not caused by sensorimotor problems. Aphasia may impact all levels of language structure, language production and comprehension, and both spoken or signed and written language. Typical symptoms include anomias (word finding problems), paraphasias (“misretrieval” of phonemically or semantically similar), and agrammatism or paragrammatism (omission or substitution of function words and inflectional morphemes). Aphasia can be classified into different clinical profiles/subtypes based on the localization of the lesion and the dominant symptoms. The standard classification includes eight types that are based on the level of fluency, comprehension, repetition, and naming.

Aphasia and linguistic theory

There are good arguments for a close cooperation between linguistics and aphasiology from both fields. From both a clinician’s and an aphasia researcher’s perspective, there is a need for at least a rudimentary descriptive apparatus that is needed to describe a person with aphasia’s linguistic behavior. Such description can be theory-neutral, or theoretically shallow and only make use of basic notions that would be based on a particular descriptive tradition/convention of the language. For instance, for Czech this would be a basic inventory of word classes and inflectional categories as found in secondary education grammar books. Alternatively, one can adopt a specific theory/model of language representation and processing. Such an approach has, potentially, two merits. Firstly, a model of language may serve as a basis for generating predictions and explanations of linguistic behavior in aphasia. Furthermore, an application of a specific model of language in therapy has the potential to provide more sensitive and reliable assessment and therapy tools.

From the perspective of theoretical linguistics, the use of aphasic data also has clear benefits, pri-

marily in model building and “criticism”. If we assume that one of the aims of linguistic research is to develop a model of language representation and processing, such models should not only explain/account for data of neurotypical, adult native speakers, but also for other groups, such as children in the process of L1 acquisition, non-native speakers, or speakers with language impairments. A general model of language representation and processing should be able to explain why certain phenomena are more prone to error in aphasia to predict potential impairment of a language component based on an observation of a different impaired component that is according to a model subserved by the same mechanism. Aphasia has played such a role in some areas of psycholinguistic research where patterns of errors are used to argue for specific components of models of language processing (e.g. Dell et al.’s model of lexical retrieval 1997) and, most prominently, in neurolinguistic research where mappings between lesion sites and symptoms serve as evidence to support models of the neurobiology of language. However, it is still relatively rare for general psycholinguistics and even more so for theoretical linguistics to use/reflect on pathological data.

Aphasia in Czech

Czech may be said to be an under-studied language in the context of aphasia. Lehečková (2016; 2009; 1985) was the first to study and describe aphasia in Czech from the perspective of linguistics. Using interviews with speakers with aphasia and employing Jakobson (1980)’s regression hypothesis, she focused on typical agrammatic and paragrammatic errors and described which linguistic categories are prone to errors and what the most frequent directions of substitution errors are. For example, the nominative is typically used in place of the other cases and the feminine gender is substituted by the masculine. An important contribution to the study of Czech in aphasia was done by Flanderková and collaborators (e.g. Flanderková et al. 2014; Hudousková et al. 2014) who conducted several experiments focused on verb processing. A much needed gap that the present work strives to fill (or rather initiate the filling of) is the complete lack of detailed descriptions of connected speech on the one hand, and a tighter integration of a general linguistic theory in the study of aphasic Czech as well as the reflection of recent advances in linguistic theory.

Usage-based approaches to aphasia

Apart from the studies grounded in the Competition Model (Bates & MacWhinney 1989; Presson & MacWhinney 2011), linguistic research of aphasia has been until relatively recently dominated by rule-and-lexicon structuralist approaches that have stressed the role of structural complexity in aphasic processing. However, in recent years, studies/analyses of language in aphasia started to emerge that adopt the usage-based approach. These studies have focused on the role of performance factors, mainly frequency and increasingly also the probability of occurrence of inflected word forms and/or cooccurrence of words in constructions and multiword units, or processing strategies grounded in world knowledge. Two types of studies can be found in usage-based approaches to aphasia that bear direct relevance for the present work. First, the research has focused on the role of frequency and association/collocational strength measures in the characteristic and assessment of connected speech production in aphasia (Zimmerer et al. 2018; Bruns et al. 2019). Second, researchers have addressed the role of “usage” factors in error patterns observed in language production and comprehension in aphasia, using both observational and experimental data (Gahl 2002; Hatchard & Lieven 2019). These studies suggest that the application of the usage-based perspective has the potential to open new avenues of aphasia research and provide accounts of language in

aphasia with more explanatory power. This follows from the fact that the usage-based model is performance oriented and draws attention to such phenomena as frequency of use, similarity, semantics, textual and situational context, or recency and gives space to inter- as well as intra-categorial differences.

Rationale for a usage-based analysis of aphasic language: a practical introduction

A token from the corpus described in the present work may serve as an example of how the usage-based framework may be applied to aphasic data. Participant aa3 from the corpus has chronic conduction aphasia. Among several other paragrammatic errors, he produced the following during his sesssion: *vytáh 0.2 toho ryby 0.2 toho ry- rybu 0.5 a: sežral jí* '(he) pulled 02. the fish-pl 0.2 the fi- fish-sg out 0.5 and devoured it'. The token of interest here is the erroneously pluralized form *ryby* in the context where only a single fish is present.[3] This substitution would be difficult to explain in structuralist terms as substitutions would be expected from a marked category value to the unmarked, i.e. from plural to singular, but not the other way around. However, the usage-based framework offer a natural explanation. Since frequency of use is taken to be one of the most prominent factors in language representation and processing in usage-based linguistics, the first step would be to explore the distributional properties of the lemma ryba in a corpus. A query of the corpus of spoken Czech shows that ryba is a so called plural-dominant word, i.e. it is one of the words that are used more frequently in plural. This is not surprising, given the state of affairs in the real world where fish typically occur in grater quantites/groups and are consequently talked about this way. In fact, the word form ryby in its plural use accounts for more than 45 % of all occurrences of the lemma in the spoken corpus (Figure 1). Under the usage-based model, this means that this form is expected to have a strong memory trace and to be easier to access. This has important consequences for aphasia, since problems of language processing, such as word retrieval, are taken to reflect computational overload under limited resources that are available to speakers due to lesions to the language processing network in the brain. This means that a speaker trying to produce a substantially less frequent word form may fail and retireve a more available word form in its stead. This simple example demonstrates how employing the usage-based framework may help us explain some of the patterns of linguistic behavior we encounter in persons with aphasia. The same logic was applied in the present work on three different phenomena to show how word frequency, relative frequency, and cooccurrence frequency can be used to describe aphasic Czech and open new avenues of research with implications for the assessment and therapy of aphasia.

3 While the form itself might also be interpreted as singular genitive, the wider context justifies the analysis of this as a pluralization error.

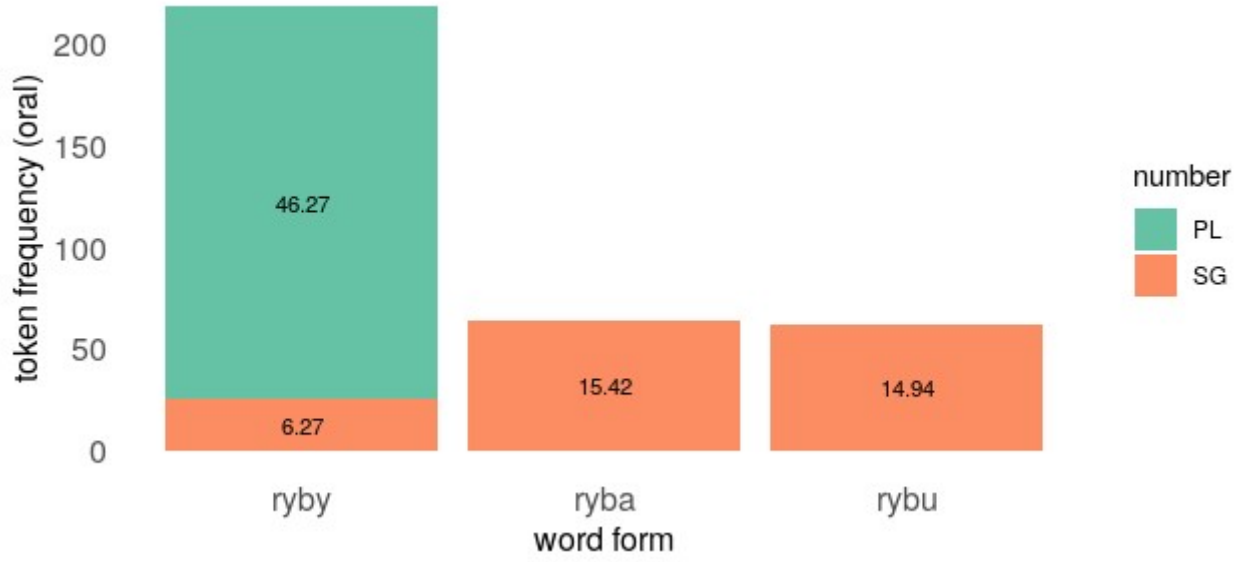

Figure 1: Distribution of word forms of the lemma ryba

Building a corpus of Czech aphasic speech

One of the aims and outcomes of the present work was the assembly of a corpus of connected speech of Czech speakers with aphasia that would be available to students, researchers, and SLTs and would thus help to facilitate and entice further interest and research of aphasia in Czech.

Samples of descriptive, narrative, procedural, and conversational discourse included in the corpus. Conversational discourse was based on participants' personal or professional life, hobbies, and speech language therapy. Descriptive and narrative discourse was elicited using a short clip and several pictures. Procedural discourse was obtained using a map task or elicitng description of meal or beverage preparation. A total of seven hours and 56 minutes was recorded with individual sessions ranging in length between 27 and 60 minutes.

Eleven native speakers of Czech with chronic aphasia participated in the interviews. General characteristics of the participants are given in #table. To provide comparison with neurotypical speakers, three individuals with no recorded history of linguistic or other cognitive impairment who matched the demographics of the participants with aphasia were additionally recorded, using a subset of three tasks that were used in the present analyses.

participant	gender	age	m/y of onset	etiology	aphasia type
aa1	M	69	n/a/07	TBI	Anomic
aa2	M	74	4/14	CVA	Mixed aphasia
aa3	M	69	12/03	TBI	Conduction
aa4	M	63	03/09	CVA	TMA

ba1	F	41	11/13	CVA	Broca's
ba2	F	63	3/04	CVA	Anomic
ba3	M	77	5/15	CVA	Anomic
ba4	M	57	12/02	TBI	Broca's
pa1	F	66	2/15	CVA	Anomic
pa2	M	52	8/01	CVA	TMA
pa3	M	84	3/15	CVA	Conduction
ac1	M		--	--	--
ac2	F		--	--	--
ac3	M		--	--	--

The transcription was performed in ELAN (Wittenburg et al. 2006). Each recording was segmented into turns and these into c-units (Loban 1966). For the purposes of the transcription, a c-unit was defined as a main-clause-like predication with all dependent predications or, in cases of a missing verb, as a semantically cohesive whole.

Transcription rules were based on the system used for the transcription of recordings for corpora of spoken Czech with adjustments that enabled more detailed transcription of disfluencies. All silent pauses exceeding 200 ms were marked based on relevant interactional research.

The transcripts were tokenized, lemmatized and morphological tagged with the MorphoDiTa tagger (Straková, Straka & Hajič 2014). Errors caused by differences between transcription systems were corrected (semi)automatically, other errors required manual correction. The whole corpus has the parameters summarized in #table.

participant group	c-units	positions	word tokens
administrator	4355	20437	16954
speakers with aphasia	4449	31422	23164
neurotypical speakers	279	2390	1897

A subset of the corpus that was mainly used for the analyses presented in following chapters was corrected manually. The parts of the corpus that were used in the analyses include the retelling of the Lion cage scene (henceforth referred to as "chaplin"), the description of the hog-killing picture ("lada"), and the story creation based on the three picture strip with cat and cake ("comic"). The subcorpus contains 1396 c-units produced by participants and 11366 tokens of which 7648 are fully formed words (including paraphasias). The transcript, lemma, and morphological tag layers were complemented by error annotation used to mark and roughly classify aphasic errors. A simple coding scheme was used in which whole c-units (instead of individual tokens) were marked for errors.

The data was coded for a range of variables that allowed a detailed description of participants' per-

formance using measures that are typically used in quantitative discourse analyses of aphasic speech. These measures are summarized in the following section.

Analyses of aphasic discourse

Two types of descriptions and analyses were provided for the subcorpus. In the first part, I described the data in terms of a suite of measures of discourse production that are standardly used in discourse analyses of persons with aphasia and performed a comparison between a group of persons with fluent and non-fluent aphasia and the three neurotypical speakers. In the second part, I focused on the production of lexical verbs and inflected nouns in the sample to show how cumulative lemma frequencies and relative frequencies of inflected word forms can be used in aphasia research. The third analysis in this part focused on prepositional phrases (PP) in the sample with regards to the relationship between frequency of words and multiword units on the one hand and the occurrence of disfluencies on the other.

Discourse production measures

Despite the fact that (interactional) discourse is the most natural form of language from the perspective of ecological validity, linguistic research of aphasia was for a long time focused on subword and single word processing and sentence level processing, while a rise of interest in discourse production and comprehension emerged later. This is somewhat similar to linguistics as a field in general. In their review of discourse analyses, Bryant and colleagues (2016) found that the number of discourse analyses had doubled in the period after 1995 compared to 1975-1995 and there was a further increase, particularly in treatment studies, in the late 2000s. This area of aphasia research is characterized by a great variability of approaches and methods. Bryant and colleagues (2016) report over 500 measures of discourse productivity in their survey of 165 papers.

While most of the analyses of discourse in aphasia have been quantitative (using text based measures or expert ratings), some authors have also used Conversation analytical methods, focusing particularly on the organization of turn taking, adjacency pair structure and the management of repairs (Boles 1998; Wilkinson et al. 1998; Damico, Oelschlaeger & Simmons-Mackie 1999; Perkins, Crisp & Walshaw 1999). Quantitative analyses have focused on micro- and macrostructural characteristics of discourse (see Armstrong 2000; Prins & Bastiaanse 2004; or Linnik, Bastiaanse & Höhle 2016 for reviews). Microstructure analyses use a suite of measures that are based on clause- and word-based phenomena such as mean length of utterance (MLU), well-formedness of clauses, lexical diversity, or semantic and argument structural characteristics of verbs, i.e. features that revolve mostly around the concepts of fluency and grammaticality. Macrostructure has to do with the organization of discourse and derived measures are typically based on informativeness of the text, cohesion and coherence, or thematic components of discourse. Analyses have increasingly focused on combining these two levels of structure in multilevel analyses. In a representative example, Marini et al. (2011) analyzed samples of 300-400 words combined microstructural measures with macrostructural indices to describe the production of two individuals with aphasia and showed that these combined measures can reveal problems that are not detected by standard assessment tools.

Following up on the literature, I present a characteristic of Czech discourse in aphasia in this section. The measures and dimensions of description I used are focused on microstructure and can be grouped into fluency (e.g. mean length of c-unit), productivity (e.g. lexical diversity), and well-

formedness (e.g. proportion of sentence fragments) measures.

Participants were grouped into three groups according to fluency (fluent, nonfluent, neurotypical), based on SLT assessments. These groups were compared using the non-parametric Kruskal-Wallis test followed by post hoc pairwise multiple comparison using the Dunn's test with Holm correction for multiple comparisons. The comparison included a total of 18 variables. Statistically significant group differences between neurotypical participants and speakers with non-fluent aphasia were found for the following measures: MLU, mean number of fluent trigrams, mean difluencies per word, zTTR, proportion of multiclause c-units, mean number of modifiers, mean number of verbs per c-unit, proportion of story-relevant words, proportion of fragments, proportion of interventions, and proportion of c-units with errors. A significant difference between the fluent and non-fluent participants with aphasia was found for MLU, mean number of modifiers, mean number of verbs per c-unit, mean ratio of nouns to nouns and verbs, proportion of content and function words.

After exploring individual variation, these measures were submitted to a hierarchical clustering analysis to visualize the inner structure of the groups.[4] A solution with two clusters was selected as best representing the structure of the data using average silhouette width as criterion. The resulting cluster dendrogram iand distances between individual participants in two dimensional space are plotted in Figure 2. All non-fluent participants were grouped in one cluster and the fluent speakers with aphasia were clustered together with neurotypical speakers. This suggests that the selected methods can reliably identify speakers with non-fluent aphasia. The inspection of the two dimensional projection suggests that the x axis represents severity, while the y axis is most likely connected to productivity, particularly lexical diversity. The analysis also shows that there are several different language profiles with the fluency groups. These profiles correspond to patterns described in the literature for other, typologically different languages which is in accordance with Prins and Bastiaanse (2004) who conclude that the existing body of research has found no major language specific differences at this level of language production. The measures used in the present work successfully differentiated between non-fluent aphasia and neurotypical speech. The failure to detect reliable differences between speakers with fluent aphasia and neurotypical participants was, by and large, caused by a specific profile of participants aa3 and pa3 that are in some respects more similar to the non-fluent speakers aa2 and aa4. The measures could be used in clinical practice with relative ease as they do not require advanced knowledge of linguistic theory and their coding is relatively straightforward and does not necessarily require a detailed token level analysis.

4 Mean number of modifiers was excluded from the cluster analysis.

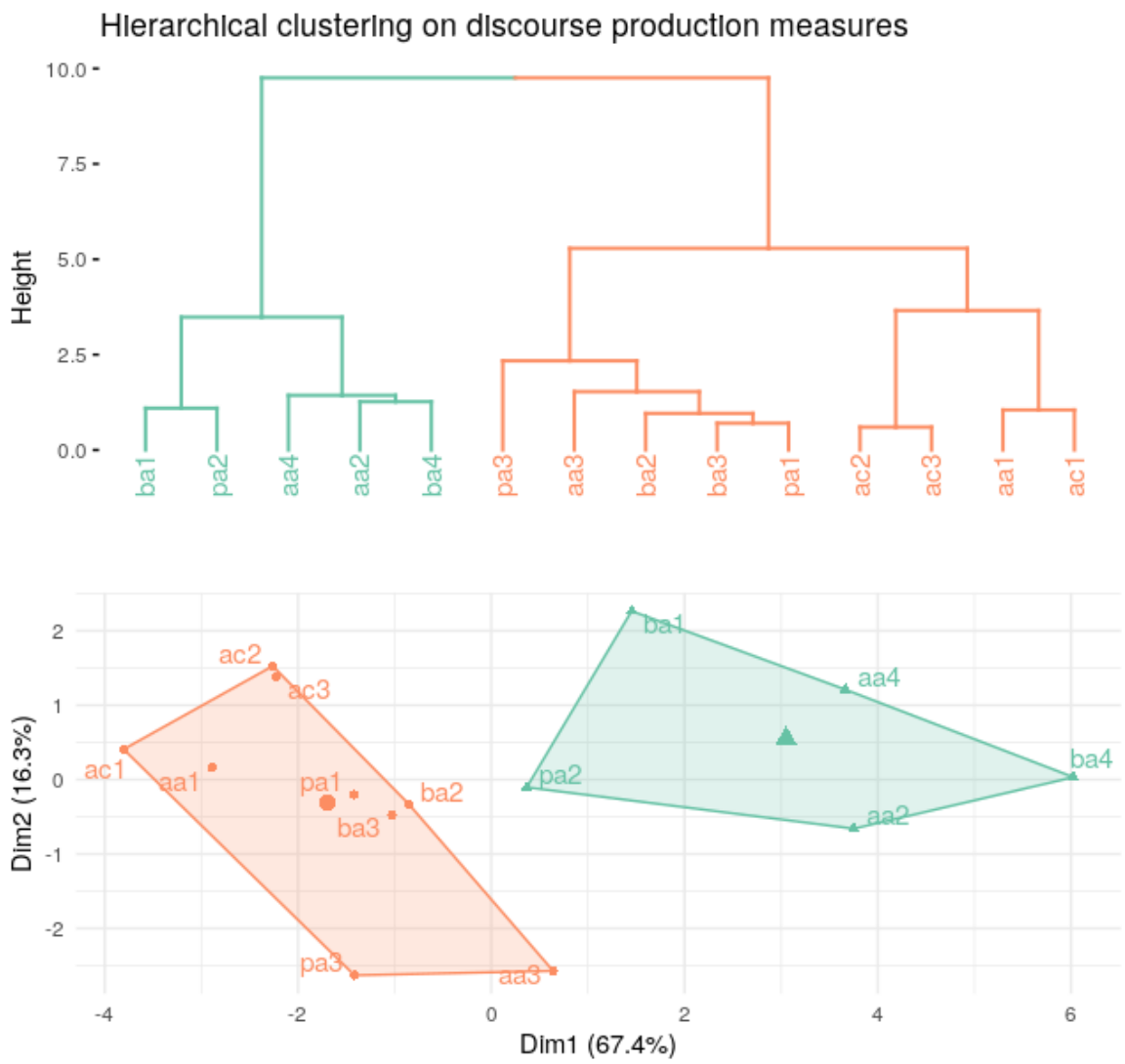

Figure 2: Cluster dendrogram of participant groups based on the production measures.

Verbs and nouns in the corpus

The first two analyses focused on the production of verbs and inflected nouns in the corpus. While verb processing has been a topic of major interest in linguistic aphasiology, nominal inflectional morphology has not received much attention, no less due to the fact that the bulk of research has been conducted on English, a morphologically impoverished language.

Verbs in the corpus

The study of verb production and processing has been one of the focal points in the linguistic study of aphasia. This can be explained by the fact that verbs and argument structure are at the core of the generative approaches which have by far been the most represented framework in linguistic aphasiology. Furthermore, it has long been established that speakers with non-fluent aphasia, particularly with Broca's aphasia produce discourses with fewer verbs, have difficulties in confrontational verb naming, and encounter problems in verb comprehension. Numerous frameworks and theories have

been proposed to explain these difficulties, such as the Argument Structure Complexity Hypothesis (Thompson 2003), the Trace Deletion Hypothesis (Bastiaanse & van Zonneveld 2005), or the adaptation strategy proposals (Hartsuiker & Kolk 1998). The aim of this section was to describe verb production in the subcorpus and to explore any potential differences between participants with different aphasia types and individual language profiles. I focus on variables that are derived from argument structure, such as transitivity, number of arguments, or grammatical voice which have been suggested in the literature to influence verb processing in non-fluent agrammatic aphasia. I was therefore interested to see if it was the case that some of the participants in the sample produced markedly fewer verbs falling within these categories, particularly the non-fluent participants aa2, aa4, and ba4 who presented with some traces of agrammatism during the interviews. Frequency of use is the other variable of interest in the present analysis. I explore not only the number and type of verbs produced by participants but also disfluencies that occurred in the context of verb retrieval in the data. The occurrence of disfluencies such as hesitation sounds or silent pauses was used as a marker of increased processing load during word finding (Goldman-Eisler 1964).

A total of 1291 tokens of lexical verbs (314 lemmas) were extracted from the subcorpus and further annotated to compare verb production of individual participants. The neurotypical speakers ac1, ac2, and ac3 were included in the analysis for reference. The verbs were annotated for the following variables: Hallidayan process types (Halliday & Matthiessen 2014: chap. 5), verb lemma (reflexive and non-reflexive uses counted as unique types, presence of modal and phase verbs, voice, number of obligatory arguments based on Vallex 4.0 (Lopatková et al. 2020), transitivity, semantic roles, presence and type of direct object in transitive verbs, expressed semantic roles, and cumulative lemma frequency (spoken and written corpora). Furthermore the presence of any disfluency preceding the verb (silent pauses, hesitations, word fragments and "intrusive" markers (*jako* 'like' or *teda* 'or, more precisely')) as well as length and type of disfluency were included in the annotation scheme.

The analysis of the various facets of argument structure did not reveal any strong influences of argument structure complexity. However, when transitive verbs were grouped together with passive and anticausative and unaccusative verb uses as per Bastiaanse and van Zonneveld (2005)'s analysis, the neurotypical group did produce relatively more verbs of this type, but no strong conclusions can be drawn from this finding. On the other hand, three results deserve a comment. While participants with non-fluent aphasia did not produce fewer transitive verbs, I found a weak correlation between transitivity and verb frequency such that the transitive verbs produced by these participants tend to be simultaneously high frequency verbs. This can be taken as support for employing the usage-based approach in aphasia research. Related to that, there was also a tendency for this group of speakers to produce more transitive verbs with omitted direct objects (cf. Figure 3). This is in accordance with the adaptation hypothesis that claims that participants with (non-fluent) aphasia omit non-essential (grammatical) information to decrease processing load (Hartsuiker & Kolk 1998). Finally, I found that participants with lower levels of fluency and lexical diversity tend to rely more on the use of the existential-presentative construction suggesting that this frequent, partially lexically filled construction is more accessible which is in accordance with the assumptions of the usaga-based model and in line with Hatchard (2021)'s findings concerning the reliance of persons with aphasia on lower level, less schematic constructions.

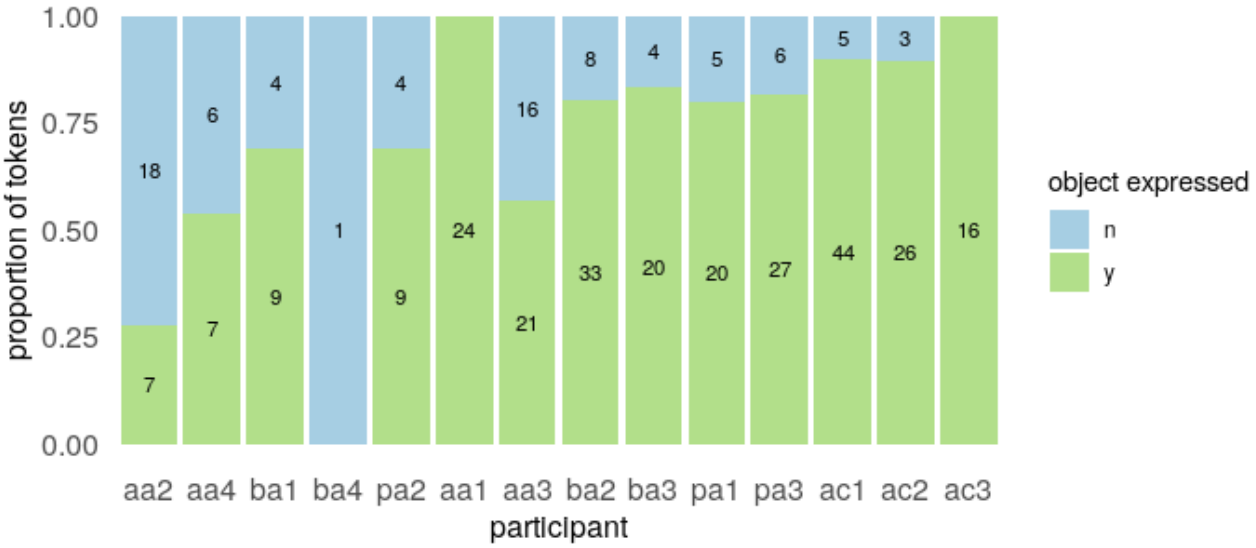

Figure 3: Proportion of direct object omission in transitive verbs; numbers in columns indicate token counts.

The analysis of disfluencies suggested that position in utterance, lemma frequency, and verb type rather than syntactic complexity explain the distribution of disfluencies in the sample (Figure 4). Specifically, I found that more fluent speakers produce less disfluencies in high frequency verbs. Non-fluent participants produce verbs more fluently when they occur closer toward the onset of the c-unit or when the verb falls in the referential process category. Note that referential predicates are mostly represented by *be* and *have* in the sample which are simlutaneously highly frequent. Example 1) shows a high frequency verb *spát* 'sleep' (fq (oral) = 382 pmw) produced fluently by a fluent participant (aa1) and a disfluent mental verb *zaujmout* 'draw attention' produced by participant with lower fluency level (ba2).

1) Factors in the fluency of verb production

   a) *a teď tam spal lev* 'and now there was a lion sleeping there' (aa1: 27)

   b) *chtěl 0.5 zaujmout tu 0.2 paní* '(he) wanted 0.5 to draw attention of the 0.2 lady' (ba2: 97)

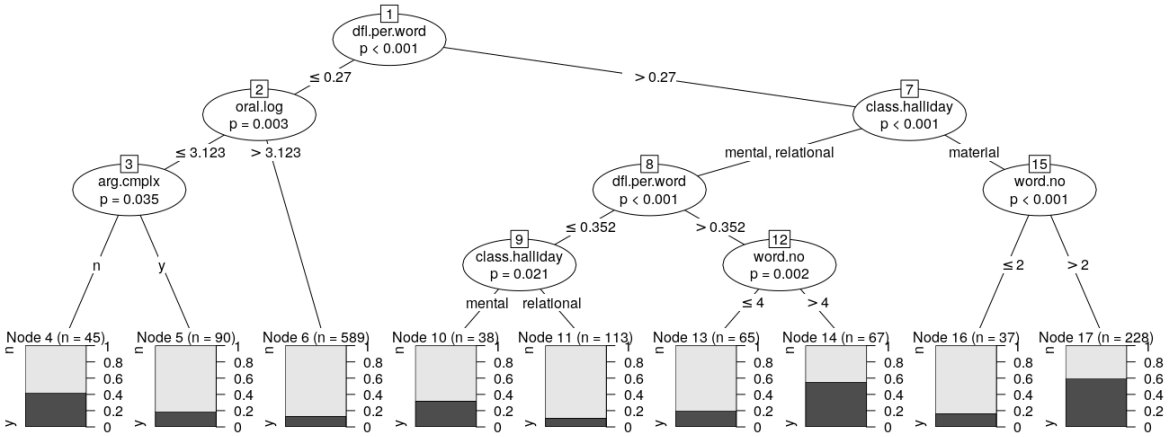

Figure 4: A classification tree of disfluent production of verbs.

Nouns in the corpus

Czech nominal morphology may be of particular interest to a usage-based study of aphasia, since some of the nouns have a high frequency of use of inflected word forms that are different from the citation form which may cause a tension between morphological complexity and frequency. The aim of this analysis was to explore the production of inflected forms by individual participants and their distributional characteristics. A very general prediction was formulated based on the usage-based assumptions. Participants with reduced levels of fluency were expected to produce a lower number of inflected forms and to produce a lower number of inflected word forms with low relative frequency. This analysis follows up on the findings of Lehečková (1986; 2001; 2009) who describes that Czech speakers with agrammatic and paragrammatic symptoms tend to use the nominative, i.e. the citation form, instead of other, context appropriate case form.

A total of 1126 fully formed generic noun tokens from the subcorpus was annotated was lemma, word form, and paradigmantic cell frequency and analyzed. Both the speakers with aphasia and the neurotypical participants were included in this analysis. Figure 5 shows that a subgroup of participants with aphasia (lower fluency and/or greater severity) indeed produced a markedly lower number of inflected word forms.

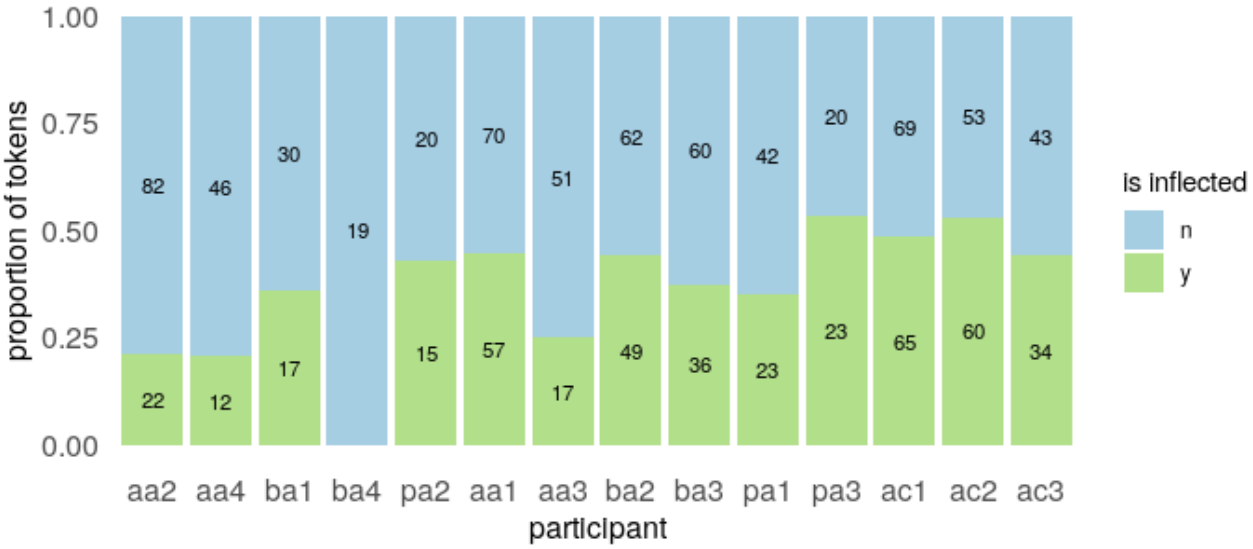

Figure 5: Proportion of inflected nouns; numbers in columns indicate token counts.

An analysis of number and case showed that this group produced a substantially higher proportion of the nominative case which is related to their having produced a higher number of isolated nouns and nouns in the existential-presentative construction. This suggest a qualitatively different production strategy that may be less demanding in terms of processing resources.

An analysis of the 425 inflected word forms revealed that participants with lower inflection rates in general produced tokens with high relative frequency that may be (parts of) fixed, formulaic expressions, such as the forms *pořádku* ‘order-sg.loc, *podstatě* ‘nature, essence-sg.loc,’, or *chvíli* ‘while-sg.acc’, the former two appearing almost exclusively in the PPs *v pořádku* ‘in order, all right’ and *v podstatě* ‘basically, in fact’, while the letter has a function of a temporal adverbial *(na) chvíli* ‘(for) a while’. These participants also produced a low number of inflected nouns with very low relative frequency. Furhtermore, when these participants produced rare forms it was in the context of particular constructions, such as the argument structure construction of escaping with the preposition *před*

'from, lit. in front of' and the tokens are found in close vicinity of each other, suggesting also the role of priming. Moreover, these low frequency inflected types also have arguably salient endings such as the dative/locative suffix *-ovi*, used for masculine animate referents. The ease of retrieval of the dative masculine animate forms may be further facilitated by the existence of the transfer argument structure construction in which recipients, prototypically sentient animate beings are marked by the dative.

Disfluencies as potential markers of chunkhood

The last analysis focused on disfluencies produced by participants with aphasia. First, I explored the placement of disfluencies within c-units. The main part of the analysis compared the occurrence and positions of disfluencies in prepositional phrases with the frequency characteristics of the noun complements and their cooccurrence with particular prepositions. Following up on Schneider (2016)'s study of the distribution of disfluencies in PPs in spoken American English, I expected that the PPs that occur more frequently will be produced more fluently. This expectation was based on the predictions of the usage-based model where frequently cooccurring items are treated as ready-made chunks stored in memory. Since disfluencies have been shown to index higher processing load (e.g. Fraundorf & Watson 2014), it follows that PPs that have a stronger association between a given preposition and a complement noun should be easier to retrieve and induce fewer disfluencies.

Overall distribution of disfluencies in the corpus

The placement of disfluencies within c-units was annotated with respect to clause and phrase boundaries. Participants produced comparable proportions of disfluencies at clause boundary positions that can be taken to mark the planning of the whole c-unit. However, two different patterns occurred with respect to phrase boundary. A subgroup of participants whose production was characterized with fewer adnominal and adverbal modifiers had a higher proportion of disfluencies at phrase boundaries relative to phrase-internal disfluencies, while participants that produced a higher number of such modifiers also produced a higher number of phrase-internal disfluencies. This pattern can be explained in terms of syntactic complexity such that c-units with more modifiers have more complex syntactic structure that is computationally more demanding and creates more room for disfluencies placed within phrases. Example 2.a) shows the production of participant ba1 who produced mostly bare phrases across c-units. We see silent pauses before each constituent. On the other hand, 2.b) represents the production of aa3 who produced longer c-units with greater inner complexity as evidenced here by the two adverbial modifiers *tam* 'there' and *pořád* 'still' and the demonstrative *toho*. The PP *u toho lva* 'with the lion' is produced with a pre-phrasal as well as a phrase-internal disfluency.

2) Pre-phrasal and phrase-internal disfluencies

   a) *ne 1 <tut> 1 kůň 0.5 honil 0.2 klauna* 'no 1 <tut> 1 the horse 0.5 was chasing 0.2 the clown' (ba1: 17)

   b) *no ale von z- tam zustal 0.2 u toho l- 0.2 lva <hes> 0.2 pořád 0.2* 'well but he s- stayed there 0.2 with the l- 0.2 lion <hes> still 0.2' (aa3: 194)

A comparison of contexts where the production was abandoned following a failed retrieval with contexts with reformulations showed an expected difference between fluent and non-fluent participant. Speakers with fluent aphasia are more likely to attempt a repair following word finding difficulties, whereas non-fluent participants have relatively lower retracing rate in these situations.

A closer look: Prepositional phrases

Prepositional phrases were selected for a closer analysis of disfluencies based on the existence of a similar study for neurotypical speech Schneider (2016) but also for practical reasons, since they are easily delimited and extracted from the corpus. Furthermore, PPs in Czech typically govern cases with word forms different from the citation form that may also have varying relative frequencies, as discussed in the previous section. Data from the whole corpus was used for this analysis, producing a total of 704 PPs with generic nominal complements (510 unique combinations of prepositions and complement word forms). These PPs were annotated for fluency. A distinction was also made between pre-phrasal and phrase-internal disfluencies, as shown in 3).

3) Disfluency placement within PPs

   a) fluent production: *pro pivo* ‘for beer’ (pa1: 125)

   b) pre-phrasal disfluency: <*hes*> *z: historie* ‘<hes> from: history’ (aa3: 6)

   c) phrase-internal disfluency: *s <hes> 1 nádobou* ‘with <hes> 1 a container’ (aa1: 152)

It was expected that the stronger the association between a preposition and a complement is the less likely a disfluency is to occur within the phrase. PPs were also annotated for presence of prenominal modifiers and for argumenthood, i.e. whether a given PP was an argument or an adjunct in a particular c-unit. Frequency data was obtained from three corpora of Czech, a corpus of written language, a spoken corpus, and a corpus of movie subtitles for the preposition, the complement noun, and their cooccurrence (forward and backward transitional probability). The frequency measures were combined into two different indices. First, the measures were submitted to hierarchical clustering on principal components. This method resulted in four clusters of PP types. Two of these clusters grouped PPs with high frequency complements and/or with strong associations such as *v kleci* ‘in a cage’ which has a high backward transitional probability (0.29) and *před rokem* ‘a year ago’ which has a high frequency complement (1361 pmw in the spoken corpus). These PPs were expected to be easier to retrieve and produced more fluently. A second index was derived directly from the frequency measures. Using the data from the subtitle corpus, a PP was marked as high frequency when the values of at least three of the five measures were higher than the third quartile. Similarly, PPs were coded as low frequency when three or more of the measures were lower than the first quartile.

The analysis did not find any relationship between disfluencies and argumenthood. A comparison of modified and unmodified PPs showed that modified PPs were produced with more disfluencies and that there was a higher proportion of phrase-internal disfluencies when a modifier was present (#figure). This was true for the whole sample as well as for individual participants.

When frequency was accounted for, there was a visible trend in the direction of more fluent production of PPs from the high frequency clusters both for the whole sample as well as for individual participants.[5] A higher proportion of these tokens was produced fluently and, when a disfluency did oc-

5 The picture is admittedly much less clear on individual level as some participants produced very few PPs overall.

cur, the proportion of pre-phrasal placement was higher for the high frequency PPs. This was true not only for the unmodified PPs but also for tokens that were used with a modifier (Figure 6). Moreover, when individual tokens in the "non-high frequency" clusters were compared using the high and low frequency "scores", additional variation in the data is explained, because tokens with high frequency scores are produced more fluently and with a lower proportion of phrase-internal disfluencies and, conversely, low frequency scores are associated with a higher proportion of disfluently produced PPs (Figure 7).

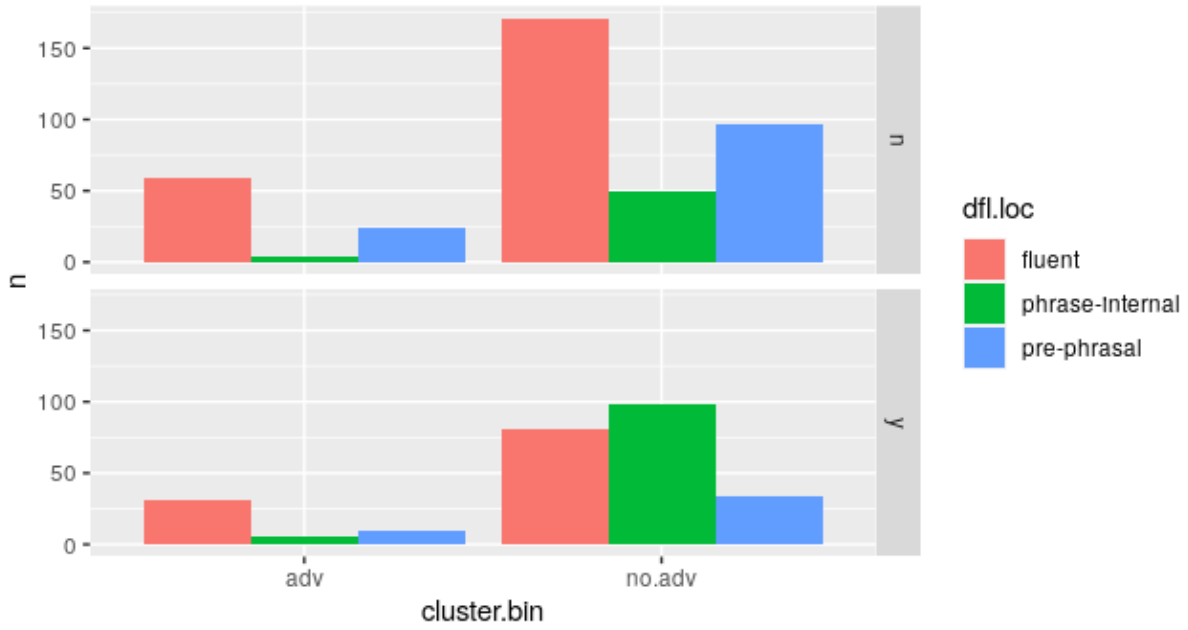

Figure 6: DIstribution of disfluencies in unmodified (upper panel) and modified (lower panel) PPs with respect to cluster membership (adv is the high frequency cluster, no.adv the non-high frequency cluster).

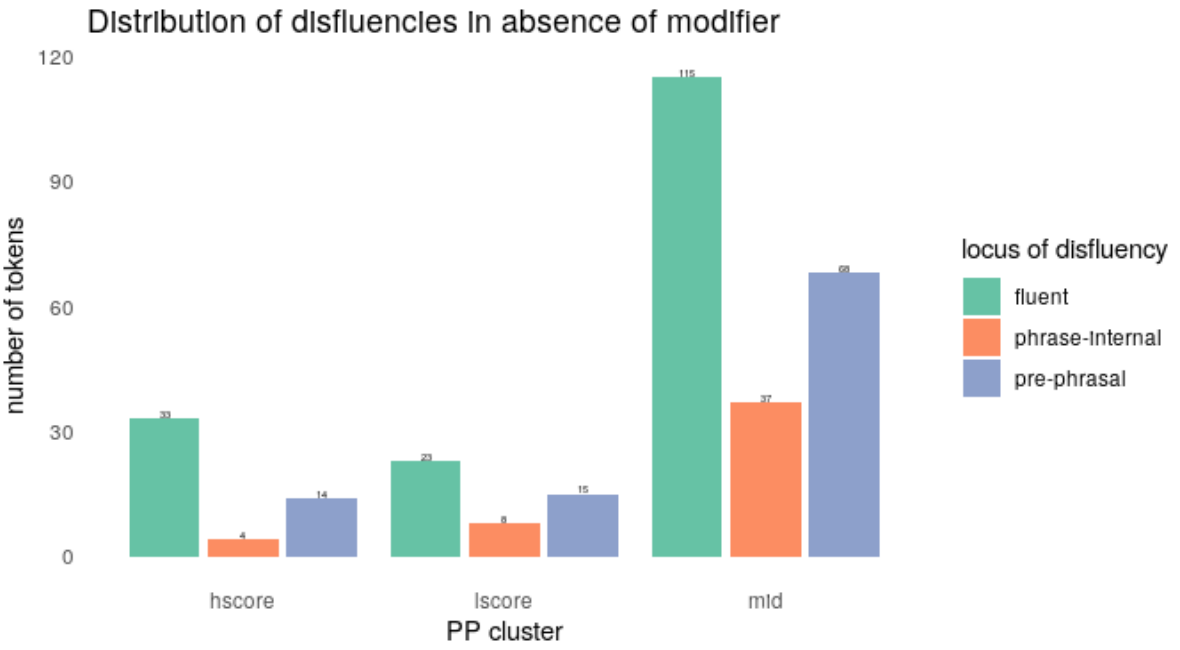

Figure 7: Distribution of disfluencies in unmodified PPs with respect to frequency scores.

Figure 8 proposes a simple model with four different types of PPs inspired by the frequency effects observed in the data. First, there are collocations with extremely high association measures, represented by *v pohodě* ‘alright, lit. in ease’ in the figure. These expressions can be assumed to be retrieved as chunks more akin to single word units leaving little space for intrusive disfluencies. Secondly, there are cases represented by *v kleci* ‘in a/the cage’. The lemma *klec* appears in approximately 30 % of all its uses in this PP. The high level of association between the preposition and the noun may also predict a direct retrieval of this bigram. However, the higher compositionality of this PP in comparison to *v pohodě* makes a disfluency slightly more likely to occur phrase-internally, as suggested by the dashed line in the figure. The phrase *k lvovi* ‘to a/the lion’ represents the situation where there is no strong association between the preposition and the complement. This is visualized by showing that the PP is composed of a slot for the preposition which is combined with an inflected noun. In more detail, we can conceptualize this such that there is a PP construction with a lexically specific preposition induced, e.g. by an argument structure construction of coming/arriving, and an open slot for a noun in the inflected form governed by the preposition. This PP construction is superimposed with the Inflected noun construction in language production and it is precisely this step that can be assumed to produce more phrase-internal disfluencies. The last schema with the example *s vařicí vodou* ‘with hot water’ shows how an additional step that concerns the combination of the complement noun with a prenominal modifier opens up more space for retrieval problems indexed by disfluencies.

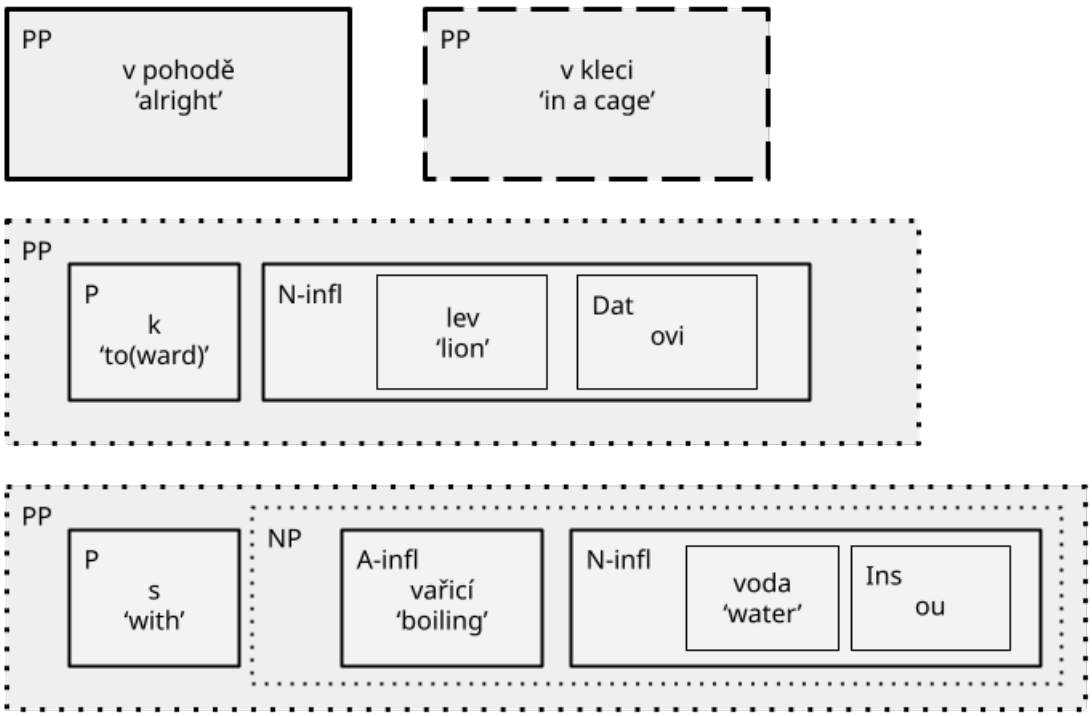

Figure 8: A model of PP production with respect to frequency

Conclusions

The present work had two main sets of aims. In the first part, I described the process of assembling a corpus of Czech aphasic speech. The corpus consists of samples of descriptive, narrative, procedural, and conversational discourse of eleven speakers with different aphasia types with the total

size of almost 17000 words produced by speakers with aphasia. Using the data from the corpus, I performed an analysis of aphasic discourse production. The majority of the microstructural measures selected for this analysis successfully differentiated between neurotypical speakers and participants with non-fluent aphasia. Furthermore, the combination of these measures used to visualize the structure of the sample with hierarchical clustering was able to identify subgroups of participants with similar language profiles. To follow up on these two outcomes, the corpus will be made available for researchers and clinicians.[6] A simple protocol containing the stimulus materials used to collect the data for the corpus as well as a transcription manual and an ELAN template file to enable other specialist to contribute their own data. The measures presented in this work may be converted into a tool for discourse analysis that could be used to help SLTs in assessment. However, the validation of the measures is needed on a larger sample.

The application of the usage-based framework on the corpus data and the results of the analyses presented in the second part have implications both for linguistics and for the study of aphasia. I hope to have shown the importance of various frequency measures in the analysis of language processing in aphasia. Cumulative lemma frequency was used for the analyses of verb production in the corpus. High frequency was found to increase the probability of fluent production. It was also apparent that persons with more severe lexical retrieval problems rely more on high frequency verbs. There was also a weak correlation between transitivity and frequency in the group of speakers with nonfluent aphasia, showing the importance to account for frequency and similar variables in the analysis of structural phenomena.

The results of the analyses of nominal inflection and disfluency placement in prepositional phrases is of particular interest both for linguistic theory and clinical practice. Both of these analyses suggest that the relative frequencies of particular inflected word forms are related to ease of access and, consequently, success of retrieval in aphasic language production. This is in line with previous research that has shown that the effects of probabilistic processing are present in both neurotypical speakers and persons with aphasia and the differences between the groups are qualitative rather than quantitative in nature. The fact that cooccurrence frequency and association strength between particular prepositions and nominal complements can be associated to fluent or disfluent production of prepositional phrases provides additional evidence for the usage-based model. Turning to clinical practice, the present work provides arguments for a tighter cooperation between linguists and clinicians and, in particular, for the introduction of corpus linguistic data in clinical practice. If we take the patterns observed in the analysis of nouns, these have the potential to be directly applied in SLT. When working with a patient in therapy on nominal morphology in Czech, a SLT can focus on specific high frequency word forms in typical usage contexts rather than using more standard protocols that focus on citation forms or whole paradigms.

In conclusion, I hope to have contributed to the growing body of literature that shows that the usage-based framework has potential benefits for the study of language in aphasia.The present study has several limitations and follow up studies with more participants and more controlled elicitation are necessary. However, my objective was to demonstrate the potential of the usage-based framework to discover interesting trends and patterns in the data that generate research questions and, if corroborated, have important clinical implications. It is precisely these new avenues of research that are subjectively the most important contribution of the present work.

6 A publication under the LINDAT/CLARIAH-cz infrastructure is planned in the future.

References

Armstrong, Elizabeth. 2000. Aphasic discourse analysis: The story so far. *Aphasiology* 14(9). 875–892. https://doi.org/10.1080/02687030050127685.

Barking, Marie, Ad Backus & Maria Mos. 2022. Individual corpus data predict variation in judgments: testing the usage-based nature of mental representations in a language transfer setting. *Cognitive Linguistics*. De Gruyter Mouton. https://doi.org/10.1515/cog-2021-0105.

Bastiaanse, Roelien & Ron van Zonneveld. 2005. Sentence production with verbs of alternating transitivity in agrammatic Broca's aphasia. *Journal of Neurolinguistics* 18(1). 57–66. https://doi.org/10.1016/j.jneuroling.2004.11.006.

Bates, Elizabeth & Brian MacWhinney. 1989. Functionalism and the competition model. In *The crosslinguistic study of sentence processing*, vol. 3, 73–112.

Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman & Tom Schoenemann. 2009. Language Is a Complex Adaptive System: Position Paper. *Language Learning* 59. 1–26. https://doi.org/10.1111/j.1467-9922.2009.00533.x.

Boles, Larry. 1998. Conversational discourse analysis as a method for evaluating progress in aphasia: a case report. *Journal of Communication Disorders* 31(3). 261–274. https://doi.org/10.1016/S0021-9924(98)00005-7.

Bruns, Claudia, Rosemary Varley, Vitor C. Zimmerer, Marcella Carragher, Gwen Brekelmans & Suzanne Beeke. 2019. "I don't know": a usage-based approach to familiar collocations in non-fluent aphasia. *Aphasiology* 33(2). 140–162. https://doi.org/10.1080/02687038.2018.1535692.

Bryant, Lucy, Alison Ferguson & Elizabeth Spencer. 2016. Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics & Phonetics* 30(7). 489–518. https://doi.org/10.3109/02699206.2016.1145740.

Bybee, Joan. 2010. *Language, Usage and Cognition.* Leiden: Cambridge University Press. http://public.eblib.com/choice/publicfullrecord.aspx?p=534754. (15 January, 2019).

Bybee, Joan L. 2007. *Frequency of use and the organization of language*. Oxford ; New York: Oxford University Press.

Croft, William A. 2001. *Radical Construction Grammar*. Oxford University Press.

Dąbrowska, Ewa. 2012. Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism* 2(3). 219–253. https://doi.org/10.1075/lab.2.3.01dab.

Damico, Jack S., Mary Oelschlaeger & Nina Simmons-Mackie. 1999. Qualitative methods in aphasia research: conversation analysis. *Aphasiology* 13(9–11). 667–679. https://doi.org/10.1080/026870399401777.

Dell, Gary S., Myrna F. Schwartz, Nadine Martin, Eleanor M. Saffran & Deborah A. Gagnon. 1997. Lexical access in aphasic and nonaphasic speakers. *Psychological Review* 104(4). 801–838. https://doi.org/10.1037/0033-295X.104.4.801.

Diessel, Holger. 2015. Usage-based construction grammar. In Ewa Dabrowska & Dagmar Divjak (eds.), *Handbook of Cognitive Linguistics*. Berlin, München, Boston: DE GRUYTER. https://doi.org/10.1515/9783110292022-015.

Diessel, Holger. 2020. A Dynamic Network Approach to the Study of Syntax. *Frontiers in Psychology*. Frontiers 11. https://doi.org/10/gh3s2g.

Divjak, Dagmar. 2019. *Frequency in Language: Memory, Attention and Learning*. 1st edn. Cambridge University Press. https://doi.org/10.1017/9781316084410.

Divjak, Dagmar, Ewa Dąbrowska & Antti Arppe. 2016. Machine Meets Man: Evaluating the Psychological Reality of Corpus-based Probabilistic Models. *Cognitive Linguistics* 27(1). 1–33. https://doi.org/10.1515/cog-2015-0101.

Fillmore, Charles J. 1988. The Mechanisms of "Construction Grammar." In *Proceedings of the*

Fourteenth Annual Meeting of the Berkeley Linguistics Society.
Flanderková, Eva, Barbara Mertins, Ondřej Bezdíček, Eva Baborová & M. Černá. 2014. Posuzování gramatičnosti v Brocově afázii - příklad dvou pacientů. *Česká a slovenská neurologie a neurochirurgie* 77/110(2). 202–209.
Fraundorf, Scott H. & Duane G. Watson. 2014. Alice's adventures in um-derland: psycholinguistic sources of variation in disfluency production. *Language, Cognition and Neuroscience* 29(9). 1083–1096. https://doi.org/10.1080/01690965.2013.832785.
Fried, Mirjam & Jan-Ola Östman (eds.). 2004. *Construction grammar in a cross-language perspective* (Constructional Approaches to Language v. 2). Amsterdam ; Philadelphia: John Benjamins Pub.
Gahl, Susanne. 2002. Lexical biases in aphasic sentence comprehension: An experimental and corpus linguistic study. *Aphasiology* 16(12). 1173–1198. https://doi.org/10.1080/02687030244000428.
Goldberg, Adele E. 1995. *Constructions: a construction grammar approach to argument structure* (Cognitive Theory of Language and Culture). Chicago: University of Chicago Press.
Goldberg, Adele E. 2019. *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton, NJ: Princeton University Press.
Goldman-Eisler, Frieda. 1964. Hesitation, Information, and Levels of Speech Production. In A. V .S. de Reuck & Maeve O'Connor (eds.), *Disorders of Language*, 96–114. J. & A. Churchill. https://doi.org/10.1002/9780470715321.ch6.
Halliday, M. A. K. & Christian M. I. M. Matthiessen. 2014. *Halliday's Introduction to Functional Grammar*. 4th edn. Routledge. https://www.jstor.org/stable/414947?origin=crossref. (15 January, 2019).
Hallowell, Brooke & Roberta Chapey. 2008. Introduction to Language Intervention Strategies in Adult Aphasia. In Roberta Chapey (ed.), *Language Intervention Strategies in Aphasia and Related Neurogenic Communication Disorders*, 3–19. 5th edn. Baltimore, MD: Lippincott Williams & Wilkins.
Hartsuiker, Robert J. & Herman H.J. Kolk. 1998. Syntactic Facilitation in Agrammatic Sentence Production. *Brain and Language* 62(2). 221–254. https://doi.org/10.1006/brln.1997.1905.
Hatchard, Rachel. 2021. *Cognitive Aphasiology – A Usage-Based Approach to Language in Aphasia* (Constructional Approaches to Language). Vol. 31. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/cal.31.
Hatchard, Rachel & Elena Lieven. 2019. Inflection of nouns for grammatical number in spoken narratives by people with aphasia: how glass slippers challenge the rule-based approach. *Language and Cognition* 11(3). 341–372. https://doi.org/10.1017/langcog.2019.21.
Hudousková, Andrea, Eva Flanderková, Barbara Mertins & Kristýna Tomšů. 2014. Hypotéza vymazávání stop a její platnost v češtině u pacientů s Brocovou afázií. *Slovo a slovesnost* 19.
Jakobson, Roman. 1980. *Child Language, Aphasia and Phonological Universals*. (Trans.) Allan R. Keiler. 3rd edn. The Hague: Mouton Publishers.
Lehečková, Helena. 1985. Využití lingvistiky při rehabilitaci afázií. *Československá otolaryngologie* 34. 116–120.
Lehečková, Helena. 1986. Agramatismus v afázii. *Slovo a slovesnost* 47(2). 138–147.
Lehečková, Helena. 2001. Manifestation of aphasic symptoms in Czech. *Journal of Neurolinguistics* 14(2–4). 179–208.
Lehečková, Helena. 2009. Afázie jako zdroj poznatků o fungování jazyka. *Slovo a slovesnost*. Ústav pro jazyk český Akademie věd České republiky 70(1). 23–35.
Lehečková, Helena. 2016. Afázie v lingvistice, lingvistika v afáziologii. *Časopis pro moderní filologii* 98(1). 7–22.
Linnik, Anastasia, Roelien Bastiaanse & Barbara Höhle. 2016. Discourse production in aphasia: a current review of theoretical and methodological challenges. *Aphasiology* 30(7). 765–800. https://doi.org/10.1080/02687038.2015.1113489.

Loban, Walter. 1966. *Language Ability. Grades Seven, Eight, and Nine*. U.S. Department of Health, Education, and Welfare.

Lopatková, Markéta, V Kettnerová, A Vernerová, E Bejček & Z Žabokrtský. 2020. VALLEX 4.0. LINDAT/CLARIAH-CZ Digital Library at the Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University. http://hdl.handle.net/11234/1-3524. (15 August, 2022).

Marini, Andrea, Sara Andreetta, Silvana del Tin & Sergio Carlomagno. 2011. A multi-level approach to the analysis of narrative language in aphasia. *Aphasiology* 25(11). 1372–1392. https://doi.org/10.1080/02687038.2011.584690.

Perkins, Lisa, Jenni Crisp & David Walshaw. 1999. Exploring conversation analysis as an assessment tool for aphasia: the issue of reliability. *Aphasiology* 13(4–5). 259–281. https://doi.org/10.1080/026870399402091.

Presson, Nora & Brian MacWhinney. 2011. The Competition Model and Language Disorders. In *The Handbook of Psycholinguistic and Cognitive Processes*. Routledge. https://doi.org/10.4324/9780203848005.ch2.

Prins, Ronald & Roelien Bastiaanse. 2004. Analysing the spontaneous speech of aphasic speakers. Review. *Aphasiology* 18(12). 1075–1091. https://doi.org/10.1080/02687030444000534.

Schneider, Ulrike. 2016. Hesitation placement as evidence for chunking: A corpus-based study of spoken English. In *Experience Counts: Frequency Effects in Language*, 61–90. De Gruyter. https://doi.org/10.1515/9783110346916-004.

Straková, Jana, Milan Straka & Jan Hajič. 2014. Open-Source Tools for Morphology, Lemmatization, POS Tagging and Named Entity Recognition. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 13–18. Baltimore, Maryland: Association for Computational Linguistics. https://doi.org/10.3115/v1/P14-5003.

Taylor, John R. 2012. *The mental corpus: how language is represented in the mind*. Oxford ; New York: Oxford University Press.

Thompson, Cynthia K. 2003. Unaccusative verb production in agrammatic aphasia: the argument structure complexity hypothesis. *Journal of Neurolinguistics* 16(2–3). 151–167. https://doi.org/10.1016/S0911-6044(02)00014-3.

Tummers, Jose, Kris Heylen & Dirk Geeraerts. 2005. Usage-based approaches in Cognitive Linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory*. De Gruyter Mouton 1(2). 225–261. https://doi.org/10/b5fbj5.

Verhagen, Véronique. 2020. *Illuminating variation: Individual differences in entrenchment of multi-word units*. Tilburg: Tilburg University Dissertation. https://www.lotpublications.nl/illuminating-variation-individual-differences-in-entrenchment-of-multi-word-units. (11 February, 2020).

Wilkinson, Ray, Karen Bryan, Sarah Lock, Kate Bayley, Jane Maxim, Carolyn Bruce, Anne Edmundson & Diana Moir. 1998. Therapy Using Conversation Analysis: Helping Couples adapt to Aphasia in Conversation. *International Journal of Language & Communication Disorders* 33(S1). 144–149. https://doi.org/10.3109/13682829809179412.

Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann & Han Sloetjes. 2006. ELAN: a Professional Framework for Multimodality Research. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. Genoa, Italy: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2006/pdf/153_pdf.pdf. (24 August, 2022).

Zimmerer, V. C., L. Newman, R. Thomson, M. Coleman & R. A. Varley. 2018. Automated analysis of language production in aphasia and right-hemisphere damage: frequency and collocation strength. *Aphasiology* 32(11). 1267–1283. https://doi.org/10.1080/02687038.2018.1497138.

Appendix: Academic CV

Work and teaching experience

- 2016-2019 Department of Middle Eastern Studies, assistant lecturer in Turkish studies
- 2019-present Department of Linguistics, assistant lecturer in General linguistics
- teaching experience
    - Úvod do lingvistiky
    - Úvod do lingvistické analýzy
    - Jazyk a kognice
    - Lexikální sémantika
    - Turečtina pro lingvisty
    - Gramatický systém turečtiny
    - Četba starších a současných tureckých textů
    - Příprava úloh pro Českou lingvistickou olympiádu
    - Bakalářský seminář (v programu Obecná lingvistika)
    - Seminář lingvistické turkologie
    - Rozšířený seminář k úvodu do lingvistiky
    - Syntax turečtiny
    - Morfologie a lexikologie turečtiny
    - Praktická jazyková cvičení (turečtina)
    - Četba a gramatická analýza tureckých textů
    - Jazykový systém turečtiny
    - Vybraná témata z turkologické sociolingvistiky
    - Aphasia: an introduction from linguistic perspective

Supervised undergraduate theses

- Role ikoničnosti v porozumění vedlejším větám časovým: příprava protokolu pro mluvčí s afázií (*The role of iconicity in the comprehension of temporal adverbial clauses: preparation of an aphasia research protocol*)
- Sémantika aktantů a relativní frekvence v porozumění pasivní konstrukci u osob s afázií (*Argument semantics and relative frequency in the comprehension of the passive construction in persons with aphasia*)
- Reflexive alternation in language processing by persons with aphasia
- Ikoničnost a časová souslednost: experiment s umělým jazykem (*Temporal sequence iconicity: An artificial language learning experiment*)
- Distribuce osmanských a Öztürkçe synonym v současné turečtině (*The Distribution of Ottoman and Öztürkçe Synonyms in Contemporary Turkish*)

Grants awarded in connection to teaching

- Creation of the course *Aphasia: an introduction from a linguistic perspective*; awarded by Charles University. 2016-2017.
- Creation of the course *Příprava úloh pro Českou lingvistickou olympiádu (Creation of contest problems for Czech linguistic olympiad)*; awarded by Charles University 2015.

Articles in peer reviewed publications

- Křivan, J & Láznička, M. forthcoming. "Inalienable Possession", in: Encyclopedia of Slavic Languages and Linguistics Online, Editor-in-Chief Marc L. Greenberg.
- Láznička, M. 2020. Lingvistické úlohy jako prostředek poznávání lingvistické analýzy a jazykové rozmanitosti. *Cizí jazyky*, 63(3): 28-38. Preprint.
- Láznička, M. 2019. Možnosti uplatnění lingvistické teorie v afaziologii na příkladu usage-based lingvistiky. *Speciální pedagogika*, 29(3): 220-237. Preprint.
- Křivan, J & Láznička, M. 2018. Diachronní korpusová analýza: slovosled českých posesivních adjektiv uvnitř nominální fráze. *Studie z aplikované lingvistiky/Studies in Applied Linguistics*, 9 (special issue): 42–65. Published version.
- Láznička, M. 2016. Příprava korpusu afatické češtiny. *Studie z aplikované lingvistiky/Studies in Applied Linguistics*, 7: 184–187.
- Láznička, M. 2014. Paul Broca: Poznámky k sídlu schopnosti mluveného jazyka následované pozorováním afémie (ztráty řeči) (translation). *Studie z aplikované lingvistiky/Studies in Applied Linguistics*, 2: 119–135.

Non-scientific articles

- Láznička, M. 2020. Když se pletou psi a osli. Vesmír, 99: 296-297.

Conference presentations

- 2022. *Effects of frequency in the distribution of disfluencies in language production of Czech speakers with aphasia*. AMLaP, York.
- 2021. *The role of relative frequency in the production of prepositional phrases in aphasia in Czech*. Academy of Aphasia 59th Annual Meeting.
- 2021. *Distribution of pauses in discourse production of Czech speakers with aphasia*. The Slavic Cognitive Linguistics Conference, Tromsø.
- 2019. *Grammatical profiling of Czech nouns: what do cases tell us about nouns' meaning.* The 15th International Cognitive Linguistics Conference, Nishinomiya. (with V. Janda)
- 2019. *Grammatical profiles of Czech nouns: case and gender*. The 7th Conference of the Scandinavian Association for Language and Cognition, Aarhus.
- 2018. *Frequency effects in Czech case homonymy*. 4th Usage-based Linguistics Conference, Tel Aviv.
- 2017. *Case homonymy in Czech: corpus data and sentence production*. The 14th Interna-

tional Cognitive Linguistics Conference, Tartu,

- 2017. *Cognitive biases and individual constructions: the case of Czech possessive adjectives*. The 14th International Cognitive Linguistics Conference, Tartu. (with J. Křivan)
- 2017. *The ordering of main and temporal adverbial clauses in Czech*. The 3rd Usage-based Linguistics conference, Jerusalem.
- 2017. *Iconicity and syntacticity in Czech temporal adverbial clauses*. 11th International Symposium on Iconicity in Language and Literature, Brighton.
- 2016. *A corpus of Czech aphasic speech: development and possible applications*. Poznan Linguistic Meeting, Poznan.
- 2016. *Functional explanations of prenominal possessors in attributive possessive constructions: Dynamic evidence from Czech*. 49th Annual Meeting of the Societas Linguistica Europaea, Naples. (with J. Křivane & E. Lehečková)
- 2016. *Frequency-based grammatical profiles of Czech nouns*. 2nd Usage-based linguistics conference, Tel Aviv. (with E. Lehečková & V. Janda)
- 2015. *Building a corpus of aphasic Czech*. IV International Congress of Clinical Linguistics, Barcelona.
- 2015. *Investigating distance iconicity in alienability marking: an artificial language learning study*. 13th International Cognitive Linguistics Conference, Newcastle.
- 2015. *Effects of tense iconicity in sentence processing*. MiLanguage language spring school, Milan.
- 2015. *Distance Iconicity in Possessive Constructions: An Artificial Language Learning Study*. 10th International Symposium on Iconicity in Language and Literature, Tübingen.

Summer schools

- 2019. *Corpus Linguistics Summer School 2019*, Birmingham.
- 2019. *Birmingham Statistics for Linguistics Summer School 2019*, Birmingham.
- 2017. *The first summer school on statistical methods for linguistics and psychology*, Potsdam.
- 2015. *MiLanguage language spring school*, Milan.
- 2012. *Türkçe Yaz Okulu (Turkish language summer school),* Ankara – Edirne – Istanbul.

Research grants

- Principal investigator, *Příprava korpusu afatické řeči (Creating a corpus of aphasic speech)*; awarded by Charles University. 2015.

Conference organizing

- Linguistics Prague 2017, 2018, and 2019, 2022, co-organizer. (conference website)

Research group membership

- Empirical Perspectives on Communication and Cognition (website)

- Experimental research on Central European languages (website)

Other activities

- Co-organizer and secretary of the Czech linguistics olympiad; co-organizer of International Linguistics Olympiad 2018 in Prague
- Member of a science communication and public outreach group Library of Languages, co-organized of the festival *Evropský den jazyků*