

**Charles University in Prague**

**Faculty of Arts**

Department of Linguistics

**BACHELOR THESIS**

Daniel Suchý

**Reflexive alternation in language  
processing by persons with aphasia**

Reflexivní alternace ve zpracování jazyka u lidí s afázií

I would like to thank my thesis advisor Michal Láznička, who guided me through the whole process of writing the thesis and provided the majority of participants needed for the experiments. I would also like to thank my parents for helping with my decision-making and for supporting me financially during my studies.

I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources.

In Prague on April 30th, 2020

Daniel Suchý

**Klíčová slova:**

reflexivní alternace, získaná afázie, agramatismus, relativní frekvence, zpracování jazyka, korpusová lingvistika

**Keywords:**

alternation, acquired aphasia, agrammatism, relative frequency, language processing, corpus linguistics

**Abstract:**

The goal of this thesis is to design an experiment that researches how frequency of use and grammatical structure affect sentence comprehension in speakers with aphasia. An additional objective is to verify this design in an experiment on healthy speakers. Introduction to the topic of aphasia is provided from a linguistic perspective, followed by a description of usage-based approach to aphasiology, which will be utilised in the practical part of this thesis. The practical part of the thesis consists of a corpus-based research on the use of reflexive constructions in Czech, followed by an experiment that tests comprehension of such sentences in healthy speakers. Using corpus data, a set of experimental items that can be used to test the comprehension of reflexive sentences was created. The experiment verified that the healthy participants evaluate the grammaticality of these sentences as expected in the vast majority of cases. The results of the experiment show a weak effect of sentence structure and no effect of frequency of use and verb lemma on comprehension. This indicates that healthy speakers find these sentences easy to process and that the design is valid. Based on these results, a design of an experiment that tests reflexive sentence comprehension in Czech speakers with aphasia is proposed.

**Abstrakt:**

Cílem této práce je vytvořit dizajn experimentu, který zkoumá vliv frekvence užití a typu gramatické struktury na porozumění větám s reflexivním slovesem u mluvčích s afázií. Dále je cílem ověřit tento dizajn pomocí experimentu na zdravých mluvčích. Téma afázie je zde představené z lingvistického pohledu. Práce dále obsahuje popis kontextového přístupu k afaziologii, který bude použitý v praktické části práce. Praktická část této práce se skládá z korpusové studie, která zkoumá užívání reflexivních konstrukcí v češtině a na základě které byl proveden experiment testující porozumění větám s reflexivním slovesem u zdravých mluvčích. Experiment ověřuje, že zdravý mluvčí v naprosté většině případů hodnotí gramatičnost těchto vět tak, jak se očekávalo před zahájením experimentu. Výsledky experimentu odhalují, že na porozumění těmto větám má syntaktická struktura slabý vliv, zatímco frekvence užití a užitá slovesná lemma nemá vliv žádný. Závěrem experimentu je, že pro zdravé mluvčí je porozumění těmto větám jednoduché a dizajn experimentu je validní. Na závěr je v práci představen návrh experimentu zkoumajícího zpracování reflexivních vět i u českých mluvčích s afázií.

# Contents

<b>1</b>	<b>Introduction</b>	<b>6</b>
1.1	Goals of this thesis . . . . .	6
1.2	Brief introduction to aphasia . . . . .	7
1.3	Motivation for aphasia research . . . . .	7
1.4	Approaches to aphasia treatment . . . . .	8
1.5	Linguistic research of aphasia . . . . .	9
1.6	Usage-based approach to linguistics . . . . .	10
1.7	Frequency effects in aphasia . . . . .	11
1.8	Agrammatism . . . . .	12
1.9	Reflexive verbs in Czech language . . . . .	14
1.9.1	Other functions of the reflexive pronoun . . . . .	15
1.9.2	Inherently reflexive verbs . . . . .	16
<b>2</b>	<b>Method</b>	<b>17</b>
2.1	Corpus research . . . . .	17
2.1.1	The corpus . . . . .	17
2.1.2	Data extraction . . . . .	19
2.2	Experimental items design . . . . .	21
2.3	Verification of stimulus sentences (experiment I) . . . . .	22
2.3.1	Results . . . . .	23
2.4	Verification of experiment design (experiment II) . . . . .	24
<b>3</b>	<b>Results</b>	<b>26</b>
<b>4</b>	<b>Conclusion</b>	<b>30</b>
4.1	Discussion . . . . .	30
4.2	Concluding summary . . . . .	31
	<b>Bibliography</b>	<b>34</b>
<b>A</b>	<b>List of queries used in the corpus research</b>	<b>38</b>
<b>B</b>	<b>Preliminary experiment sentences</b>	<b>39</b>
	<b>List of Figures</b>	<b>41</b>
	<b>List of Tables</b>	<b>42</b>

# Chapter 1

## Introduction

### 1.1 Goals of this thesis

The goal of this thesis is to explore the topic of syntactic processing difficulty in aphasia. Most speakers with aphasia experience an increased difficulty of comprehension of both spoken and written language. This topic is subject of extensive interdisciplinary research, which aims to discover which linguistic and non-linguistic factors influence the difficulty of syntactic processing the most. One of the factors that play a role in sentence processing is the frequency of use. Frequently used words or grammatical constructions are easier to comprehend. This has been well documented by a large body of research on healthy speakers and to a lesser extent on speakers with aphasia. So far, the focus with aphasia has been mostly on single words so far, but research on combinations of words or syntactic constructions is also starting to emerge. It is also known that aphasia can manifest itself in many ways, depending on the grammatical system of the language that the speaker uses. Therefore, a variety of speakers coming from different language backgrounds should be studied. This area of research on speakers of Czech is starting to develop (e.g. Flanderková (2019)), however not much is known so far. Therefore, in this thesis, I aim to explore the topic of reflexive syntactic constructions in Czech and measure processing difficulty of these constructions for healthy native Czech speakers, based on the frequency of occurrence of these constructions. In addition, I aim to design an experiment on the difficulty of processing reflexive sentences in Czech aphasic speakers.

This chapter provides a short theoretical overview of the topic of aphasia and the topic of syntactic processing in both healthy speakers and speakers with aphasia. First, I provide a very brief introduction to aphasia, with a focus on linguistic research of this topic. I discuss how aphasia research can be enhanced by a linguistic viewpoint and how linguistic knowledge can help in aphasia classification, diagnostics and therapy. I discuss the most relevant symptom of aphasia for this thesis - agrammatism, and how it is influenced by grammatical systems of different languages. Next, I continue with an introduction to usage based approach to psycholinguistics and aphasiology, an approach that focuses primarily on how frequency of use influences language comprehension. In the last section, I explain the function of reflexive constructions in the Czech language.

This theoretical insight in diverse areas is needed to conduct the practical part of this thesis, which is covered by the remaining three chapters. In chapter

two, I discuss the steps that I took when researching reflexive constructions use in Czech. I also discuss the preparation of an experiment that measures the processing difficulty of these constructions in healthy speakers. The results of this experiment are presented in chapter 3. The last chapter discusses limitations of this experiment, proposes improvements, and discusses how this experiment can be modified to measure processing difficulty in speakers with aphasia.

## 1.2 Brief introduction to aphasia

Aphasia is an acquired neurogenic language impairment, caused by a lesion in certain areas of the brain. It is one of the most common impairments resulting from an ischemic or hemorrhagic stroke. The median prevalence of aphasia after stroke is estimated to be 30 percent of all stroke survivors, with frequencies ranging from 15 to 42 percent, depending on evaluation methodologies (Flowers et al., 2016). Its typical symptoms include *anomia* (inability to recall words), *agrammatism* (impairment in use and comprehension of syntax and morphology), difficulty in language comprehension and in some cases slow and effortful speech. Depending on the type of the lesion, a person with aphasia will have problems with expressing themselves, understanding others, or both. The exact definition of aphasia varies from author to author and has changed over time. Papathanasiou and Coppens (2017) define aphasia using the following criteria:

- it is primarily a language-level problem
- it includes receptive and expressive components
- it is multimodal in nature
- it is caused by a central nervous system dysfunction

Aphasia therefore differs from other impairments, such as dementia, in that it is language-specific. Other cognitive capabilities (e.g. memory) remain intact. It affects language capabilities as a whole and is not restricted to perception, production, spoken, written or signed language. Depending on the type of lesion it can be primarily expressive or receptive, but both components are typically affected, especially since anomia (inability to recall words and their meanings) is present in all types of aphasia.

Other authors, such as Code (2010) define aphasia differently, including also impairments that result from progressive disorders, such as dementia. Even then, it is important to differentiate between aphasia and motor impairments, such as dysarthria or aphonia.

## 1.3 Motivation for aphasia research

Many patients diagnosed with aphasia recover within weeks to months post-stroke. However, many other suffer from long-term symptoms (Laska et al., 2001). Aphasia is associated with lowered quality of life, that comes with a communicative impairment. Patients report lowered quality of life in areas of daily and work activities, social relationships and opportunities of learning and acquiring new



information (Ross and Wertz, 2003). Aphasia is also associated with higher risk of mortality. Fortunately, therapy is now widely considered beneficial for chronic patients (Watila and Balarabe, 2015). This often means speech therapy sessions of varying length and intensity. Other alternative methods of treatment such as transcranial magnetic stimulation (Ren et al., 2014) or drug treatments (Bakheit, 2004) are also being researched and show promising results.

Multidisciplinary research is needed in order to discover which factors contribute to higher risk of chronic aphasia. It is also needed to determine which forms of therapy are efficient and how they can be improved. Research can be conducted from a neurological or psychosocial perspective. Linguistic research of aphasia, focusing on diagnosis and therapy, is also considered beneficial. Moreover, aphasia is also interesting from a theoretical perspective. It can help us understand what areas of the brain are responsible for language faculties and how they function.

## 1.4 Approaches to aphasia treatment

As mentioned in the sections above, the main goal of scientific research of aphasia is to find out how to treat the disorder, what to focus on during the treatment and find out what methods of treatment provide tangible benefits for the impaired speakers. Two prominent approaches have emerged during the history of aphasia treatment. The older, but still popular approach is called the *impairment based method*. This method is based on the fact, that an experience, such as a therapy session, directly affects the structure of the brain. This process, known as *neuroplasticity* has been discovered thanks to the advances in neuroimaging technology (Martin et al., 2008). The impairment based method emphasizes that aphasia is primarily a language disorder and the therapy thus focuses on treating the specific deficits in a patient's production and comprehension. Therefore, an accurate description of aphasic language comprehension is needed in order to determine what elements cause the most difficulty. Therefore, in this thesis, I will attempt to design an experiment protocol, which will enable an analysis of a particular element of Czech grammar.

The other method, called the *consequences approach* focuses on the fact that aphasia is a functional (social) disorder. It focuses on practical language use in everyday communication and it tries to provide a better accountability for language therapy, give a freedom of choice to patients and help not just the patient, but also their family or other affected people (Martin et al., 2008). This approach works well together with a more general *usage based* approach in linguistics, as both focus on the fact that language use always takes place in a specific context and language competence is thus shaped by this context (Gahl and Menn, 2016). Therefore, in this thesis I will attempt to address the relationship between context (in this case frequency of use) and a grammatical expression.

The conclusion is that aphasia is both a language impairment and a social impairment. Both approaches can and should be used in a successful therapy and neither of the two aspects of aphasia can be ignored.

## 1.5 Linguistic research of aphasia

As mentioned in the previous section, aphasia can be viewed from different angles. Now, I will move on to discuss the linguistic approach to aphasia and how it can help us to better deal with the disorder. During the first years of its history, research in aphasiology was viewed from a physiologic or psychiatric point of view. Linguistics became more integrated in the first half of the 20th century (e.g. Pick, Jakobson), but especially with the development of psycholinguistics in the 1960s (Papathanasiou and Coppens, 2017, p. 25). Today, the linguistic approach is considered necessary to obtain a full picture of the impairment. Lehečková (2016) provides an overview of the benefits of linguistic approach to the diagnosis and treatment of aphasia:

- In what way is a patient's use of language pathological?
- How should therapists plan out and conduct language therapy?
- How are aphasia symptoms in different languages related to each other?

Linguistics provides an apparatus that enables us to accurately describe a speech impairment, not only in general, but also for a particular individual. This is important if researchers want to design and evaluate test batteries. Determining what constitutes a standard, grammatical expression in language is a requirement, but in reality it is not always straightforward. Even healthy individuals differ considerably in their speech production and their evaluations of what is considered 'correct'. This needs to be taken into account when designing the tests, because a deviation from the norm is not always pathological. In addition, it is a requirement that the test needs to be quick and easy to administer, while maintaining the depth and quality that is needed to correctly evaluate all aspects of patient's language.

When a person is diagnosed with aphasia, speech therapy should follow. This is primarily a task for a speech pathologist, but a linguistic approach is also beneficial when designing effective speech therapy and evaluating its effectiveness. Lehečková (2016) mentions several uses of linguistic theories in speech therapy, such as creating clinically relevant aphasia models or research in universal symptoms of aphasia.

The last area of contribution of linguistics in aphasiology is research of aphasia typology and agrammatism. Agrammatism (as described in section 1.8) can be simply viewed as an incorrect use of syntax. However, it was sufficiently defined only after researchers started to describe aphasia symptoms using a precise linguistic terminology. In isolating languages such as English, agrammatism may simply be viewed as a syntactic disorder but research in a variety of typologically different languages has shown that the issue is more complex. In the international research project *Agrammatic Aphasia: A Cross-language Narrative Sourcebook* that compares aphasic data in 14 structurally different languages, the authors conclude, that aphasic speech shares similarities across all these languages, but also varies according to the particular grammar of the language (Menn and Obler, 1989).

Therefore, one of the goals of this thesis is to explore the effects of agrammatism on aphasic comprehension in Czech, a synthetic language with free word

order. Then the results can be compared with similar structures in other languages. The manifestations of agrammatism in Czech were already described e.g. in Lehečková (2001) and later in Flanderková (2019) and in this theses I will attempt to expand them by designing a way to study the effects of a practical grammatical constructions.

Finally, when discussing the relationship between linguistics and aphasiology, one cannot omit the benefit that the research of aphasia has brought to the theories in general linguistics of healthy language. The data obtained from patients with aphasia has proved useful when reserching theoretical questions such as language universals (Lehečková, 2016) and can help cognitively oriented grammarians decide between different modes of analysis. For example, Comrie (1989) analyses the topic of *Word Order in the Germanic Languages* with a help from this type of data. Here, the author proposes two hypotheses for the word order. The first is that the verb is always found as a second constituent in a sentence. The second, more complex hypothesis has two parts: subjects occur sentence-initially, immediately followed by the verb, and one constituent may be preposed, in which case the subject must be positioned after the verb. The speakers with aphasia in Comrie's sample do not follow the simpler hypothesis - the verb is not always the second constituent in their utterances. However, they follow the second, complex hypothesis more closely - they follow its first premise (prepose additional constituents before the verb), but not always the second one (place the subject after the verb). This may indicate that one of the rules was left intact, while the other has been wiped out by aphasia. Therefore, the author argues, that the simple hypothesis should be rejected, and the more complex one is true even for healthy speakers.

## 1.6 Usage-based approach to linguistics

Many different models, methodologies and frameworks have been developed in order to better conceptualize the complexity of both healthy language and its impairments. In this thesis, I will utilize the methodology of *usage based approach* to language and grammar. In this methodology, language is viewed as a tool, that emerges from the practical human need of communication. All people share certain general cognitive abilities (such as ability to categorize and memorize) and a common need to cooperate with each other. These abilities, when applied in a social environment, result in an emergence of an abstract and complex set of symbols, a natural language (Beckner et al., 2009).

According to the usage based approach, the most important factor in development of grammar (and therefore any natural language) is *frequency* (Diessel, 2007). Certain elements of language, such as words, but also larger constructions are more frequent than others. Speakers who encounter them frequently can predict when these expressions occur and thus they pay less attention to them. They may even reduce their phonological form or merge them with other words. This makes effortless language processing possible and leads to the emergence of complex grammar. Speakers do not have to pay close attention to each and every word and can therefore focus on complex meanings of the whole utterances. As a result, however, infrequent types of expressions are harder to process. Speakers need more time to process them and they are more likely to misunderstand them.

This tendency may be increased in people with aphasia. The general mechanism is likely the same, but the ability to use frequency as a hint in language processing is lowered in people with aphasia. As a result, these speakers may need longer reaction times to process the same information and may be more error-prone when processing infrequent grammatical constructions. If this effect is significant, they may have to rely on simple words and constructions, which makes expressing complex concepts difficult. The effect of frequency on language processing in aphasia remains an open research problem.

In my experiment, I will attempt to explore the relationship between probability of occurrence of a certain grammatical construction and the difficulty of its processing. First, before discussing this effect in people with aphasia, I will provide evidence for it for neurotypical speakers. This effect is very well documented for healthy adults and has been known for a long time. Oldfield and Wingfield (1965) ran an experiment measuring the time it takes to name an object, comparing it with the frequency the word for this object has in a corpus. They found that less frequent words have greater latency in naming - the less frequent the word is, the more time it takes to recognize an object with that name. Since then, this effect has been shown for different parts of speech and even grammatical constructions. Jurafsky (2002) provides an example of frequency effect in verbal constructions:

1. The doctor remembered [<sub>NP</sub> the idea].
2. The doctor remembered [<sub>S</sub> that the idea had already been proposed].
3. The doctor suspected [<sub>NP</sub> the idea].
4. The doctor suspected [<sub>S</sub> that the idea would turn out not to work]

The word frequencies acquired from a corpus reveal that the verb *remember* usually occurs with a noun phrase complement (NP) and the verb *suspect* usually occurs with a sentential complement (S). This creates a bias and as a result, the sentences (1) and (4) are easier to process. When experiment participants encounter less usual sentences such as (2) and (3), their reading times are longer, which suggest higher difficulty in processing.

## 1.7 Frequency effects in aphasia

Frequency effects play a central role in the usage-based approach to language and aphasia. The approach to aphasic language processing based on frequency (also called a probabilistic approach) has often lacked visibility, despite the fact that it has been well demonstrated in psycholinguistic research on neurotypical speakers (Gahl and Menn, 2016). Given that the existence of this effect in healthy speakers is uncontroversial, it should be expected that it applies just as well

to people with aphasia. Alternatively, if the probabilistic effects do not play a role in aphasic comprehension, it should be possible to explain why it is so. However, the literature has tended to look elsewhere when analyzing sentence level comprehension problems in aphasia, even in cases in which probabilistic effects were clearly demonstrated for neurotypical speakers (ibid.).

The effect of probabilistic verb bias in aphasia was analyzed by Gahl (2002). In this experiment, both healthy and participants with aphasia were presented with a *sentence plausibility judgement task*, a type of experimental method, in which participants are presented with a series of sentences and their task is to determine whether the presented sentences make sense or not. Two types of verbs were used in the sentences - verbs with an *intransitive verb bias*, that typically occur in intransitive sentences and verbs with *transitive verb bias*, that usually occur in transitive sentences. An example of a verb with an intransitive bias is the verb *melt* and an example of a verb with a transitive bias is the verb *dissolve*. The frequencies were extracted from a corpus of written language. Three types of sentences were used: active transitive sentences ('The cook melted the butter'), passive transitive sentences ('The butter was melted by the cook') and intransitive sentences ('The butter melted in the pot'). The sentences were presented in concordant conditions (e.g. the verb *melt* in an intransitive sentence) or discordant conditions (e.g. the verb *melt* in a transitive sentence).

As expected, the results show that healthy participants make very few errors overall. However, their error rate for discordant items is significantly higher than for concordant items, confirming the hypothesis that verb bias plays a role in sentence comprehension difficulty for healthy speakers. Similar results were reported for a group of patients with aphasia. The study shows that verb bias influences patients with aphasia as well, however the effect varies based on the type of aphasia. The effect was weaker in a group of patients with non-fluent aphasia. However, even in this group, some of the participants made more error on discordant than on concordant items.

These results were extended by Gahl et al. (2003), a similar experiment that also took into account semantic roles typically associated with the verbs. Three types of sentences were used: transitive active, transitive passive, intransitive with an undergoer subject and intransitive with an agentive subject. This study used a mixed group of patients with aphasia and a control group consisting of healthy participants. Again, the participants were presented with the sentence plausibility judgement task. The results show that the sentences are significantly easier to process, when the verbs are used in concordant condition e.g. the verb *melt* used in an intransitive sentence. Furthermore, the results show that semantic role bias also plays a role. Verbs that are usually associated with an undergoer subject are easier to process in undergoer-subject sentences than other types of verbs. This is in contrast with previous findings (e.g. (Kegl, 1995) in Gahl et al. (2003)), that hypothesize that intransitive, undergoer-subject sentences are inherently easier to process, due to their syntactic form being more canonical.

## 1.8 Agrammatism

Agrammatism is the main symptom of non-fluent aphasia, a one that is the most relevant to the practical part of this thesis. Agrammatism is an inability

to comprehend grammatical constructions of language and is manifested both in production and perception. The term *agrammatism* is sometimes contrasted with the term *paragrammatism*. If both terms are used, agrammatism is the omission of grammatical words or morphemes and paragrammatism is their substitution. Agrammatism is thus typical for patients with Broca's aphasia while paragrammatism is typical for patients with Wernicke's aphasia. If only agrammatism is used (as in this thesis) it can mean both omission and incorrect substitution of grammatical morphemes. Different researchers use different terminologies, however the term agrammatism that describes both of these phenomena seems to be more prevalent and more in accord with modern research findings (Flanderková, 2019).

In agrammatism, certain aspects of grammar are avoided, other can be overused. All aspects of grammar are typically affected. Code (2010) describes syntactic problems such as inability to use function words and comprehend argument structure of a sentence. Morphology is equally affected in patients with agrammatism - both free and bound morphemes may be either omitted or incorrectly substituted. Agrammatism is especially visible in non-fluent forms of aphasia, but it is in lesser extent present in all forms of this impairment.

Research in agrammatism helps with developing effective aphasia diagnosis and therapy, but it is also interesting from a theoretical viewpoint. When coupled with neuroimaging methods, research in agrammatism can help us understand the relationship between neurological processes and language.

Agrammatism is often thought of as an expressive disorder. However, the perceptive aspects of agrammatism cannot be neglected as they play an important role in everyday communication. The problems patients face in agrammatic comprehension, often mirror those they have with production (Flanderková, 2019, p. 52). Therefore, in this thesis, I will focus on perceptive deficits in agrammatic participants.

It has to be noted that the exact manifestation of agrammatism varies among languages. In the beginnings, research in agrammatism was often English-centric and so, its description was skewed and it was viewed as a primarily syntactic disorder. However, morphology will also be equally affected, especially in languages where it is complex. This has been documented for native speakers of Czech (Lehečková, 2001), where both syntax and morphological categories such as case, tense and person were affected. Full description of this impairment is only possible thanks to studies on a variety of typologically different languages. Language typology determines the exact form of agrammatism (Lehečková, 1988). In isolating languages, such as English, both grammatical morphemes and function words are usually omitted. This is because word roots and bare lexical morphemes in isolating languages are typically not bound. In synthetic languages, such as Czech, grammatical morphemes will typically be substituted, but free-standing function words may be missing. In this type of grammar, roots are frequently bound morphemes, and therefore some sort of suffix needs to be added. This means that bare lexical morphemes with omitted inflectional morphemes would result in non-word forms. Let us take the word *žena* 'a woman' as an example. The root of this word is *žen-*, a bound morpheme that has to be supplemented with a suffix, which provides information about gender, number and case. The suffix *-a* signifies that it is a feminine noun, in singular, in nominative case. Therefore,

the omission of the suffix would result in a non-word, and it is thus more likely to be incorrectly substituted in aphasic speech. This contrasts with agglutinative languages, in which both substitution and omission will be frequently observed.

The type of errors that occur in agrammatical speech is not random. Frequency of use has a major role in determining how likely an error is to occur. This will be discussed further in section 1.7. Other than frequency, other patterns of substitution have been proposed. Lehečková (1988, 2001) describes similar patterns of substitution in Czech and Finnish, two unrelated and typologically different languages. The author proposes that the number of incorrectly substituted morphemes is related to the order of their acquisition in childhood. The morphemes that are learned first by children are also those, that are less likely to get substituted incorrectly by adults with aphasia.

## 1.9 Reflexive verbs in Czech language

In this section, I will explain the notion of *reflexive verb* in Czech language. The practical part of this thesis will examine comprehension of sentences that contain these verbs and so, a brief introduction to this part of Czech grammar is needed. Put simply, reflexive verb is a verb, whose direct object is the same as its subject. In English, a verb can be made reflexive by attaching a prefix *self-* to it, for example the verb *self-destruct*. Alternatively, *self* can be used as a suffix that modifies a pronoun which is a direct object of a verb, for example in a sentence *He poured himself a cup of coffee*. Reflexive verbs are typologically common and can be found in Germanic languages (such as the pronoun *sich* in German), Romance languages (such as the pronoun *se* in French) or Slavic languages, as I will demonstrate. In Czech, reflexivity can be expressed using a variety reflexive pronouns (see Karlík (2017) for more information), but for the purpose of this thesis only the reflexive pronoun *se* is relevant. Other forms, such as *sebe* or *svůj* also mean ‘self’ and are used to express the similar functions in different constructions. However, they work as direct objects or adjectives, and do not create a *reflexive alternation* that changes the syntactic structure of a sentence. *Se* is a verbal clitic, that modifies a verb, but is not attached to it. I will demonstrate its use with a simple example:

- |    |   |           |             |
|----|---|-----------|-------------|
| 5. | Tatínek   | umyl      | nádobí.     |
|    | Dad.NOM.sg  | wash.past | dish.ACC.pl |
|    | ‘Dad washed the dishes.’                            |           |             |
|    |   |           |             |
| 6. | Tatínek   | se        | umyl.       |
|    | Dad.NOM.sg  | refl      | wash.past   |
|    | ‘Dad washed himself.’ (meaning ‘Dad had a shower.’) |           |             |

The sentence (5) is a canonical transitive sentence. It consists of a subject in nominative case, a verb and an object in accusative case. The sentence (6) is a canonical reflexive sentence. This sentence has the same structure as (5), except the noun phrase denoting the object has been replaced by the reflexive pronoun *se*. Unlike most nouns, the reflexive particle is not marked for case. It can be

analysed as a derivational morpheme of the verb or as a direct object (Veselý, 2018).

It has to be noted that the pronoun *se* is polysemous. It can carry a variety of meanings when used in the syntactic structure described in (6), based on the verb and the nouns with which it is combined. The reflexive function is considered to be the basic and diachronically primary function (Svoboda, 2014). As described in the next chapter, most but not all verbs will form a reflexive construction when combined with *se*. The semantic role of the subject also has an important role when determining the meaning of sentences with the reflexive pronoun. The subject of a reflexive sentence is an animate entity and has a semantic role of an agent. If an inanimate entity is used, the sentence can still be meaningful, but will carry a different meaning. Let us take the verb *zapsat* ‘to write down’ as an example. This verb can be used with an animate subject in a sentence such as *Vrátný se zapsal do seznamu* ‘The doorman wrote himself (his own name) into the list’. This is a typical reflexive sentence. If I change the subject of the sentence to *záznam* ‘entry’, the meaning of the sentence will change. It will become an anticausative sentence (described in section 1.9.1) meaning ‘An entry was written down into the list’.

In the experiment conducted as a part of this thesis, I will only be concerned with the reflexive function of *se*. Therefore, it is needed to distinguish between this function and other functions of *se*. These other functions will be discussed in the next section.

### 1.9.1 Other functions of the reflexive pronoun

The examples (5) and (6) demonstrate the most common and diachronically primary function of *se*. However it can have other meanings or functions, that depend on which verb it modifies and even on other syntactic constituents that occur in the same sentence. An overview of these functions is presented by Svoboda (2014).

#### Reciprocal function

One of the more common functions is the reciprocal function. An example is provided below:

7. Iva a Zuzka se často navštěvují  
Iva:NOM and Zuzka:NOM REFL:ACC often visit:3.PL  
‘Iva and Zuzka often visit each other’

This sentence (7) cannot be translated as ‘Iva and Zuzka visit themselves’, which would be an ungrammatical sentence. Rather, it has to be translated as ‘Iva and Zuzka visit each other’. Syntactically the sentence is the same as the sentence (6). The reciprocal meaning is inherent to the verb and is expressed when it is modified by the reflexive pronoun. The verb cannot be used with a reflexive meaning. The correct type of subject is also necessary. This function is expressed when the subject consists of multiple entities that engage in a mutual activity. Another examples of verbs typically used with reciprocal function are verbs such as *líbat* ‘to kiss’ or *objímat* ‘to hug’. The difference between the reflexive and



reciprocal function is purely semantic and can be reliably discriminated by native speakers.

### Anticausative function

This function is often used with deagentive verbs in order to eliminate the semantic role of *cause* in the sentence.

8. Loď se převrátila (na vlnách).  
boat:NOM REFL:ACC capsized:3.SG on waves:LOC  
'The boat capsized (in waves).'

Unlike the reflexive sentences, such as (6), the example (8) does not carry a reflexive meaning. It has to be translated not as 'the boat capsized itself', but rather as 'the boat was capsized'. The sentence has the same syntax and morphology as the sentence with a prototypical reflexive sense. The reflexive pronoun *se* and the verb form are the same as in (6). The anticausative meaning is created when the verb *převrátit* is combined with a reflexive particle and an inanimate agent. Another common examples of this functions are verbs such as *rozbít se* 'to break' or *otevřít se* 'to open'. Typically these verbs cannot be used with an animate subject.

### 1.9.2 Inherently reflexive verbs

Some Czech verbs can only occur with a reflexive pronoun. These are called *inherently reflexive verbs* or *reflexiva tantum* (Panevová and Karlík, 2017). It is a small, but semantically heterogeneous group of verbs without any semantic sense common to all the verbs. In this case, no specific function of the reflexive verb can be described. It is analysed as meaningless and is viewed simply a part of the lexicon (Kettnerová and Lopatková, 2014). There are around 120 such verbs in written Czech (Lopatková et al., 2016).

In addition, some verbs can occur in transitive sentences and in sentences with a reflexive pronoun, however their meanings are not related by any derivational relation. A good example is a verb *jmenovat* 'to appoint, nominate' and *jmenovat se* 'to be called'. Another example is *hodit* 'to throw' vs. *hodit se* 'to match'. These verbs are analyzed as two distinct lexemes, one of which is inherently reflexive (Kettnerová and Lopatková, 2014).

# Chapter 2

## Method

In this chapter, I will provide a detailed overview of all steps that I took when designing and conducting an experiment on processing of reflexive constructions in speakers of Czech. The first part consists of examining the use of reflexive verbs in Czech in order to obtain realistic material for the experiment. After that, I ran two experiment trials on healthy speakers of Czech, in order to obtain baseline results of syntactic processing of reflexive constructions, that can serve as a baseline when examining reflexive construction processing in aphasia.

### 2.1 Corpus research

The first step was to gather and examine data of reflexive constructions use in healthy language. A quick corpus search reveals 20 553 different verb lemmas used with the reflexive verb *se*. However, not all of these are reflexive verbs. As described in section 1.9, verbs used with the reflexive pronoun can have different functions, based on the semantics of the verb. In addition, some of these verbs are only used in constructions with the reflexive pronoun and are not used in regular transitive sentences. Out of those verbs that are used in both constructions, some are irregular and their meaning is different when used in reflexive and transitive construction. Therefore, the goal of the corpus research was to create a list of roughly 20 verbs, that frequently occur in both transitive constructions, and in construction with a reflexive meaning <sup>1</sup>. They enabled me to directly compare the difficulty of processing of both constructions, with other variables being equal and served as basis for the experiment.

#### 2.1.1 The corpus

The data was extracted from a corpus of written Czech language *SYN2015*, which is a part of the *Czech national corpus* (Křen et al., 2015). The extracted data was analysed using R programming language (R Core Team, 2021) The corpus contains over 100 million tokens (without punctuation) gathered between years 2010 and 2014. This corpus is representative - it contains a wide variety of

---

<sup>1</sup>terminology note: I use the term *construction* as it is used in construction grammar - as a combination of a certain syntactic and semantic properties, that together form and abstract grammatical meaning

genres and topics used in modern Czech. It is also balanced - all genres or styles are represented in approximately equal amounts.

This corpus is lemmatized and annotated for syntax and morphology. Morphological annotation contains information such as part-of-speech, case, number, tense or mood for all relevant words. It enabled me to search for a particular word in all its shapes and forms, for example find all word occurrences in accusative case. Syntactic annotation enabled me to view and process the syntactic structure of phrases and sentences. Figure 2.1 provides an example of syntactic annotation in SYN2015. Syntactic trees are automatically generated for most of the sentences in the corpus. Using a combination of morphological and syntactic annotation, I was able to create queries such as “find all transitive occurrences of this verb with arguments in accusative” or “find all occurrences of this verb when modified with a reflexive pronoun”. This enabled me to easily search for verbs that are used with the reflexive pronoun, but also without it.

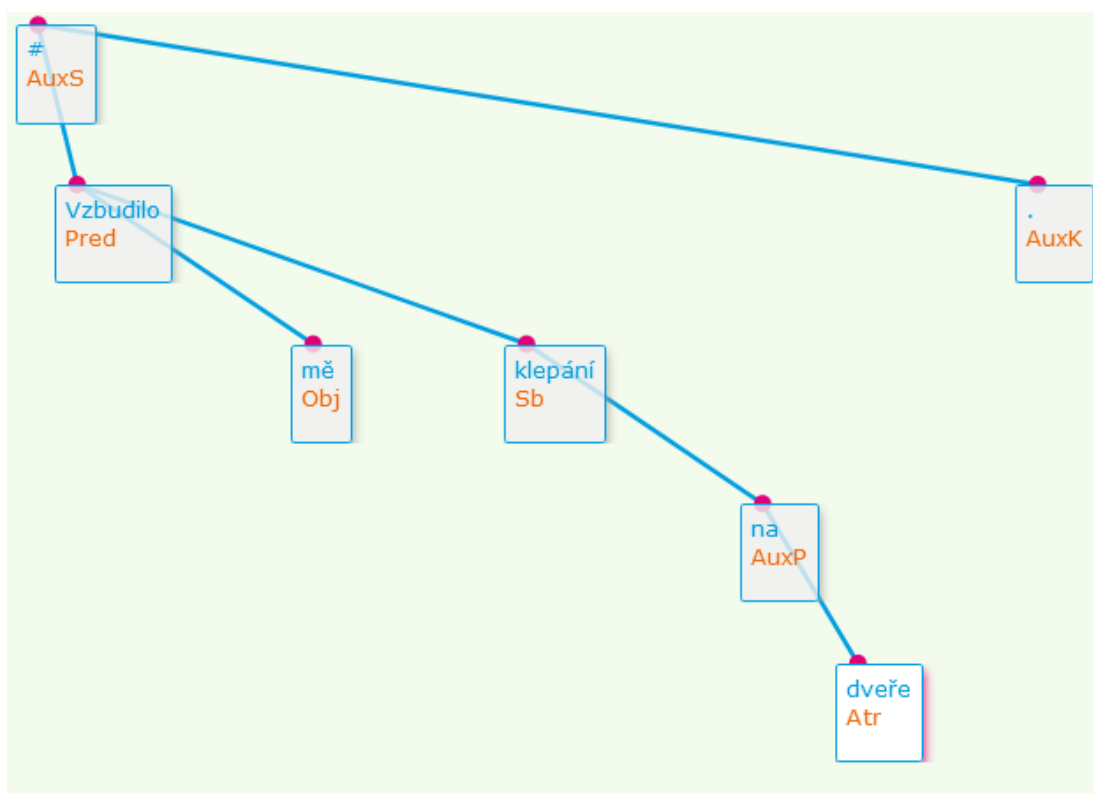


Figure 2.1: An example of syntactic annotation in SYN2015. This is an automatically created syntactic tree for a simple sentence *Vzbudilo mě klepání na dveře* meaning ‘Knocking on the door woke me up.’ Syntactic functions, such as predicate, subject and object are marked in the annotation.

However, using the corpus SYN2015 for research on aphasia carries some disadvantages. After all, it is a corpus of written language and the difference between spoken and written Czech can be significant at times. Frequency of word use in different constructions can differ between written language and spoken language that the patients encounter in their day-to-day lives. As of this date however, a large-scale syntactically annotated corpus of spoken Czech does not exist and so a corpus of written language has to be used.

### 2.1.2 Data extraction

After choosing the most suitable corpus, I proceeded with finding the verbs that frequently occur in both reflexive and transitive constructions. I started with creating a list of verbs that occur in transitive constructions.<sup>2</sup> These are the verbs that occur with a noun in accusative case in their argument structure. Syntactic labelling used in SYN2015 made creating such a list fairly straightforward.<sup>3</sup> The same method was used to create a list of reflexive verbs - these are simply the verbs that occur with the reflexive pronoun in their argument structure. I intersected the two lists, creating a list of verbs that can be used in both types of structures. This resulted in a list of almost 10 000 verbs.

In the next step, I proceeded with calculating a ratio of reflexive vs. non-reflexive uses for each word. This is simply a percentage of reflexive occurrences of the verb, compared to total number of occurrences for the verb. From now on, I will call this value *reflexivity bias*, analogous to *transitivity bias* in Gahl and Menn (2016).

The next step was to automatically filter the list of verbs, based on their total frequency of use and reflexivity bias. Low frequency verbs are unsuitable for the experiment, because the participants may be unfamiliar with them. Verbs with extremely high frequency are also unsuitable. The most frequent words contained in the list were verbs such as *být* ‘to be’ and *mít* ‘to have’, that are not used in reflexive constructions, but were included due to errors in corpus annotation, that inevitably come with such high frequencies. Other extremely frequent verbs were modal verbs, that cannot be used alone and only modify other verbs. In addition, I am interested in general syntactic abilities of the participants, not in the mastery of a few extremely common verbs. Extremely frequent verbs are likely to be mastered even by highly disfluent participants, may be overused and appear as outliers in the analysis. As a result, I filtered out the verbs, that had fewer than 20 or more than 500 instances of use per million words in the corpus.

I also filtered out the verbs that had their reflexivity bias lower than 0.2 or higher than 0.8. These are the verbs that might be unfamiliar to the participants in both constructions, or verbs that have different meanings when used with and without a reflexive pronoun. I also verified this list by comparing it with a list of inherently reflexive verbs, provided by the Valency lexicon of Czech verbs (Lopatková et al., 2016), which is also based on corpus data. These verbs occur only with the reflexive pronoun and therefore are unsuitable for the experiment (see section 1.9.2 for more details). My list does not contain any of these exclusively reflexive verbs. This automatic filtering resulted in a list of 201 verbs and made the subsequent manual processing feasible.

In the next step a carefully considered each verb in the list and annotated it for the following variables:

- Semantic compatibility - semantically compatible verbs are “regular” verbs, that carry roughly the same meaning in both transitive constructions and constructions with the reflexive pronoun. An example of a semantically

---

<sup>2</sup>this list, as well as all the other data collected from the corpus is publically available here: <https://github.com/DanielSuchy/reflexive-alternation-in-aphasia>

<sup>3</sup>the exact list of queries used to obtain the data is presented in Appendix B

compatible verb would be *mýt* ‘to wash (something)’ and its reflexive counterpart *mýt se* ‘to wash oneself’. An example of semantically incompatible verb would be *jmenovat* ‘to appoint’ which changes its meaning to *jmenovat se* ‘to be named’ when modified with a reflexive pronoun.

- Anticausativity - these are the verbs that usually carry an anticausative meaning when used with the reflexive pronoun (Svoboda, 2014, p. 45). This is an alternative function of the reflexive pronoun, that is irrelevant to the topic of this thesis (see section 1.9.1 for more information).
- Reciprocity - these are the verbs that usually carry a reciprocal meaning when used with the reflexive pronoun (Svoboda, 2014, p. 36). Again, this is a marginal function of the reflexive and it poses a different kind of processing difficulty for the participants (see section 1.9.1 for more information about the reciprocal constructions).

I filtered out the verbs that are not semantically compatible or that are primarily used to express the anticausative or reciprocal functions. This resulted in a list of prototypical, regular reflexive verbs, according to the criteria outlined in section 1.9. This step ensured that the difficulty of syntactic processing is of the same kind for all of these verbs and the results for them can be directly compared. In addition I looked at the distribution of verbs compared to reflexivity, shown in the density function plot in the figure 2.2. This graph shows that most verbs have a low reflexivity bias around 0.3, but verbs with higher reflexivity bias (verbs that are used mostly in reflexive constructions) are also present in sufficient numbers.

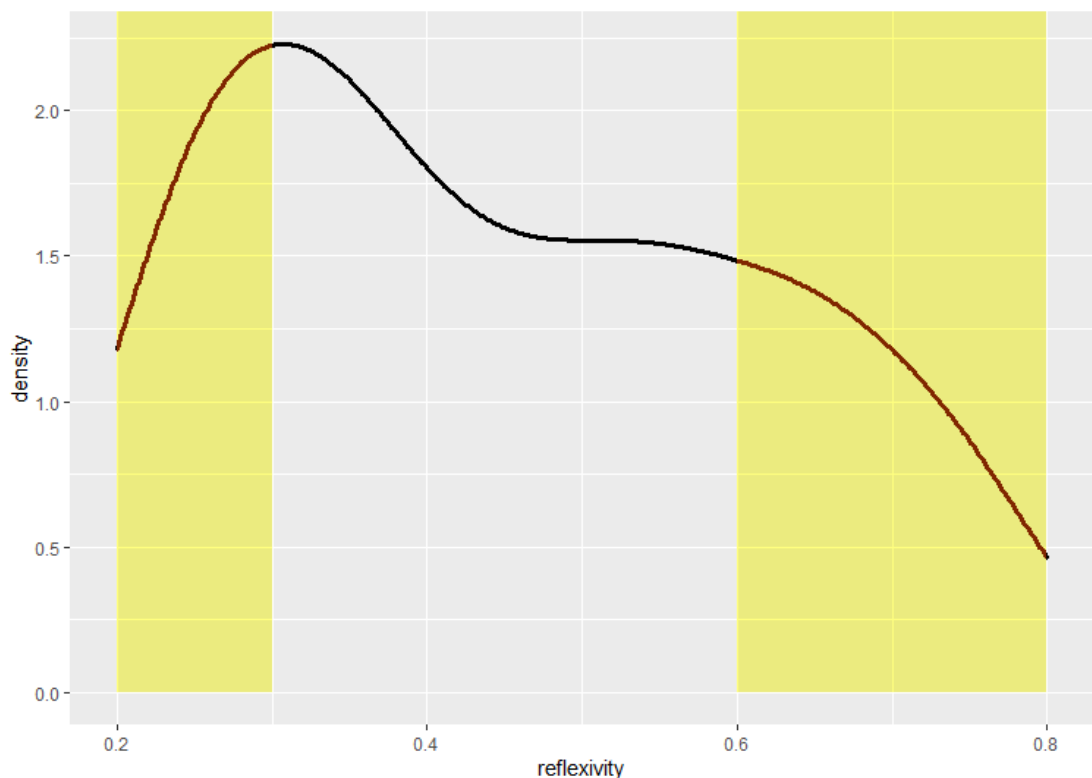


Figure 2.2: Density function of verb reflexivity bias. The two groups with high or low reflexive bias are highlighted.

As a result I decided to split this list into two groups - verbs with low reflexive bias (between 0.2 and 0.3) and verbs with high reflexive bias (between 0.59 and 0.8). This enabled me to clearly see the effect of reflexivity bias on sentence comprehension. From each of these groups I chose a list of 8 verbs based on the criteria above (semantic compatibility and lack of anticausativity and reciprocity), 16 verbs in total. These verbs serve as a basis for the experiment. The List of these lemmas, alongside their reflexive bias and relative corpus frequency is presented in table 2.1.

lemma	reflexive use %	ipm	refl. bias group
ponořit	59	23.40	high
naklonit	72	40.94	high
přestěhovat	80	29.82	high
rozesmát	75	29.28	high
vypravit	66	27.56	high
zaměřit	61	61.87	high
zbavit	66	82.18	high
probudit	62	66.33	high
hájit	25	20.88	low
ohlásit	22	21.95	low
oddělit	21	26.10	low
stáhnout	26	70.41	low
trápit	28	57.14	low
vzbudit	21	42.76	low
zapsat	29	41.76	low
tlačit	21	40.56	low

Table 2.1: List of lemmas selected for the experiments. The abbreviation *ipm* stands for instances per million - how frequently on average this lemma occurs in a million word instances in the corpus.

## 2.2 Experimental items design

After I identified the most suitable verbs to be used in the experiment I was able to proceed with designing experimental items that can be used to test comprehension of healthy speakers and participants with aphasia. These experimental items are simple sentences that contain the target verb in various grammatical constructions and with various noun and prepositional phrases used in their argument structure. The task for the participants is to evaluate these sentences, judge whether the sentence “makes sense” or not. Their judgements will be compared to the predictions about sentence grammaticality outlined below, in order to understand whether the participants interpret the sentences as predicted. Four types of sentences are used in the experiment:

- grammatical transitive sentences, such as: <sup>4</sup>

Dělník      přestěhoval      nábytek      do      bytu  
 workman.NOM    move.PAST    furniture.ACC    into    apartment.ACC

‘The workman has moved the furniture into the apartment’

- grammatical reflexive sentences, such as:

Dělník      se      přestěhoval      do      bytu  
 workman.NOM    REFL    move.PAST    into    apartment.ACC

‘The workman has moved into the apartment’

- ungrammatical sentences missing both a direct object and a reflexive pronoun, such as:

\*Dělník      přestěhoval      do      bytu  
 workman.NOM    move.PAST    into    apartment.ACC

- ungrammatical sentences containing both an object and a reflexive pronoun, such as:

\*Dělník      se      přestěhoval      nábytek      do      bytu  
 workman.NOM    REFL    move.PAST    furniture.ACC    into    apartment.ACC

For each of the target verbs, 8 such sentences were created - two of each grammatical type, with different noun and prepositional phrases. A complete list of these 128 sentences is presented in Appendix A. In order to avoid an interference caused by anomia or lack of lexical knowledge, simple, highly frequent nouns were used. Common professions or simple kin terms were used as subjects. Subjects were always animate, because only animate subjects can be used in both transitive and reflexive sentences. If possible, the selected noun phrases form frequent collocations with the verbs and describe common events or actions.

## 2.3 Verification of stimulus sentences (experiment I)

Before presenting the experiment to participants with aphasia, the design of the experiment as well as the experimental sentences needed to be verified by

---

<sup>4</sup>The provided morphological annotation contains only the most important labels, relevant for the experiments. ‘NOM’ stands for the nominative case, ‘ACC’ stands for the accusative case, ‘PAST’ stands for a past tense verb form and REFL stands for the reflexive pronoun *se*

healthy participants. In order to do this I have created a simple preliminary experiment that consists of a series of sentence acceptability judgements. The goal was to verify that a group of Czech native speakers agrees on the grammaticality of the stimulus sentences. In this experiment, I aim to verify that the participants understand the sentences in the way I predicted and their understanding of the sentences is consistent. The transitive and reflexive sentences should be unanimously considered as grammatical, while the oddly-formed sentences should be perceived as ungrammatical.

First, I have created a simple web based experiment environment using PenController for IBEX (Schwarz and Zehr, 2018), a tool that enables creating psycholinguistic experiments and presenting them online. In this interface, participants are asked to enter personal details that will be relevant for an analysis of the results. This includes their native language, age, occupation and any problems with reading. After that, the participants are presented with a simple acceptability judgement task. All of the 128 experimental sentences described in the section above are presented to the participant in random order. Their task is to judge, whether the presented sentences are meaningful or not. Participants are told to judge, whether the sentence “makes sense grammatically”. They are instructed to chose from three options - *dává smysl* ‘sentence makes sense’ or *nedává smysl* ‘sentece does not make sense’. Alternatively, they can also use a third option, saying *nejsem si jistá/ý* ‘I am not sure’. At the end of the experiment, a participant had an opportunity to provide feedback on the experiment. This can be used to improve the design or indicate any problematic items.

I expect that the presented sentences will not pose any problems to healthy native speakers of Czech. Errors should be randomly distributed, attributed to a lack of attention during the experiment, which can take around 10 minutes to complete. Reaction times should be low and consistent. Alternatively, if multiple systematic errors or significantly longer reaction times occur in a single item, it could mean that the item is problematic and needs to be replaced before presenting the experiment to the participants with aphasia. The same holds true for an accumulation of multiple indifferent (“I am not sure”) answers.

### 2.3.1 Results

The results of experiment I were analysed using the R programming language (R Core Team, 2021) and are publicly available.<sup>5</sup> In this phase, 35 students taking an introductory course to linguistics participated in the experiment for partial credit. Their mean age was 20.54, standard deviation was 2.09 and it ranged between 19 and 27. They were all Czech native speakers, students of language-related majors. Three of them reported that they experience some problems with reading. Each participant evaluated the same set of 128 unique sentences, presented in a random order. In total, 4480 sentence evaluations were collected during the whole experiment. None of these were excluded, as none of the participants markedly deviated from the general trend. 6 of the 35 participants judged all items as expected, only 3 differentiated from the expected judgements in more than 10 percent of cases. On average, participants judged

---

<sup>5</sup>the results can be accessed using this link: <https://github.com/DanielSuchy/reflexive-alternation-in-aphasia>



the sentences as expected in 96 percent of cases, the standard deviation was 3.92. The lowest scoring participant provided concordant answers in 82 percent of items. The three participants with reading problems did not differentiate from this trend, with their ‘success rate’ being only half a percent lower.

The results are presented in the table 2.1. In total, 122 discordant judgements were made, in which a participant judged an item as grammatical instead of ungrammatical or vice versa. In 52 cases, the participants expressed uncertainty, marking an item with an ‘I am not sure’ answer. As expected, less frequent / less canonical types of sentences are more confusing. The grammatical sentences did not pose any problems to the participants. Transitive sentences are very rarely problematic, reflexive sentences tend to be problematic slightly more often. The ungrammatical sentences that lack both an object and a reflexive particle (last row in the table) are the most problematic and form the majority of incorrect judgements. This type of a sentence is not grammatical, but could still be interpreted as odd but meaningful by a willing listener in a highly specific context. However, the participants are instructed to judge, whether the sentence is grammatical or “oddly-formed”, not interpretable. This is clearly visible from the results - these sentences, expected to be ungrammatical, were sometimes judged as grammatical. This contrasts with the other type of ungrammatical sentences that contain both object and a reflexive. This type, despite being ungrammatical, did not confuse the participants.

sentence type	expected	grammatical	ungrammatical	unsure
transitive	grammatical	1111	8	1
reflexive	grammatical	1096	20	4
w. object and reflexive	ungrammatical	12	1105	3
w/o object or reflexive	ungrammatical	82	994	44

Table 2.2: Plausibility judgement task for experiment I

The experiment showed that participants behave as predicted when presented with the judgement task. It verified that the items are valid and ready to be used in experiments with both control and clinical group. It managed to verify all possible combinations of verbs lemmas, and syntactic structures, providing a large sample of judgements despite the small group of participants. However, it is unsuitable beyond this point - it repeats the same stimuli multiple times, such as the same lemma in multiple constructions. It cannot provide an adequate evidence for the effect syntactic structure and verb bias on sentence comprehension. Also, it is too long for both patients with aphasia, who tire quickly, but also for healthy participants, who can easily guess the purpose of the experiment, when presented with the same patterns multiple times.

## 2.4 Verification of experiment design (experiment II)

In the second phase, I designed the final version of the experiment, based on the results of the experiment I. Its design needs to be suitable to be used in both healthy participants (control group) and participants with aphasia (clinical

group). The goal of this next version of the experiment is to measure any possible effects of syntactic structure and verb bias on sentence comprehension. This version was shortened - unlike in the previous version, a participant is presented with a limited combination of all used independent variables (verb lemma, grammaticality, syntactic structure and other arguments attached to the verb), which will ensure a higher reliability of collected data. As an additional point the design needed to be shortened, in order to avoid loss of attention from the participants, and in order to be presented to heavily impaired speakers.

The subset of sentences presented to a single participant is balanced. In total, a participant evaluates 32 sentences. Each of the verbs used in the experiment is presented twice, each time with different arguments. Each type of sentence structure is presented 8 times, each time with different variables such as verb lemma and other syntactic arguments. In order to achieve this, so called *latin square* design is utilized. In this type of design, the first participant is presented with a transitive version of the first sentence, reflexive version of the second sentence and so on. The second participant is presented with different types of sentences for each of the verbs, e.g. reflexive version of the first sentence, transitive version of the second sentence, etc. This type of design ensures that all types of items will be tested in similar numbers, while keeping the experiment short.

The items are the same as in the previous version, with one change only. I changed the sentences for a single verb - *zapsat* 'to write down', which caused the largest amount of discord in sentences without object or reflexive. Sentences with this verb were responsible for 18 errors out of 82 (see table 2.1). In these sentences, I changed the prepositional phrase, giving it a different semantic role (locative instead of goal), so that it cannot be confused with a direct object. For example, I change an ungrammatical sentence *\*Vrátný zapsal do kalendáře* 'The doorman wrote down the calendar' to *\*Vrátný zapsal u vchodu* '\*The doorman wrote down by the entrance'. The previous version of the sentence is ungrammatical, the verb *zapsat* is transitive, requires a direct object. However, it could potentially be interpreted as grammatical, by a willing and imaginative participant, who could assume that the sentence would be unusual, but interpretable. Such participant may assume that the object of this sentence could be somehow obvious, for example if mentioned in a previous context of a conversation. I am not interested in such possible interpretations, but in judgements of the simple, prototypical sentences. Therefore the context in which the action of *zapsat* takes place was changed, in order to make such an imaginative interpretation less likely.

The experiment is presented using a similar web-based interface as in the first experiment. An important difference is that the indifference option was removed. participants now have to decide whether the sentence "makes sense grammatically", without any third option. The verbal ratings ("yes, makes sense" "no, it does not") were replaced with symbols - a green check mark that stands for a positive judgement and a red x mark that stands for negative. This should remove the interference of reading difficulty in heavily impaired participants.

# Chapter 3

## Results

The experiment described in 2.4 ran over a period of 26 days, during which 115 participants completed the experiment. Its results were analysed using the R programming language (R Core Team, 2021) and are publicly available.<sup>1</sup> A dummy test version of this experiment is also publicly available.<sup>2</sup>

The participants were recruited using multiple ways. Some are university students, attending philology and linguistics-related classes at the Faculty of Arts. Another large group consists of high school students, participating in *Czech Linguistic Olympiad*, an annual linguistic competition for high school students. The rest are acquaintances of the author or were recruited using online social networks. It has to be noted that the participant sample is not balanced. Many of the participants are students who focus on language study and as a result their sensitivity towards language phenomena may be increased. In addition, higher education may be associated with higher social and economical status. Overall, the sample is not representative of the population of Czech speakers as a whole. This fact will be further elaborated in the discussion in the next chapter.

The participants are quite young overall (mean age is 19.4, standard deviation is 3.3), with only 6 being aged 25 or older. The gender distribution is roughly equal, 62 females and 51 males completed the experiment. Two participants indicated other/non-binary as their gender. The participants were also asked to report their highest achieved education and their current occupation. The results are quite diverse when it comes to education, with 72 participants reporting elementary as their highest completed level of education, 32 reporting high-school and 11 reporting university education. However, most of the participants indicated that they are still studying. This means that most of the participants are high school or university students.

The sample is relatively homogeneous when it comes to native languages of the participants, with 108 reporting Czech and 7 reporting ‘other’ as their native language. The non-native speakers do not markedly differ from the main group, their ‘success rate’ in the experiment is only 3 percent lower. The best non-native speaker concurred with the expected judgements in all experimental items, while the worst concurred in 87 percent, which is better than two of the participant

---

<sup>1</sup>you can access the results using this link: : <https://github.com/DanielSuchy/reflexive-alternation-in-aphasia>

<sup>2</sup>use this link:  
[https://expt.pcibex.net/ibexexps/suchyda/hodnoceni\\_vet\\_pozastavene/experiment.html](https://expt.pcibex.net/ibexexps/suchyda/hodnoceni_vet_pozastavene/experiment.html)

from the group of native speakers. However, I still decided to exclude the group of native speakers. The reason behind this is the simplicity of the experimental sentences. The sentences use very common words, and are canonical examples of one of the most common syntactic structures in Czech. Therefore, in my opinion, even a participant with less than complete knowledge of Czech grammar, but with a developed sense for linguistic analysis could successfully complete the experiment. In the experiment, I am only interested in the intuitive knowledge of native speakers.

In addition, 4 participants reported that they experience some sort of reading problems. The experiment form did not ask about the severity or the type of problems that these participants encounter. These participants do not differ from the main group of speakers. Their mean success rate is 97 percent, one percent lower than the rest of speakers. Two of the speakers scored 100 percent success rate, the other two scored above 90 percent. I decided to keep these participant in further analysis. As expected, their reading problems did not affect their results and these sentences should be easy enough to process even for speakers with dyslexia.

In total, the included participants completed 3456 experimental items. Each item consists of a presented sentence and a judgement about sentence plausibility, completed by the participant. The judgement can be either *concordant*, meaning that it matches the predicted judgement for the sentence, or *discordant*, meaning the the participant evaluated the sentence differently than expected (e.g. labeled a sentences as ungrammatical, when grammatical judgement was expected.) Each participant evaluated 32 unique sentences. Each unique sentence was on average presented to 27 different participants (minimum 18, maximum 37 participants). This uneven distribution is caused by the experimental interface - if two participants work on the experiment during the same time, they are assigned the same 'seed', are are therefore given the same set of sentences.

Out of the 3456 judgements, 3374 were judged according to the expectations - they were judged as either grammatical or ungrammatical, in accord with the expected judgement that I assigned to the sentences when designing the experiment (see section 2.2 for details on which sentences are considered grammatically correct or incorrect). 82 items were judged differently than expected - they were judged as grammatical when ungrammatical judgement was expected or vice versa. To sum up, judgements of experimental items provided the expected results in 98 percent of cases.

Next, I analyzed each participant individually, in order to find any outliers. The minimum "success rate", the ratio at which participants judged the sentences as expected was 81 percent. Only 4 participants scored bellow 90 percent, while 61 participants scored 100 percent, meaning that they judged every single sentence as expected. As a result, I decided not to exclude any participants from further analysis. This high success rate implies also that the task was extremely easy for healthy participants.

In the next step, I analyzed the effect of sentence type on the rate of discordant judgements. The results are summed up in table 3.1. The three types of sentences - transitive, reflexive and ungrammatical sentences that contain both a reflexive particle and a direct object posed no problem to the participants. This very low rate of discordant judgements, can be attributed to random error, when a

participant is presented with a large number of similar sentences. However, the sentences without both the reflexive particle and the direct object are markedly more error-prone. These sentences were expected to be judged as ungrammatical, but were in some cases judged as grammatical during the experiment. This result is not unexpected, given the results gathered in the Experiment I, in which the participants often marked these sentences as ambiguous as they could not decide whether the sentence is grammatical or not. Ungrammatical sentences such as these are expected to yield a slightly higher percentage of discordant judgements and confirm the hypothesis that ungrammatical sentences are more difficult to process, even for healthy participants.

sentence type	total judgements	discordant judgements	% discordant
transitive	864	16	1.9
reflexive	864	8	0.9
w. object and reflexive	864	10	1.2
w/o object or reflexive	864	48	5.6

Table 3.1: Number of discordant judgements per sentence type

Next, I attempted to analyze whether this higher rate of discordant judgements for sentences w/o object or reflexive is indeed a result of higher processing difficulty and not, for example, caused by an ill-chosen experimental item. However, the analysis shows that the distribution of discordant judgements between items does not show any significant trends. The mean number of discordant judgements was 1.5 per item, even for this most problematic sentence type. None of these sentences were exceptionally problematic, most were correctly judged by all participants. The highest rate of discordant judgements per sentence w/o object or reflexive is 6. This means that even this type of sentence is judged as expected in the vast majority of cases, regardless of the verbs or nouns that the item contains.

As the last step, I looked at the distribution of discordant judgements, when accounted for the reflexivity bias. As described in section 1.7, frequency of use (or in this case frequency of occurrence in a corpus) plays a major role in sentence processing difficulty, with less common sentences being more difficult to process. Verbs with a high reflexivity bias (those frequently used in reflexive sentences) should be easier to process when they occur in a reflexive sentence, while those with a low reflexivity bias should be easier to process in transitive sentences. In an experiment, this should manifest itself as a higher rate of discordant judgements, when a verb is used in a sentence opposing its reflexive bias. The results of this experiment, compared with reflexive bias are presented in table 3.2.

These results do not show any effects of the reflexive bias. In total 12 discordant judgements were made in sentences with a matching condition, while the same amount (12) were made in sentences with a mismatching condition. If reflexive bias played a major role in sentence comprehension for healthy speakers in these types of sentences, the rate of discordant judgements should be significantly higher in mismatching condition. This shows that the effect of reflexive bias on healthy speakers is low, if any. It is not surprising, given the high overall “success” rate of the participants and it may indicate that the sentences are extremely easy for healthy speakers. Importantly, the numbers presented here are very low, as

sentence type	reflexive bias	condition	n. discordant
reflexive	high	matching	2
transitive	low	matching	10
reflexive	low	mismatching	6
transitive	high	mismatching	6

Table 3.2: Number of discordant judgements (mis)matching condition. Condition is matching if sentence type matches the reflexive bias. Only sentences that are expected to be judged as grammatical are included.

the speakers are unlikely to make a discordant judgement when presented with a grammatical sentence. However, it will be interesting to compare these numbers with data from speakers with aphasia in order to determine, whether they are more sensitive to verb bias.

# Chapter 4

## Conclusion

### 4.1 Discussion

In the preceding chapters, I researched the use of reflexive verbs in Czech, designed an experiment that examined the difficulty of processing such sentences and ran two experiment trials on healthy speakers. However, the presented work is not without flaws and the results show that further work needs to be done before presenting the experiment to speakers with aphasia.

First, the results of the second experiment need to be discussed. The grammatical sentences and the sentences with both object and a reflexive pronoun provided expected results. The rate of discordant judgements was low, participants evaluated the sentences as expected. However, the sentences without an object and a reflexive were more problematic, forming more than a half of all discordant judgements. In theory this could be caused by the fact that these sentences are ungrammatical and therefore confusing and harder to process. However, another factor, also briefly discussed in section 2.4 could also be a potential source of an error. For the sake of simple instructions, the participants were instructed to judge, whether the presented sentence is a ‘normal Czech sentence’ (*normální česká věta*). This could lead some of the participants to judge sentence interpretability not grammaticality. This is not a problem for the grammatical sentences, which were always easily interpretable. It is also not a problem for the sentences with both an object and a reflexive, which cannot be interpreted, as it is not clear, whether the presented sentence is transitive or reflexive.

However, this may not be true for the sentences which lack both a reflexive pronoun and a direct object. In order to comply with the experiment instructions, an imaginative participant could feel ‘forced’ to attempt to interpret such sentence, even if it is oddly formed. Such a participant may judge that the object missing in a sentence *Vrátný zapsal u vchodu* ‘The doorman wrote down by the entrance’ is somehow clear from a previous context. For example it could be an answer to a question such as *Kdo zapsal tu návštěvu?* ‘Who wrote down that those visitors entered?’. An answer such as the one presented above is indeed ungrammatical and oddly-formed, but would have a clear interpretation in this context.

Such an interpretation is of course not the purpose of the experiment, which should measure the comprehension of simple, ordinary, out-of-context sentences. Therefore, the instructions should be modified to reflect this more accurately. It

should be more clearly stated that the participants should judge the sentences without hesitation and exactly as they are presented, without context.

Another possible sources of inaccuracy could be the participant sample. As noted in the previous chapter, the sample is not balanced and is composed mostly of young, linguistically gifted students. This could possibly lead to distortion in results, especially if the current sample produces fewer judgement errors than is typical of healthy Czech population. As a further step, a more diverse sample should be recruited, preferably a one that will be matched when it comes to age and education to a population of patients with aphasia. Different generations of people can perhaps process the sentences in different ways, leading to different results. It is important to note however, that the presented experimental items are short and simple sentences, which were intentionally filled with some of the most common words. In my opinion, the discussed sample provided sufficient evaluation of the experiment design and the experimental items, but a serious application of the experiment would require recruiting a reliable control group.

Another weak point of the presented work could be the method of obtaining results. Therefore, the next step should be an extension of experiment II, in which reaction times are measured. I programmed the interface for experiment II so that it can be completed by pressing keys on a keyboard, a method by which reaction times can be measured. However, I decided against using this feature in the experiment, since the participants complete the experiment at home, which is an uncontrolled, distracting environment. Therefore, in this step, a different set of participants should provide judgements on the experimental items designed in the previous steps, but will enter their input using a keyboard. Preferably, this should be done with participants who are physically present in a controlled environment, to avoid outside distractions, that could prevent a reliable measurements of reaction times. Reaction times will be measured for each sentence - this method should allow for directly measuring the difficulty of sentence processing, a much more reliable method than just plausibility judgements used in previous phases. A longer reaction time would indicate a more demanding decision making process, regardless of whether the resulting answer concurs with the expected judgement or not.

The group of participants involved in this step should serve as a control group for the last step, which will involve conducting the experiment with speakers with aphasia. The design utilised in experiment II should be simple to understand and short enough, so that even heavily impaired participants can be surveyed. Again, reaction times should be measured and compared with the control group. This will enable to reliably measure the effect of both sentence structure and verb bias on speakers with aphasia and compare them with a control group composed of healthy speakers.

## 4.2 Concluding summary

Here, I provide a brief summary of all the steps that I took when writing this thesis. My goal was to examine the effects syntactic structure and frequency of occurrence on syntactic processing in healthy speakers of Czech, and create a basis for evaluating these effects in speakers with aphasia. First, I described the topic of aphasia in general, before moving on to syntactic processing in aphasia



and even more specifically on probabilistic models of syntactic processing. After providing a theoretical overview of the topic, I proceeded with conducting a corpus based research on the use of reflexive sentences in Czech. I discovered that different types of verbs occur in reflexive sentences with different probabilities. This enabled me to design an experiment that tests the role of relative frequency in the comprehension of reflexive sentences. The experiment consists of a sentence plausibility judgement task, in which the participants evaluate, whether a sentence ‘makes sense’ or not. I ran two trials of the experiment on healthy speakers, each time with a slightly different design. The goal of the first trial was to evaluate the experimental items, in order to determine whether the judgements of participants concur with what I predicted before the trial. Each experimental item was evaluated individually, in order to find those, where my predictions diverged from the results. No systematic divergences from the expected results were found and the participants behaved as expected. Participants were more likely to err when judging a single particular sentence (in all its syntactic forms). This sentence was corrected after the trial.

After the first trial, the experiment was modified so that the stimuli do not repeat and in order to be suitable to be used when evaluating both healthy speakers and speakers with aphasia. The second trial was conducted on a larger group of healthy speakers and attempted to discover any effect of sentence structure and verb bias on comprehension. It also served as a verification of the final design of the experiment, which has to be suitable to be used with both healthy speakers and speakers with aphasia. As expected, the results show that the items were easy to process for healthy participants, who concurred with the predicted judgements of grammaticality in the vast majority of cases. Variance between the speakers was very low, as was the variance of discordant judgements across sentences. This indicates that the experimental items are ready to be used in an experiment with speakers with aphasia and a matched control group.

The results showed a clear effect of sentence structure on the rate of concordant judgements. The participants agree almost unanimously when evaluating grammatical sentences, but judgements of ungrammatical sentences can cause a disagreement in a small number of cases. Some of the sentences that were expected to be labeled as ungrammatical were sometimes judged as grammatical by the participants. However, these sentences should be evaluated as ‘confusing’ instead of ‘difficult’ and the results may indicate that the participants attempt to make sense of the sentence, even though it seems unusual.

In this trial, I also attempted to evaluate the effect of frequency of use on sentence processing. Certain verbs occur in reflexive sentences more frequently than the others and thus have a high ‘reflexive bias’. Others rarely occur in reflexive sentences and thus their reflexive bias is low. Based on the literature described in the introduction, I predicted that a verb with a high reflexive bias will be easy to process in a reflexive sentence, while a verb with a low reflexive bias will be easier to process in a transitive sentence. However, this prediction was not confirmed by the experiment on healthy participants. The rate of ‘correct’ judgements of the sentences was not markedly influenced by the verb bias. This is not unexpected, given that the sentences were very easy to process for healthy speakers, and the overall error rate was low. These results need to be compared with a sample of speakers of aphasia, who should find it more difficult to process these sentences and could

thus show a pronounced effect of verb bias on sentence comprehension.

# Bibliography

- Amo Bakheit. Drug treatment of poststroke aphasia. *Expert Review of Neurotherapeutics*, 4(2):211–217, March 2004. ISSN 1473-7175, 1744-8360. doi: 10.1586/14737175.4.2.211. URL <https://www.tandfonline.com/doi/full/10.1586/14737175.4.2.211>.
- Clay Beckner, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, and Tom Schoenemann. Language Is a Complex Adaptive System: Position Paper. *Language Learning*, 59:1–26, December 2009. ISSN 00238333, 14679922. doi: 10.1111/j.1467-9922.2009.00533.x. URL <http://doi.wiley.com/10.1111/j.1467-9922.2009.00533.x>.
- Chris Code. Aphasia. In Jack S. Damico, Nicole Müller, and Martin J. Ball, editors, *The Handbook of Language and Speech Disorders*, pages 317–336. Wiley-Blackwell, Oxford, UK, March 2010. ISBN 978-1-4443-1897-5 978-1-4051-5862-6. doi: 10.1002/9781444318975.ch14.
- Bernard Comrie. Chapter 18. Word order in the Germanic Languages — Subject-verb or verb second?: Evidence from aphasia in Scandinavian languages. In Lise Menn, Loraine K. Obler, Gabriele Miceli, and M. O’Connor, editors, *Agrammatic Aphasia*, pages 1357–1364. John Benjamins Publishing Company, Amsterdam, January 1989. ISBN 978-90-272-2045-5 978-1-55619-024-7 978-90-272-7351-2. doi: 10.1075/z.39.30com. URL <https://benjamins.com/catalog/z.39.30com>.
- Holger Diessel. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology*, 25(2):108–127, August 2007. ISSN 0732118X. doi: 10.1016/j.newideapsych.2007.02.002. URL <https://linkinghub.elsevier.com/retrieve/pii/S0732118X07000207>.
- Eva Flanderková. *Čeština v afázii: teorie a empirie*. 2019. ISBN 978-80-246-4246-8. OCLC: 1104141435.
- Heather L. Flowers, Stacey A. Skoretz, Frank L. Silver, Elizabeth Rochon, Jiming Fang, Constance Flamand-Roze, and Rosemary Martino. Poststroke Aphasia Frequency, Recovery, and Outcomes: A Systematic Review and Meta-Analysis. *Archives of Physical Medicine and Rehabilitation*, 97(12):2188–2201.e8, December 2016. ISSN 00039993. doi: 10.1016/j.apmr.2016.03.006. URL <https://linkinghub.elsevier.com/retrieve/pii/S0003999316300417>.
- Susanne Gahl. Lexical biases in aphasic sentence comprehension: An experimental and corpus linguistic study. *Aphasiology*, 16(12):1173–1198, December

2002. ISSN 0268-7038, 1464-5041. doi: 10.1080/02687030244000428. URL <http://www.tandfonline.com/doi/abs/10.1080/02687030244000428>.
- Susanne Gahl and Lise Menn. Usage-based approaches to aphasia. *Aphasiology*, 30(11):1361–1377, November 2016. ISSN 0268-7038, 1464-5041. doi: 10.1080/02687038.2016.1140120. URL <https://www.tandfonline.com/doi/full/10.1080/02687038.2016.1140120>.
- Susanne Gahl, Lise Menn, Gail Ramsberger, Daniel S. Jurafsky, Elizabeth Elder, Molly Rewega, and L. Holland Audrey. Syntactic frame and verb bias in aphasia: Plausibility judgments of undergoer-subject sentences. *Brain and Cognition*, 53(2):223–228, November 2003. ISSN 02782626. doi: 10.1016/S0278-2626(03)00114-3. URL <https://linkinghub.elsevier.com/retrieve/pii/S0278262603001143>.
- Dan Jurafsky. Probabilistic Modeling in Psycholinguistics: Linguistic Comprehension and Production. 2002.
- Petr Karlík. REFLEXIVNÍ ZÁJMENO. In *CzechEncy - Nový encyklopedický slovník češtiny*. 2017. URL <https://www.czechency.org/slovník/REFLEXIVN\unhbox\voidb@x\bgroup\let\unhbox\voidb@x\setbox\@tempboxa\hbox{I\global\mathchardef\accent@spacefactor\spacefactor}\accent19I\egroup\spacefactor\accent@spacefactorZ\unhbox\voidb@x\bgroup\let\unhbox\voidb@x\setbox\@tempboxa\hbox{A\global\mathchardef\accent@spacefactor\spacefactor}\accent19A\egroup\spacefactor\accent@spacefactorJMENO>.
- Václava Kettnerová and Markéta Lopatková. Reflexive Verbs in a Valency Lexicon: The Case of Czech Reflexive Morphemes. 2014.
- M Křen, V Cvrček, T Čapka, and A Čermáková. SYN2015: reprezentativní korpus psané češtiny. *Ústav Českého národního korpusu FF UK*, 2015.
- A. C. Laska, A. Hellblom, V. Murray, T. Kahan, and M. Von Arbin. Aphasia in acute stroke and relation to outcome. *Journal of Internal Medicine*, 249(5):413–422, May 2001. ISSN 0954-6820, 1365-2796. doi: 10.1046/j.1365-2796.2001.00812.x. URL <http://doi.wiley.com/10.1046/j.1365-2796.2001.00812.x>.
- H. Lehečková. Aphasia in linguistics, linguistics in aphasiology. *Casopis pro Moderní Filologii*, 98:7–22, 2016.
- Helena Lehečková. Linguistic theories and the interpretation of agrammatism. *Clinical Linguistics & Phonetics*, 2(4):271–289, January 1988. ISSN 0269-9206, 1464-5076. doi: 10.3109/02699208808985260. URL <http://www.tandfonline.com/doi/full/10.3109/02699208808985260>.
- Helena Lehečková. Manifestation of aphasic symptoms in Czech. *Journal of Neurolinguistics*, 14(2-4):179–208, April 2001. ISSN 09116044. doi: 10.1016/S0911-6044(01)00014-8. URL <https://linkinghub.elsevier.com/retrieve/pii/S0911604401000148>.

- Markéta Lopatková, Václava Kettnerová, Eduard Bejček, Anna Vernerová, and Zdeněk Žabokrtský. *VALLEX 3.0*. 2016. URL <http://hdl.handle.net/11234/1-2307>.
- Nadine Martin, Cynthia K. Thompson, and Linda Worrall, editors. *Aphasia rehabilitation: the impairment and its consequences*. Plural Pub, San Diego, 2008. ISBN 978-1-59756-162-4. OCLC: ocn170203804.
- Lise Menn and Loraine K. Obler. Chapter 20. Cross-language data and theories of agrammatism. In Lise Menn, Loraine K. Obler, Gabriele Miceli, and M. O'Connor, editors, *Agrammatic Aphasia*, pages 1369–1389. John Benjamins Publishing Company, Amsterdam, January 1989. ISBN 978-90-272-2045-5 978-1-55619-024-7 978-90-272-7351-2. doi: 10.1075/z.39.32men. URL <https://benjamins.com/catalog/z.39.32men>.
- R. C. Oldfield and A. Wingfield. Response Latencies in Naming Objects. *Quarterly Journal of Experimental Psychology*, 17(4):273–281, December 1965. ISSN 0033-555X. doi: 10.1080/17470216508416445. URL <http://journals.sagepub.com/doi/10.1080/17470216508416445>.
- Jarmila Panevová and Petr Karlík. REFLEXIVNÍ SLOVESO. In *CzechEncy - Nový encyklopedický slovník češtiny*. 2017. URL <https://www.czechency.org/slovník/REFLEXIVN\unhbox\voidb@x\bgroup\let\unhbox\voidb@x\setbox\@tempboxa\hbox{I\global\mathchardef\accent@spacefactor\spacefactor}\accent19I\egroup\spacefactor\accent@spacefactorSLOVESO>.
- Ilias Papathanasiou and Patrick Coppens. *Aphasia and related neurogenic communication disorders*. Jones & Bartlett Learning, Burlington, MA, second edition edition, 2017. ISBN 978-1-284-07731-5.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL <https://www.R-project.org/>.
- Cai-Li Ren, Guo-Fu Zhang, Nan Xia, Chun-Hui Jin, Xiu-Hua Zhang, Jian-Feng Hao, Hong-Bo Guan, Hong Tang, Jian-An Li, and De-Liang Cai. Effect of Low-Frequency rTMS on Aphasia in Stroke Patients: A Meta-Analysis of Randomized Controlled Trials. *PLoS ONE*, 9(7):e102557, July 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0102557. URL <https://dx.plos.org/10.1371/journal.pone.0102557>.
- Katherine Ross and Robert Wertz. Quality of life with and without aphasia. *Aphasiology*, 17(4):355–364, January 2003. ISSN 0268-7038, 1464-5041. doi: 10.1080/02687030244000716. URL <https://www.tandfonline.com/doi/full/10.1080/02687030244000716>.
- Florian Schwarz and Jérémy Zehr. PennController for Internet Based Experiments (IBEX). 2018. doi: 10.17605/OSF.IO/MD832. URL <https://osf.io/md832/>. Publisher: Open Science Framework.

Ondřej Svoboda. Functions of the Czech reflexive marker se/si. *Research master thesis*, 2014.

Vojtěch Veselý. K slootovorné funkci reflexivních morfémů se, si [On the derivative function of the reflexive morphemes se, si]. 2018.

M.M. Watila and S.A. Balarabe. Factors predicting post-stroke aphasia recovery. *Journal of the Neurological Sciences*, 352(1-2):12–18, May 2015. ISSN 0022510X. doi: 10.1016/j.jns.2015.03.020. URL <https://linkinghub.elsevier.com/retrieve/pii/S0022510X15001513>.

# Appendix A

## List of queries used in the corpus research

This section presents a list of queries used to perform the data extraction from *SYN2015*. The queries were formed in *CQL*, a specialized, but simple query language, that can be used to obtain data from all parts of the Czech National Corpus. The queries are replicable and will yield the same result every time.

- find all verbs in the corpus: [tag="V.\*"]
- find all nouns in accusative, that are syntactically dependent on a verb: [tag="NN..4.\*" & p\_tag="V.\*"]
- find all occurrences of the pronoun "se", that are syntactically dependent on a verb: [word="se" & tag="P.\*" & p\_pos="V.\*"]

# Appendix B

## Preliminary experiment sentences

The table below contains a list of all sentences used in the preliminary experiment.

lemma	sentence type	sentence
hájit	transitive-grammatical	Předseda hájil zaměstnance před komisí
hájit	both object and reflexive	Předseda se hájil zaměstnance před komisí
hájit	reflexive-grammatical	Předseda se hájil před komisí
hájit	no object or reflexive	Předseda hájil před komisí
naklonit	transitive-grammatical	Soudce naklonil hlavu ke straně
naklonit	both object and reflexive	Soudce se naklonil hlavu ke straně
naklonit	reflexive-grammatical	Soudce se naklonil ke straně
naklonit	no object or reflexive	Soudce naklonil ke straně
oddělit	transitive-grammatical	Průvodce oddělil turisty od davu
oddělit	both object and reflexive	Průvodce se oddělil turisty od davu
oddělit	reflexive-grammatical	Průvodce se oddělil od davu
oddělit	no object or reflexive	Průvodce oddělil od davu
ohlásit	transitive-grammatical	Pilot ohlásil přelet přes vysílačku
ohlásit	both object and reflexive	Pilot se ohlásil přelet přes vysílačku
ohlásit	reflexive-grammatical	Pilot se ohlásil přes vysílačku
ohlásit	no object or reflexive	Pilot ohlásil přes vysílačku
ponořit	transitive-grammatical	Sportovec ponořil obličej pod hladinu
ponořit	both object and reflexive	Sportovec se ponořil obličej pod hladinu
ponořit	reflexive-grammatical	Sportovec se ponořil pod hladinu
ponořit	no object or reflexive	Sportovec ponořil pod hladinu
probudit	transitive-grammatical	Opilec probudil kamaráda na konečné
probudit	both object and reflexive	Opilec se probudil kamaráda na konečné
probudit	reflexive-grammatical	Opilec se probudil na konečné
probudit	no object or reflexive	Opilec probudil na konečné
přestěhovat	transitive-grammatical	Ředitel přestěhoval rodinu do Polska
přestěhovat	both object and reflexive	Ředitel se přestěhoval rodinu do Polska
přestěhovat	reflexive-grammatical	Ředitel se přestěhoval do Polska
přestěhovat	no object or reflexive	Ředitel přestěhoval do Polska



rozesmát	transitive-grammatical	Moderátor rozesmál účinkující na jevišti
rozesmát	both object and reflexive	Moderátor se rozesmál účinkující na jevišti
rozesmát	reflexive-grammatical	Moderátor se rozesmál na jevišti
rozesmát	no object or reflexive	Moderátor rozesmál na jevišti
spojit	transitive-grammatical	Kanclér spojil prezidenta s ministrem
spojit	both object and reflexive	Kanclér se spojil prezidenta s ministrem
spojit	reflexive-grammatical	Kanclér se spojil s ministrem
spojit	no object or reflexive	Kanclér spojil se zahraničím
stáhnout	transitive-grammatical	Výrobce stáhl léky z trhu
stáhnout	both object and reflexive	Výrobce se stáhl léky z trhu
stáhnout	reflexive-grammatical	Výrobce se stáhl z trhu
stáhnout	no object or reflexive	Výrobce stáhl z trhu
trápit	transitive-grammatical	Starosta trápil obyvatele nerozhodností
trápit	both object and reflexive	Starosta se trápil obyvatele nerozhodností
trápit	reflexive-grammatical	Starosta se trápil nerozhodností
trápit	no object or reflexive	Starosta trápil nerozhodností
vypravit	transitive-grammatical	Strýček vypravil děti do školy
vypravit	both object and reflexive	Strýček se vypravil děti do školy
vypravit	reflexive-grammatical	Strýček se vypravil do školy
vypravit	no object or reflexive	Strýček vypravil do školy
vzbudit	transitive-grammatical	Tatínek vzbudil sousedy pozdě
vzbudit	both object and reflexive	Tatínek se vzbudil sousedy pozdě
vzbudit	reflexive-grammatical	Tatínek se vzbudil pozdě
vzbudit	no object or reflexive	Tatínek vzbudil pozdě
zaměřit	transitive-grammatical	Vědec zaměřil úsilí na výzkum
zaměřit	both object and reflexive	Vědec se zaměřil úsilí na výzkum
zaměřit	reflexive-grammatical	Vědec se zaměřil na výzkum
zaměřit	no object or reflexive	Vědec zaměřil na výzkum
zapsat	transitive-grammatical	Vrátný zapsal návštěvu u vchodu
zapsat	both object and reflexive	Vrátný se zapsal návštěvu u vchodu
zapsat	reflexive-grammatical	Vrátný se zapsal u vchodu
zapsat	no object or reflexive	Vrátný zapsal u vchodu
zbavit	transitive-grammatical	Lékař zbavil pacienta závislosti
zbavit	both object and reflexive	Lékař se zbavil pacienta závislosti
zbavit	reflexive-grammatical	Lékař se zbavil závislosti
zbavit	no object or reflexive	Lékař zbavil závislosti

# List of Figures

2.1	An example of syntactic annotation in SYN2015. This is an automatically created syntactic tree for a simple sentence <i>Vzbudilo mě klepnání na dveře</i> meaning ‘Knocking on the door woke me up.’ Syntactic functions, such as predicate, subject and object are marked in the annotation. . . . .	18
2.2	Density function of verb reflexivity bias. The two groups with high or low reflexive bias are highlighted. . . . .	20

# List of Tables

2.1	List of lemmas selected for the experiments. The abbreviation <i>ipm</i> stands for instances per million - how frequently on average this lemma occurs in a million word instances in the corpus. . . . .	21
2.2	Plausibility judgement task for experiment I . . . . .	24
3.1	Number of discordant judgements per sentence type . . . . .	28
3.2	Number of discordant judgements (mis)matching condition. Condition is matching if sentence type matches the reflexive bias. Only sentences that are expected to be judged as grammatical are included. . . . .	29