SENTIMENT ANALYSIS IN CZECH

Kateřina Veselovská

ÚSTAV FORMÁLNÍ
A APLIKOVANÉ LINGVISTIKY

STUDIES IN COMPUTATIONAL
AND THEORETICAL LINGUISTICS

Kateřina Veselovská

SENTIMENT ANALYSIS IN CZECH

Published by the Institute of Formal and Applied Linguistics
as the 16th publication in the series
Studies in Computational and Theoretical Linguistics.

This book has been printed with the support of the grant GA15-06894S of the Grant Agency of the Czech Republic and of the institutional funds of Charles University.

Printed by Printo, spol. s r. o.

ISBN 978-80-88132-03-5

Motto:

Sentiment without action is the ruin of the soul.

— Edward Abbey

to all emotional people

Contents

Acknowledgement

Searching for emotions is something we all do constantly, so this book should definitely not be seen as the end of my research. However, one important chapter is over and there is no study on emotional meaning without a little emotional acknowledgement. Therefore, I make use of emotional language I describe further in this work to thank the following people:

Thanks to Professor Eva Hajičová for her constant courage to explore new topics in computational linguistics, no matter how peculiar they may seem at first sight. For her will to lead me through this adventurous way. For all the puckered eyebrows and important comments she has given to me. For being inspirational, supportive and always young at heart.

Thanks to Jana Šindlerová, Honza Hajič Jr., Aleš Tamchyna and Honza Mašek, the sentimental team, for their contribution to papers this publication would have never come into being without. Special thanks to the guys for the implementation works. I highly appreciate this friendly cooperation.

Thanks to The Office 423 – the loveliest bunch of nerds ever. People who shared their feelings, their stories and their snacks with me in good times and in bad. Each one of you has taught me a lot, guys. It was a good ride.

Thanks actually to all my colleagues from the Institute of Formal and Applied Linguistics. Not only for their ideas concerning my work and the great times we had at the conferences and other events, but also for being so dedicated to scientific research. Of course you are brilliant minds. But you are also exceptional people.

Thanks to Mrs. Libuška Brdičková, the most lively and patient secretary of all times, for both the paperwork and personal advice. I would have never made it without you. Seriously.

Thanks to Vláďa Kuboň, Jarka Hlaváčová, Ondra Plátek, Markéta Lopatková and Stefan Gindl for their feedback – you have saved me from a lot of negative emotions and disappointments concerning not only commas after having this printed.

Thanks to Eda Bejček for his patience when helping me out with the LaTex issues.

Thanks to my students for being curious.

Thanks to Deloitte crew for letting me put my sentiments into practice.

Thanks to my best friend Veronika and to all my ladies. This work takes just a short period of time in the history of our friendships, but your support turned out to be crucial many times.

Thanks to Jirka for being The One. For supporting me, helping me, sharing his passions, his thoughts and his life with me over many many glasses of wine. I will cherish every single minute of our discussions till the end of my days.

Finally, thanks to my parents Zuzana and Petr, the Best Team, who have invested in me all the love one can dream of. Who have taught me most about emotions and who have always been there for me when I needed to discuss my feelings. Who have always trusted in me, whatever I did. Also, thanks to the rest of my family who have repeatedly showed a sincere effort to understand what I am actually working on at all the family gatherings. Děkuju – mám vás ráda.

Once again, thank you all for sharing your smiles, your wisdom and first of all your emotions with me. I will never be grateful enough to have met you.

This study is based on projects investigated at the Institute of Formal and Applied Linguistics, Charles University in Prague. It was supported by the Grant Agency of the Czech Republic (project GA15-06894S) and by the Ministry of Education of the Czech Republic (project LH14011). It has been using language resources and tools developed, stored and distributed by the LINDAT/CLARIN project of the Ministry of Education of the Czech Republic (project LM2015071).

1
Introduction

Expressing emotions is one of the major functions of human language. Conveying feelings, moods and affects is common to all humans. It is one of the first things children learn to articulate, since development of emotions precedes the development of thinking. Emotions involve the way people feel. By expressing emotions we show our opinions and feelings towards others and the world around us. Emotions constitute the basis of communication and as such, they can be considered foundations of interpersonal relationships.

Not to be able to express or comprehend emotions is often perceived as a deficiency or even a disability. Life without emotions is perceived as empty. On the other hand, emotions can also be misused for the purposes of manipulation e.g. in media or advertising. Since language is one of the main tools for expressing emotions, it is important to understand which linguistic means are used to communicate emotional meaning, and thus to get closer to the automatic processing of emotional language.

Computational processing of emotions in language is focused on automatic detection, extraction and classification of emotional meaning. It helps us to learn about emotions, entities which expresses emotions and entities towards which emotions are expressed, on larger datasets of different nature. Using methods of natural language processing, we can find generalized patterns of emotional expressions, usable in various practical applications.

In the present study, we have two main goals. The first one is to give a compact description of basic means of emotional language in Czech. The second one is to employ the findings concerning emotional language in computational applications, namely sentiment analysis, i.e. automatic extraction of emotions from text.

The outcomes of the publication are aimed at better understanding of Czech both as a language system and in communication. The results of this work are also expected to be applicable not only in sentiment analysis (as described in detail in Chapter 9), but also in many other areas of natural language processing, such as question-answering, recommendation systems, automatic summarization of a text, automatic dialogue systems or emotionality modelling.

1.1 Emotional Meaning

The central term of this monograph is *emotional meaning*. The study is based on the assumption that there is a distinction between descriptive and emotional meanings of language structures (Hare, 1961). Whereas descriptive meanings express objective facts (*this car has four wheels*), emotional meanings convey subjective attitudes instead (*this car is ugly*).[1]

Although there have been some attempts to define emotional meaning cross-disciplinary (see Fontaine et al., 2013), the concept of emotional meaning usually differs with each field it concerns, e.g. with psychology, sociology, cultural anthropology etc. In the linguistic analysis of emotional meaning, the definition also depends on the particular linguistic discipline. Moreover, we need to take into account the fact that emotions are rather complex phenomena on their own. Therefore, we decided to approach emotional meaning as a compositional meaning, consisting of many linguistic aspects. At the moment, we can only say with certainty what emotional meaning is *not*: it is not an objective description of any event, situation or mental state.

The present study surveys the ways emotional meaning is modelled in Czech using means from different layers of language. We observe the mutual interaction of lexical and structural meaning and the role of specific areas of linguistic research, such as the system of evaluative expressions' valency, idiomatics etc. For details, see Section 1.4.

Also, since we believe that evaluation always means expressing emotions and emotions usually have an evaluative nature (meaning that they can mostly be categorized within the scale between positive – negative), we use the terms *emotional* and *evaluative* interchangeably, as also suggested by e.g. Čáp (1980), p. 97, or Malrieu (2002). When distinguishing between the terms *emotional* and *expressive*, we agree with Mikulová (2010) that although these two terms are very close, there is a difference. Whereas expressivity can be perceived as a general term including all the utterances in which the speaker deviates from the norm (formal, lexical, stylistic etc.), emotionality can be seen as an expression of feelings. We are aware of the fact that this approach is rather simplifying given the complex nature of the investigated phenomenon. However, we deliberately use this terminology in accordance with both linguistic and computational emotion research tradition.

1.2 Motivation

Current theoretical and computational linguistics witnesses a massive increase of interest in emotional language research. The area has received much attention over the past decade with the rise of Web 2.0 (see Wiebe et al., 2004) and with the newly-emerged user-generated evaluative content obtained from weblogs or social networks. These data serve as a basis not only for studies in pragmatics and cognitive linguistics (see e.g. Eisenchlas, 2011; Paradowski and Jonak, 2011), but also for numerous natural language processing experiments, including information extraction and text categorization (see Liu, 2012). However, there are only few attempts to describe emotional

[1] However, we are aware of the fact that one sentence can sometimes be perceived as either neutral or positive/negative, depending on a situation. For example, in the sentence *this car costs $30,000*, the price can be considered average/high/low depending on a car type, country, whether the speaker is the buyer or seller etc.

meaning in morphologically rich languages (e.g. Jang and Shin, 2010). Though there are many resources of Czech emotionally oriented texts, and corpus linguistics offers both data and tools to handle it, there is still lack of systematic approaches to Czech emotional language.

Generally, there are many practical applications of emotional meaning research, from automatic processing of product reviews to prediction of market trends or election outcomes. Emotional language can be employed in better understanding of people's sentiments in public opinion surveys or social media monitoring. Moreover, it can be applied in forensic linguistics, namely in cyberbullying (i.e. harassing other people using social networks) or hate speech detection. Also, it plays an important role in diverse psycholinguistic or even neurolinguistic applications helping people with disorders involving emotion comprehension, such as autism or Asperger's syndrome.

1.3 Objectives

The main goal of this monograph is to build a comprehensive and coherent model of emotional language means in written Czech. Making use of the available theoretical and empirical resources, the study addresses the following research tasks:

1. The main task is to describe all the components construing emotional meaning in Czech, including both lexical and grammatical means. Not only do we focus on specific Czech features of emotional expressing, but we also explore components of emotional language which can be investigated crosslinguistically, comparing our findings with the work broadly described in Chapter 2.
2. Since emotional meaning is one of the main constituents of the interpersonal function and interpersonal communication is the main function of language shared by all natural languages, we explore it also in terms of pragmatics and theory of communication.
3. Although Czech corpus linguistics has been gathering large data over a long period of time, these data remain unexplored considering linguistic means of expressing emotions. Moreover, data from the Internet reviews have been collected, annotated and made accessible for research only recently. We use manually annotated data from various sources, including treebanks, plain text corpora, and online reviews from real-life Internet users (see Chapter 8) to find evidence for our hypotheses concerning emotional language.
4. This work also contributes to the Prague Dependency Treebank project by adding a new layer for capturing phenomena from the field of pragmatics, namely evaluation. This is done by implementing evaluative items from the Czech Subjectivity Lexicon (see Section 8.1) to the treebank and refining it manually, following the newly established guidelines for dependency annotation of evaluative sentences (see Appendix D).

We are aware of the fact that intonation, tone, rhythm, stress and even facial expressions and gestures play an important role in expressing of emotions. However, this area is rather wide on its own and exceeds the scope of this work which builds exclusively on written data. Therefore, the study does not focus on emotional language means from the field of prosody and multidimensional analysis taking into account non-linguistic emotional means. Also, we do not describe topic-focus articulation in emotional structures. Moreover, we do not concentrate on surveying emotional language from neurolinguistic or psycholinguistic perspective. However, we believe that the outcomes of our research could be applied in broader multidisciplinary investigation of emotions, e.g. in connection with mental disorders or mental well-being research.

1.4 Roadmap

This publication has two major parts, **Linguistic Structure** (Chapter 3 – Chapter 7) and **Computational Applications** (Chapters 8 and 9). The former focuses on linguistic issues of emotional meaning whereas the latter describes practical usage of findings concerning the means of emotional language. The monograph answers specific research questions in the following chapters:

- In Chapter 2, **Theoretical Background**, we introduce basic linguistic and computational approaches to the research of emotional meaning and give a brief overview of the data this work builds upon.
- In Chapter 3, **Lexical Aspects**, we discuss the lexical nature of Czech evaluative items, whether inherently or adherently emotional. We discuss frequently used augmentatives, diminutives, vulgarisms and other emotional expressions, making use of the Czech Subjectivity Lexicon (see Section 8.1). Also, we focus on idiomatic and euphemistic statements and alternative ways of expressing emotions on the lexical level.
- In Chapter 4, **Morphosyntactic Aspects**, we survey the morphology of Czech emotional items and their part-of-speech categorization. We investigate the influence of the part of speech categories of evaluative expressions on emotional structures and point out the role of gender mismatch in evaluation. Also, we focus on the syntactic nature of emotional structures. Not only do we describe their basic patterns, but we also explore compound emotional structures, including bigger chunks of text. Moreover, we pay special attention to the issues of negation, which can be crucial for the polarity of emotion expressed.
- In Chapter 5, **Semantic Aspects**, we analyze the semantic core of evaluative language. We pay special attention to emotional verbs. First, we investigate their valency characteristics, considering the long tradition of the research on Czech in the domain (Panevová, 1980). Second, we suggest their semantic classification, inspired by Levin (1993). Also, we investigate the discourse structure of

larger evaluative text spans, namely the relationship between individual sentences. We pay special attention to discourse connectives, attitude markers and coreference relations.
- In Chapter 6, **Pragmatic Aspects**, we describe the position of evaluative structures in the field of pragmatics. We examine the role of the context of emotional utterances, explain the basics of irony and sarcasm and situate the role of emotional expressing within the speech acts theory. Moreover, we draw the relationship between emotional expressing and conversational maxims.
- Chapter 7, **Formal Representations of Emotional Structures**, develops the ways of representing evaluative structures by means of two linguistic frameworks, namely Functional Generative Description (Sgall, 1967) and Construction Grammar (Fried and Östman, 2004). For Functional Generative Description, we propose annotation scheme of evaluative structures in the Prague Dependency Treebank. Within the Construction Grammar, we address a new type of construction, the Subjective Construction, and integrate it into the construction formalism, proposing also relevant attributes.
- In Chapter 8, **Creating Evaluative Resources**, we describe how we created and evaluated the data for further experiments in the field of sentiment analysis. The chapter concerns not only the Czech Subjectivity Lexicon, but also manually annotated data, problems in annotation and annotator statistics.
- Chapter 9, **Sentiment Analysis**, focuses on practical applications of emotional language research. We explain the methods and discuss the results we reached for automatic polarity detection in Czech and automatic opinion target identification in Czech, English and several other languages.
- Appendix A, **Data Overview**, provides statistics about the data size, origin, format etc.
- Appendix B, **Annotation Guidelines**, presents the guidelines based on which the manually annotated data used in this study was processed.
- Appendix C, **Computational Basics**, explains the basic natural language processing metrics used in the study.
- Appendix D, **TrEd Extension**, describes how we created an extension of the tree editor TrEd to complement the data by the annotation of emotional structures.

1.5 Typographical Conventions

The following list explains typographical conventions used in this book:
- in-text examples and terms are in italics
- English translations are in single quotes
- positive expressions are marked as [positive expression]+
- negative expressions are marked as [negative expression]–
- other expressions relevant to a given discussion are marked as [expression]

Unless stated otherwise, the examples are created based on author's Czech native competence for the purposes of the study.

2

Theoretical Background

As the linguistic part of this study describes in detail, emotional meaning can be conveyed on different levels of language description. Therefore, it is understandable that the problem of expressing emotions has been discussed in linguistic literature from various points of view, emphasizing various components of language meaning.

In this chapter, we introduce both linguistic and computational approaches to the research of emotional meaning. Whereas linguistic theories deal mostly with language structure and theoretical models, computational approaches aim for real-life applications. Linguistic approaches to emotions are related to all major areas of language description, from grammar and lexicon to pragmatics. However, focus on one aspect of emotional expression does not mean that the authors disregard the other components.

While pragmatic approaches view language primarily as a communication tool and emphasize language use, other theories concentrate more on particular linguistic aspects expressing emotional meaning. These aspects are then commonly used in automatic detection of emotions and other computational disciplines concerned with emotional meaning, as described in Chapter 9.

2.1 Linguistic Approaches

In this section, we introduce particular approaches to emotions in language, from more general to more specific ones.

2.1.1 Pragmatics

Theories stemming from pragmatics are above all based on observing language as a tool of human interaction. Therefore, they pay special attention to both communication and situational context, which, in the framework of pragmatics, also contributes to meaning. The following approaches define emotional meaning functionally, i.e. with respect to the communication purpose it aims for.

The earliest approaches of this sort are those connected to philosophy and especially to the area of ethics. R. M. Hare (1961), a moral philosopher, distinguishes between evaluative and descriptive meaning, constructing the former as based on the latter. Hare determines objects as good/bad members of a given class functionally, using his strawberry example: He argues that if we know the descriptive properties of a strawberry and we also know the criteria to call a strawberry a good one then we can say about a particular strawberry whether it is a good one or not. 'The goodness in strawberries' is defined by being good members of a class (i.e. being suitable for their function as strawberries – being large, red, juicy etc.). In other words, Hare says that even descriptive constructions can convey evaluative (and so emotional) meaning in certain context. The prototype theory by Rosch (1975) works on the same assumption. The author claims that some members of a category are more central (i.e. 'better' in representing the given category) than others.

Whereas Hare and Rosch focus more on the semantic context, important for perceiving a given term as functioning well, most of the functional approaches focus on the term *function* within the meaning of communication. This branch is represented by Appraisal Theory. Appraisal Theory, developed by Martin and White (2005), is based on the tradition of the Systemic Functional Linguistics (SFL, Halliday, 1973). SFL regards language as inherently functional, consisting of three major components or *metafunctions* which work rather differently (see Halliday, 1994):

- *ideational*, i.e. reflective, by which we make sense of reality
- *interpersonal*, i.e. active, used to act on the others
- *textual*, which "breathes relevance into the other two" (Halliday, 1994, p. xiii), i.e. organizing instances of discourse

From the point of view of Appraisal Theory, the most important of the three is the interpersonal function, i.e. a function construing relationships between people and using, among other means, emotional utterances. *Appraisal* in the usual meaning of the term, i.e. the act of judging the value, condition, or importance of something, is introduced as one of the three components construing the interpersonal meaning (together with involvement and negation). Appraisal is further divided into three interacting domains: *attitude*, *engagement* and *graduation*. Whereas attitude is concerned with emotional reactions, engagement deals with sourcing attitudes in the discourse. Graduation is then defined as intensifying the feelings. Besides, the attitude system is divided into *affect*, *appreciation* and *judgment*, depending on what kind of meaning the speaker uses. For clarity, see Fig. 2.1.

We can illustrate the difference between these three types of attitude by the following invented examples:

- *I love their strawberry cakes!* – affect, attitude expressed by emotional response
- *Their strawberry cakes have balanced flavors.* – appreciation, attitude expressed by assessment of taste quality
- *The baker really knows his job!* – judgment, attitude expressed by construing the view of the baker as a social being doing his job well

In this study, we make use of this fine-grained attitude classification in Chapter 5.

In addition to the attitude system, Appraisal Theory develops and extends the SFL account of interpersonal meaning by attending to three axes along which the speaker's/writer's intersubjective stance may vary, namely:

$$\left\{\begin{array}{l} \text{Appraisal} \left\{\begin{array}{l}\text{Attitude}\left\{\begin{array}{l}\text{Affect}\\ \text{Appreciation}\\ \text{Judgement}\end{array}\right.\\ \text{Engagement}\\ \text{Graduation}\end{array}\right.\\ \text{Involvement}\\ \text{Negation}\end{array}\right.$$

Figure 2.1: Interpersonal Function

- *affect* – "the means by which writers/speakers positively or negatively evaluate the entities, happenings and states-of-affairs with which their texts are concerned" (Martin and White, 2005, p. 2).
- *modality* in a broader sense – speaker/writer certainty and questions of how the textual voice positions itself
- *intensification* – providing a framework for the description of how speakers/writers increase and decrease the force of their assertions and how they sharpen or blur the semantic categorizations with which they operate

Apart from strictly formal theories, Appraisal Theory is known for being focused on the meaning of utterance rather than on its form. It explores 'semantic resources', i.e. which kinds of meanings are constructed, with less focus to particular linguistic features. We partly adopt this approach in Chapter 6 dedicated to pragmatics in the second part of this study.

Within the usage-based perspective on language, we can also find the term *stance*, in the meaning of evaluation. This terminology is widely used e.g. in Englebretson (2007). In this work, stance is understood as a heterogeneous term related to different scientific fields and thus a topic of interactional research. Stance-taking is considered rather an activity than a set of linguistic features or lexical expressions.

Stance-taking in spoken dialog is described by the Stance-triangle Theory introduced by Du Bois (2007). Du Bois introduces the stance triangle as a way of representing the components of the stance act, and articulating their multiplex interrelations. He consistently distinguishes between evaluating subject, evaluated object and evaluation as the item giving value to the object, as illustrated on the following invented example:

(1)	Susan	likes	strawberries.
	EVALUATING SUBJECT	EVALUATION	EVALUATED OBJECT

This 'triangle' is crucial for most of the present-day computational models of evaluation. In the present study, we also use this distinction when describing semantic participants of evaluative structures in the linguistic part or when detecting subjects and objects of evaluation within the rule-based systems for opinion target identification in the computational part (see Section 9.3).

Besides the explicitly mentioned participants of evaluative structures, we also need to take into account the possibility that the evaluating subject can also be the author of a given text. This case is described e.g. in the Metadiscourse Theory, where *metadiscourse* means "self-reflecting linguistic material referring to the evolving text and to the writer and imagined reader of that text" (Hyland and Tse, 2004, p. 156). Hyland and Tse (2004), following Beauvais (1989) and Crismore and Farnsworth (1990), inter alia, work with the assumption of interaction between the writer and the reader where *attitude markers* (such as *unfortunately*, *I agree* or *surprisingly*) are viewed as items expressing the interpersonal function of metadiscourse.

These attitude markers are broadly used also in Critical Discourse Analysis (CDA, Fairclough, 1995). CDA views language as a form of social practice and focuses on the ways social and political beliefs are reproduced in text and talk. In addition to the international scene dealing with emotional meaning in CDA (see Graham, 2003), there is also the Czech CDA research represented e.g. by Nekvapil (2006). Moreover, the last decade has seen a rise of minor studies dealing with domain dependent issues of emotional meaning in various text types (Brdečková, 2006; Pazderová, 2007). As discussed in Chapter 6, domain and situational dependency are the typical features of evaluative meaning belonging to pragmatics.

2.1.2 Lexicon and Grammar

In the previous section, we introduced approaches considering primarily the content, i.e. the meaning of emotional utterances. However, there are a number of studies focused rather on form, i.e. on linguistic properties of evaluative structures. In this section, we consider approaches which deal with the influence of lexicon and grammatical structure on emotional expressing. Here, we take into account grammar in the narrow sense, discussing theoretical views on morphology and syntax from which we proceed in Chapter 4 of the linguistic part of this study.

Conrad and Biber (2000) inquire into particular evaluative items, namely stance adverbials, and take into consideration both the grammatical form and the kind of stance expressed. For example, they identify *probably* as typical for epistemic stance which indicates how certain the speaker or writer is, or where the information comes from. Besides, they apply the quantitative approach developed also by Hunston (2011), on large amounts of data. When comparing three kinds of registers (conversations, academic writings and news reports), they are able to quantify e.g. the proportion of different parts of speech in different evaluative texts, confirming the hypothesis that emotional expressing is domain-dependent, i.e. it differs depending on particular text type.

More fine-grained morphological analysis is proposed by Kučerová (2000, p.c.), who suggests that there exists a specific type of agreement indicating a presence of expressive meaning in Czech. Kučerová gives example of the following sentence:

(2) Petr, kluk jeden pitomá, začal zpívat.
Peter$_{\text{MASC}}$ boy$_{\text{MASC}}$ one$_{\text{MASC}}$ silly$_{\text{FEM}}$ started to-sing
'Peter, a silly boy, started to sing.'

She argues that weakening of the specified grammatical gender (masc. *kluk* – 'boy' in (2)) by a different gender (fem. *pitomá* – 'silly' in (2)) is motivated by the expressive function of this syntagma (and thus by emotional meaning).

Apart from isolated studies investigating particular morphological realizations of emotional meaning, there exists a distinct area of evaluative morphology (see e.g. Stump, 1993; Bauer, 1997), exploring morphological rules which serve to express diminution or augmentation, endearment or contempt, and distinguishing e.g. *pejorative morphemes* (like *-ard* in English in *coward*, *dullard*, *sluggard*, *drunkard* etc.). In this study, we discuss these principles in Chapter 3, devoted to lexical meaning of emotional expressions.

Concerning syntactic features of emotional meaning, attention was given to issues of the syntactic role of negation in evaluative sentences, since it usually switches their polarity. This study handles negation following e.g. Veselovská (2011a) or Yoon (2011).

The current Czech research into syntactic patterning of evaluative constructions (see Šindlerová et al., 2014) suggests that the structure of evaluative utterance is tightly connected to verbal valency. In this study, we investigate verbs from the Czech Subjectivity Lexicon, i.e. a list of positive and negative evaluative items (Veselovská, 2013), using a system of verb classes inspired by Levin (1993). Levin distinguishes *verbs of psychological state* (*amuse*, *admire*, *marvel* and *appeal verbs*), *verbs of desire*, *judgment verbs* or *verbs of assessment*, using classification based on alternations they participate in. However, for a number of languages, emotion verbs have been classified for the most part semantically and syntactically – see e.g. Belletti and Rizzi (1988) or Mathieu et al. (2010). The syntactico-semantic properties of Czech emotional utterances based on verb classification are described in Chapter 4.

In addition to research of evaluative lexical items based on communication discourse or preselected authentic texts, there is an influential field of research based on corpus research, carried out e.g. by Hunston (2011). Hunston investigates lexical semantics and phraseology of evaluative expressions, i.e. their tendency to occur in some environments more frequently than in others. She argues that phraseology, in a broadly defined sense, plays a number of roles in the study of evaluative language. Moreover, Hunston is concerned with the quantifying point of view, giving statistics of concordances of frequent evaluative items with a special focus on modal-like expressions, including traditional modal verbs. Corpus analysis surveying evaluative language is also performed e.g. by Biber et al. (1999), who recognize lexical bundles that express stance (like *in fact*, *no doubt* etc.).

In the Czech linguistic tradition, the corpus approach to lexicon is pursued primarily by Čermák (2007), who focuses on idiomatic expressions of emotional meaning. Čermák refers to evaluative function of phrasemes and idioms, claiming that "most phrasemes are distinct means of expressing assessments, actually they are the richest source of it in the system" (Čermák, 2007, p. 116). Čermák distinguishes several categories of Czech phrasemes and idioms (namely grammatical, nominal, modification and verbal) and gives some inspiring examples of evaluative idioms for each category mentioned in the second part of this book. His definitions and examples serve as one of the bases for Chapter 3.

2.1.3 Stylistics

To the best of our knowledge, emotional language has not been in the centre of attention of Czech linguistics so far. However, evaluative use is sometimes mentioned in various Czechoslovak and Czech handbooks of stylistics, mostly in connection with pragmalinguistics and performative language. Most of the authors also take into account expressive language in general, but with no special stress on emotional language as a distinctive category. Zima (1961) in his monograph on expressivity tries to classify linguistic means on the lexical and stylistics level. Grepl (1967) accesses emotional meaning as being an actualized use of normally neutral linguistic means on different linguistic layers, with expressivity standing at the lexical level. Mistrík (1985) and Jelínek (1995), emphasize the meaning of expressive lexicon in connection with different styles. They also distinguish between objective and subjective styles according to the degree of expressivity. Čechová et al. (1997), on the other hand, pay attention to the context and stylistic homogeneity of expressive words. Hoffmannová (1997) also mentions the role of the context and the fact that it can change the meaning of potentially emotional expressions.

Moreover, there exist several isolated studies on this topic in the theoretical domain. J. V. Bečka (1975) uses the same terminology as Wilson (2008) in a different sense. He considers the term *subjective* as referring to the author of the given text, whereas the term *polar* is used when the author communicates with the addressee. On the other hand, Bečka still reflects some basic facts about expressive (including evaluative) items and demonstrates some direct and indirect ways to express evaluation in Czech. Expressive language, e.g. vulgarisms, as the opposite of cultivated language, is mentioned in Daneš (1969) and occasionally investigated e.g. in Krčmová (1981), Čmejrková and Hoffmannová (2003) or Hoffmannová and Müllerová (2007).

2.1.4 Crosslinguistic Comparison

Currently, the research into emotions in language, their impact on linguistic structure and a crosslinguistic comparison of their means of expression concerns a variety of languages and topics, as attested by collections of elaborate scientific works. Baider and Cislaru (2014) study the way different languages encode emotional information and the core role emotions play in linguistic structure and learning, emphasizing language use in social contact e.g. in English, Greek, French or Japanese. Moreover, some of the studies in this issue focus on prosodic aspects of emotional expressing, which has been a rather unheeded topic so far.

Fontaine et al. (2013) describe affective modelling as a cross-disciplinary problem, construing emotional meaning within different fields, from psychology and linguistics to cultural anthropology, sociology or history. As a part of a crosslinguistic comparison, they investigate the ways in which emotional terms translate into other languages, pointing out that there can exist significant differences in how these seemingly similar emotional terms can be applied across various languages. Also, they present a methodological approach based on the Component Process Model (CPM, Scherer, 2001). CPM postulates that emotions are processes triggered by goal-relevant events and consists of synchronized activity of several human sub-systems (cognitive appraisals, bodily reactions, expressions, action tendencies, and feelings). The authors introduce a new instrument to assess the meaning of emotional terms, namely the GRID questionnaire. The questionnaire consists of a grid of 24 emotional terms spanning the emotional domain and 142 emotional features that operationalize five emotional components introduced in CPM. Based on the answers, they describe differences in expressing emotions in various languages and cultures in general.

Mikulová (2010) draws on Appraisal Theory (see Section 2.1.1) background and searches for individual points of agreement and difference between expressing evaluation by linguistic means in Czech and German. The author distinguishes expressivity from emotionality, and emotionality from emotionalizing processes, claiming that evaluation is the headstone of emotionality.

In this study, we also apply a contrastive view when comparing some of the basic means of emotional language in Czech and English. Also, we employ our findings in practical applications, e.g. when building opinion target detection systems for both Czech and English (see Section 9.3.1 and Section 9.3.2).

2.1.5 Integration Approach

In the present study we focus on integrating different layers of language description, inspired by Klenner et al. (2012). Klenner et al. (2012) build on a compositional theory of clause-level polarity determination. Rather then taking into account only surface strings, they argue that proper polarity annotation of complex phrases requires access to their syntactic structures.

Other linguistic theories consider also multiple parts of linguistic structure when modelling emotional meaning and even try to propose its formal representation. In this work, we benefit from Construction Grammar (Fried and Östman, 2004), building on the growing body of construction grammar research concerning the expressions of subjective judgment, broadly defined (see e.g. Fried and Östman, 2005; Matsumoto, 2008; Terkourafi, 2010). Also, we draw on Functional Generative Description as proposed by Sgall (1967) and his school and extensively elaborated up to now (see e.g. Hajičová, 2012). Based on these theories, we suggest formal representations of emotional structures, see Chapter 7.

There are a number of papers dealing with compositional semantics in the field of sentiment analysis. Whereas Choi and Cardie (2008) indicate that simple heuristics based on compositional semantics can perform better than learning-based methods that do not incorporate compositional semantics, Moilanen and Pulman (2007) explain sentiment classification of grammatical constituents in a quasi-compositional way.

2.2 Computational Approaches

In addition to traditional linguistic theories, emotional language is also being explored by computational linguists concerned with the field of sentiment analysis. Sentiment analysis, also known as opinion mining, is an automatic detection of a positive or negative polarity, or neutrality of a sentence (or, more broadly, a text). This is mostly done by detecting *evaluative items*, i.e. words or phrases inherently bearing positive or negative value. These words (phrases) are collected in subjectivity lexicons, i.e. in lists of lexical items bearing an inherently positive or negative value. The implementation of polarity items from the subjectivity lexicon into the data is the first step towards sentiment analysis.

2.2.1 Subjectivity Lexicons

The issue of building a subjectivity lexicon is described e.g. in Taboada et al. (2011) or more specifically in Banea et al. (2008a). Here the authors use a small set of subjectivity words and a bootstrapping method of finding new candidates on the basis of a similarity measure. The authors get 4,000 top frequent entries for the final lexicon. A different method for obtaining a subjectivity lexicon – translation of an existing foreign language subjectivity lexicon – is described in Banea et al. (2008b). Another method is to use already existing lexical databases, such as WordNet (Miller and Fellbaum, 1998) or General Inquirer (Stone et al., 1966), to extract the subjectivity lexicon using a small seed words set and a bootstrapping method. A widely known example of this approach is SentiWordNet (Esuli and Sebastiani, 2006).

Mostly, the authors use subjectivity lexicons and sentiment analysis in general to improve machine translation systems. They are interested in how the information about polarity should be transferred from one language to another, if the polarity could differ in the corresponding text spans and if it is possible to compile a subjectivity lexicon for the target language during the translation. There are a number of papers dealing with the topic of building the subjectivity lexicons for particular languages (see e.g. Bakliwal et al., 2012; De Smedt and Daelemans, 2012; Jijkoun and Hofmann, 2009; Perez-Rosas et al., 2012). Following the work done by Veselovská (2013) and Veselovská et al. (2014), we present the process of building a Czech Subjectivity Lexicon and describe its properties in Section 8.1.

2.2.2 Morphosyntax

Apart from lexical features of evaluative texts, some researchers also take into account their morphosyntactic properties. Wiebe et al. (2004) explore learning evaluative language from corpora, pointing out not only lexical, but also syntactic means of expressing emotionality. The authors also underline the role of morphological context. They use the so-called *subjectivity clues* to identify emotional words in a corpus of Wall Street Journal texts. Such clues are for example: unique occurrence of the word in a corpus, unusual collocation or distributional similarity.

The manual and automatic identification of linguistic expressions of the so-called *private states* (speaker's attitudes) is explored also in Wilson (2008). Wilson recognizes sentence polarity, intensity and attitude as important features of subjectivity expressions, with attitudes bearing two other important, rather syntactic markers of sentiment, namely *source* (the evaluating author) and *target* (the evaluated entity).

Although almost all studies of the topic mention the impact of syntactic structures, the actual research is devoted to separate studies of individual syntactic phenomena (such as Narayanan et al., 2009), some of them using syntactic properties of evaluative structures as a basis for rule-based systems for opinion target identification (see e.g. Qiu et al., 2011; Veselovská and Tamchyna, 2014). However, only a few projects use syntactically annotated corpora. In Section 9.3, we follow the work of Qiu et al. (2011) who learn syntactic relations from dependency trees.

Concerning syntax on the text level, many studies (see e.g. Somasundaran et al., 2008, 2009) are devoted to the mutual dependency between evaluative language and discourse relations annotation. They point out that sentiment analysis is useful for the identification of discourse relations in the text, and vice versa. This topic is further investigated in Section 5.2 of this study.

2.2.3 Neural Networks

Following the latest trends in natural language processing and computer science in general, many researchers explore possibilities of employing deep learning and neural networks also in sentiment analysis tasks. Mostly, the authors focus on sentence-level sentiment analysis of short texts. Dos Santos and Gatti (2014) use Stanford Twitter Sentiment corpus (Go et al., 2009) to train deep convolutional neural network on tweets. Kumar and Rani (2016) propose a probabilistic neural network with a self-adaptive approach to perform sentiment analysis also on Twitter utterances. Concerning applications of deep learning in sentiment analysis on longer texts, it has been widely explored e.g. by Socher et al. (2013), who use recursive deep models for semantic compositionality and introduce Sentiment Treebank. Since it often turns out that traditional approaches suffer from not paying attention on domain-dependent types of sentiment expression, Vo et al. (2017) suggest to combine deep learning with domain knowledge to improve efficiency and accuracy of the systems. As for Czech, in Lenc and Hercig (2016) the authors present the first attempt at using neural networks for sentiment analysis on Czech data showing promising results. The deep learning approach has also inspired Tamchyna and Veselovská (2016), c.f. Section 9.3.3.

2.2.4 Other Studies

Apart from particular studies mentioned above, we draw on handbooks on both sentiment analysis and opinion mining (Pang and Lee, 2008 or Liu, 2012) and natural language processing methods in general (Indurkhya and Damerau, 2012). Concerning the specifics of the Czech research in sentiment analysis area, we mostly refer to Červenec (2011), Veselovská (2012) and Habernal et al. (2013). For further references and more details about sentiment analysis resources, see Chapter 8 and Chapter 9.

2.3 Data Resources

The present monograph builds upon the following data resources:

- large publicly available corpora
- Czech Subjectivity Lexicon
- manually annotated data
- online data resources
- other mostly non-digital resources

Concerning the large publicly available corpora, for domain-independent plain text examples, we use the Czech National Corpus (CNC, `http://www.korpus.cz`), namely the collection of SYN (i.e. synchronous) corpora (version 3) containing 2.2 billion words from different types of corpora (journalistic, fiction, technical) and the parallel corpus InterCorp (version 7) consisting of data from 38 languages. In this study, we use the Czech and English language pair. The great number and variety of lemmatized and morphologically annotated texts makes it possible to use a quantitative approach and complex linguistic analysis e.g. of lexical properties of emotional meaning (see Chapter 3). We gather the CNC potentially evaluative texts by querying the evaluative items from SubLex 1.0 (see Section 8.1).

For syntactico-semantic analysis, we make use of the Prague Dependency Treebank (PDT), an annotated corpus of Czech texts (Hajič et al., 2006; for guidelines, see

Mikulová et al., 2006; `http://ufal.mff.cuni.cz/prague-dependency-treebank`) containing 2 million words. PDT consists of more than 3,000 documents containing almost 50,000 sentences. The documents are annotated considering the morphological, surface syntactic and tectogrammatical[1] structure of sentences, including their information structure. In about 90% of the text data the basic anaphoric links have already been labeled, which is crucial for distinguishing sources and targets of the evaluation expressed. PDT is an advantageous source for analyzing the syntactic and semantic nature of evaluative sentences (see Chapter 4 and Chapter 5) since it is annotated with labels both for sentential constituents and for semantic roles. In this study we use PDT version 2.0.

For the purposes of discourse connective classification presented in Chapter 5.2, we use the Prague Discourse Treebank 1.0 (Poláková et al., 2013; PDiT, `http://ufal.mff.cuni.cz/pdit/`). PDiT is a manually annotated layer of linguistic description above the existing layers of the PDT, published as a part of PDT version 3.0. It portrays linguistic phenomena from the perspective of discourse structure and coherence. In terms of evaluation, it is e.g. a source of adversative or concessive constructions distinguished by inverse polarities and a valuable platform for the exploration of the relationship between discourse and evaluation in general.

Details about the build-up of the Czech Subjectivity Lexicon and information about other emotional data resources can be found in Chapter 8.

In this chapter, we have introduced theoretical resources for the research of emotional meaning. In the following parts, we make use of these foundations to provide an analysis of emotional meaning both with respect to the linguistic structure and within computational applications.

[1] Tectogrammatical layer represents the linguistic meaning of the sentence, i.e. it captures deep syntactic relationships, but it is more semantically oriented. On this level, the irregularities of the outer shape of sentences are absent (including synonymy and at least the prototypical cases of ambiguity) and it thus serves as an interface between linguistics in the narrow sense (as the theory of language systems) on one side and such interdisciplinary domains as that of semantic interpretation (logical analysis of language, reference assignment based on inferencing using contextual and other knowledge, further metaphorical and other figurative meanings), that of discourse analysis or text linguistics, and so on, on the other.

I Linguistic Structure

3

Lexical Aspects

This chapter is dedicated to the ways in which emotional meaning can be expressed on the lexical level of linguistic structure. When describing lexical features of emotional structures, we build on Zima (1961) who includes emotionality in the broader concept of expressivity. Inspired by Zima's categorization of expressive lexemes, we apply it on emotional lexemes, since we distinguish between the terms *expressivity* and *emotionality*. Whereas expressivity is understood as more general term including all the utterances deviating from the norm, emotionality is perceived as an expression of feelings (see also Mikulová, 2010).

Zima (1961) suggests that potentially all the lexemes in a lexicon can be expressive in a certain context and divides expressivity into three basic categories:

- *inherent expressivity*, a permanent part of word's meaning, a given lexeme is expressive in every context, e.g. *tlusťoch* – 'a fatty'
- *adherent expressivity*, concerning words which gain expressive meaning in certain contexts while in other contexts they remain neutral, e.g. *hrubý* – 'rough/huge/ coarse' v.s. *hrubý* – 'rude'
- *contextual expressivity*, a rather stylistic phenomenon which arises when a certain word significantly differs from the surrounding text context, e.g. when using colloquial language means in a formal text or vice versa: *Vážená paní/vážený pane, prosím pojďte se mnou dneska pařit.* – 'Dear Sir or Madam, please go boozing with me tonight.'

In the present study, we distinguish *inherent*, *adherent* and *contextual emotionality* defined on the same principle. In this chapter, we mostly focus on the first two types. Contextual emotionality is discussed in Chapter 6.

First, we characterize various inherently emotional lexical forms by which positive and negative emotional meaning can be conveyed, namely augmentatives, diminutives and other inherently emotional lexemes. In this survey, we join the evaluative morphology research which is concerned with e.g. augmentative and diminutive suffixes (see e.g. Stump, 1993). As a lexical resource, we use (among other resources) the Czech Subjectivity Lexicon, i.e. the dictionary of evaluative items in Czech (see Section 8.1).

Second, we investigate adherent emotionality. Apart from Zima, who mostly considers direct evaluative expressions, we also take into account implicit representation of emotional meaning. Namely we survey expressions with twofold polarity depending on a context. Also, we describe euphemisms and emotional idioms, drawing from the Czech lexicological tradition (see Čermák, 2007). Moreover, we look into alternative possibilities of expressing emotions graphically, or more precisely into different kinds of emoticons and other graphical symbols. Finally, we explore named entities containing emotional lexemes and describe their role in emotional meaning description.

3.1 Inherently Emotional Lexical Forms

3.1.1 Augmentatives

Augmentatives are derived words which usually express greater intensity. In Czech, they are mostly formed by the following suffixes: *-ák* (*chlapák* – 'he-man'), *-isko* (*chlapisko* – 'a hulk'), *-an* (*zoban* – 'a beak'), *-as* (*lotras* – 'a rascal') and *-izna* (*babizna* – 'a hag') (Dokulil et al., 1986). The common Czech grammars (see e.g. Dokulil et al., 1986, Karlík et al., 1996 or Karlík et al., 2002) claim that augmentatives primarily express negative (pejorative) evaluative meaning.

We agree that augmentatives are often used in negative sense, as *babizna* – 'old hag' in (3):

(3) Ta [babizna] páchne jako slaneček uleželý v sudu s dehtem!
'The [old hag] stinks like a red herring that's been stood over head in a tar barrel!'
(InterCorp)

However, based on linguistic evidence from both manually annotated evaluative corpora and CNC data, we claim that augmentatives can also be used in a positive context, as indicated by Karlík et al. (1996), p. 129, see (4):

(4) "Je to [chlapisko], ten Bunaparta," ozývá se opět Ljulj-hodža a mluví táhle, jako by nahlas myslil, "[chlapisko]."
'"He's [quite a man], this Bunaparta," the Reeling Hodja began again, in a slow drawl, as though thinking aloud, "[a great man]."'
(InterCorp)

Here, the augmentative *chlapisko* – 'a hulk' is used as a positive evaluation (as also apparent from the broader context of the example sentence in the corpus), and thus translated accordingly.

On top of that, we found examples of augmentatives which are used primarily in a positive sense, as in (5):

(5) "Jó," pravil ten starý pán, když si všiml, kam zírám, "to je [chlapák], co?"
'"Ah!" said the old gentleman, following the direction of my gaze, "[fine fellow] that, ain't he?"'

(InterCorp)

Moreover, the negative/positive or even ironic use of augmentatives is very often a matter of prosody and pragmatic context. For instance, a sentence *Pojď sem, ty [zeťáku]!* – 'Come here, you [son-in-law]!' can be considered either positive or negative, depending on the given situation.

Because of the above mentioned evidence, it is not trivial to employ augmentatives in automatic classification as indicators of emotions in a text (see Chapter 9). We cannot rely on the assumption that most words with specific augmentative suffixes can be considered negative, even if we exclude neutral lexemes with augmentative suffixes obtained automatically from a retrograde vocabulary, e.g. *podobizna* – 'a portrait' or *voják* – 'a soldier'.

To sum up, we claim that by adding an augmentative suffix, a lexeme gains:

- negative markedness,
  e.g. *pes* – 'dog' NEUTRAL → *psisko* – 'a cur' NEGATIVE
- positive markedness,
  e.g. *chlap* – 'a guy' NEUTRAL → *chlapák* – 'a great guy' POSITIVE
- stronger evaluation, if the original lexeme is already evaluative,
  e.g. *baba* – 'a hag' NEGATIVE → *babizna* – 'a witch' NEGATIVE

This concerns not just nouns, but also other parts of speech which take augmentative suffixes, i.e. adjectives and adverbs. However, in such cases the positive/negative markedness depends more on the word they modify, see e.g. *dlouhatánské čekání* – 'very long waiting' v.s. *dlouhatánský potlesk* – 'very long applause'.

3.1.2 Diminutives

Diminutives in Czech are a very productive category. They are derived forms used to convey a slight degree of the root meaning, smallness of the object or quality named, encapsulation, intimacy, or endearment. They are usually derived from concrete nouns of all types and they keep the grammatical gender of the root noun (or the stem). Diminutive suffixes in Czech are either primary (or simple), *-ek*, *-ík*, *-ka*, *-ko*, *-átko* or secondary (or compound), *-eček*, *-ečka*, *-ečko*, *-íček*, *-ička*, *-íčko*, *-ičko* and many more, since the derivation process can be endless in this case – e.g. *prstýneček, lokýnečička* – 'tiny ring, tiny ringlet'. Apart from nouns, diminutive suffixes can also be added to other parts of speech, mainly adjectives *tichounký* – 'very silent' and adverbs *tichounce* – 'very silently'. According to Czech grammars, diminutive suffixes add positive emotional meaning to the primary lexemes (see e.g. Dokulil et al., 1986, p. 301), but this does not hold for all the cases – sometimes the suffix can simply mean just a diminution. Moreover, diminutive suffixes are also used when deriving hypocorisms, i.e. intimate forms of words or given names, e.g. *Magdaléna* → *Magdička* etc.

In our data, we usually find diminutives in utterances conveying positive emotional meaning, see (6):

(6) Jenže [tatínek] byl v práci, [maminka] spala, a maňásci vypadali o tolik líp, když jim umyla [tvářičky].
'But [Daddy] was at work, and [Mommy] was sleeping, and the puppets looked so much prettier now that they had their [faces] cleaned up.'
(InterCorp)

However, we also realize that in cases when the diminutive suffix is already lexicalized, the diminutives convey no positive markedness anymore, see (7):

(7) Rodiče měli [rybičky] v akváriu, což mě absolutně nebavilo.
'My parents had [fish] in an aquarium, but I was absolutely not fond of it.'
(SYN2006pub)

In this case, *rybičky* – 'fish' means only fish in a bowl, not e.g. big sea fish, and it can be considered a neutral expression, despite being derived from *ryba* by a diminutive suffix *–ičk–*.

Apart from the positive or neutral use, diminutives are very often employed to express irony. Although it is quite difficult to detect irony without suprasegmental features and a situational context of the given utterance as explained in Section 6.2, there are some contexts in which we can conclude on ironic meaning, see (8):

(8) Takový pokrytecký [svatoušek] by zasloužil pěkně vyprášit.
'The hypocritical [saint] deserved a good thrashing.'
(InterCorp)

According to our evidence, diminutives are very often used in ironic sense when combined with an expression of the opposite polarity (see e.g. *[pokrytecký]- [svatoušek]+* – '[hypocritical]- [saint]+') in the example above.

3.1.3 Vulgarisms

Vulgarisms are lexical forms usually conveying pejorative emotional meaning. They play an important role in expressing and also automatic detection of emotional meaning (see Chapter 9), since they are usually very strong indicators of negative evaluation in text. Czech vulgarisms are collected in several lexicons, namely Bajger (1998), Obrátil (2000), Ouředník (2005) or Šimáčková (2009). Very often, vulgarisms are also parts of idioms (see Section 3.2.3), as in (9):

(9) a. Je tu tma jak v [prdeli].
'It is fucking dark here.' (neutral)
b. S Jardou je vždycky [prdel].
'Jarda is always fun to be around.' (positive)
c. A je to v [prdeli].
'We are fucked.' (negative)

To identify contexts in which vulgarisms are used in evaluative meaning, it is necessary to apply collocational analysis desambiguating the emotional structures and eventually positive or negative polarity of the given phrases.

However, when surveying emotional meaning in Czech data, we often work with user reviews full of vulgarisms expressing negative evaluation, see (10):

(10) No to si ten [zasranej] dr. dre ze mě už dělá [prdel], větší [sračku] fakt nemohl udělat [do píči]!
'That fucking dr. dre must be kidding me, he couldn't have made a bigger shit, fuck it!'

(Heureka.cz)

The rise of Web 2.0, characterized by user-generated content, led to changes in the ways in which vulgarisms are expressed. While, Internet users are now allowed to contribute to many online discussions, they have to follow various restrictive policies. Therefore, people try to get around the rules by inventing new "encoded" vulgarisms. Examples of the most common strategies illustrating the new forms of vulgarisms found in the data can be seen below.

- The author uses special characters:

  (11) To je ale [p*ča!]
  'Such a [c*nt!]'

- The author changes some letters for other ones (e.g. *hovado* → *howado*):

  (12) Ty [howado] s minimozečkem kura domácího!
  'You [beast] with a chicken brain!'

(Nova.cz)

- The author uses a wordplay:

  (13) a. *Kalousek*, a politician's surname → *Kaďousek*, evoking defecation
  b. *Peake*, a politician's surname → *Peacha*, phonetically evoking *píča* – 'cunt'

- The author relies on a general knowledge:

  (14) Svět se v ... obrací.
  'The world turns into ...'

  (Novinky.cz)

- The author uses abbreviations:

  (15) WTF?!
  'WTF?!'

There are several ways to identify the "encoded" vulgarisms automatically, so that we would not lose them within the lemmatization process. We can use the Levenshtein edit distance (Levenshtein, 1966) and compare potentially "encoded" vulgarisms with the list of the correct forms given the edit distance number (e.g. *p*ča* – 'c*nt' → *píča* – 'cunt', edit distance = 1 since it only takes one change * → *í* to get the correct form). Also, we can use lists of the most common abbreviations and correct them automatically using the paraphrasing rules. The standard state-of-the-art spell-checkers could also be the way to treat these expressions.

3.1.4 Other Inherently Emotional Lexemes

Besides the above mentioned categories, we also found dysphemisms like *chcípnout* – 'to peg out', depreciatives like *móresy* – 'bad habits', familiarisms like *zlatíčko* – 'sweetie' or melioratives like *maminka* – 'mummy' in the texts. Taking into account the fact that we are not able to detect irony (as well as understatement or overstatement) in the written text, we conclude that these items appeared in their prototypical negative or positive sense.

Apart from lexical items with evaluative affixes, there exists a set of inherently emotional Czech lexemes with no specific prefixes or endings. These lexemes, like *báječný* – 'amazing', *senzace* – 'blockbuster' or *trpět* – 'to grieve' are collected in Czech Subjectivity Lexicon (see Section 8.1). For the lexicon coverage, see Table 3.1.

Special category of emotional lexemes are interjections which are very often inherently emotional (e.g. interjections expressing negative feelings like *au* – 'ouch' or *fuj* – 'ugh'). However, sometimes we cannot be sure about their polarity, since it depends on the pragmatic context, see the following examples of positive and negative use of *ach* – 'oh':

(16) Ach, to je nádhera!
'Oh, that's beautiful!'

POS	lexemes
nouns	1,851
verbs	1,549
adjectives	773
adverbs	440
particles	12
total	4,625

Table 3.1: Subjectivity Lexicon coverage

(17) Ach, já jsem si zapomněla klíče!
'Oh, I forgot my keys!'

Apart from taking into account the inherently emotional lexemes, we also suggest that all the expressions that are neutral from the point of emotional meaning analysis can mostly be modified by evaluative prefixes like *pa–*, *kvazi–* or *pseudo–*, all meaning *pseudo–*, like in *věda* – 'science' $\rightarrow$ *pavěda* – 'pseudoscience'.

Moreover, we observed presence of *elusive elements*, i.e. words bearing evaluative power which are difficult to describe in terms of positive or negative polarity without using a complicated inference, e.g. *kontroverze* – 'controversy', *osobitý* – 'distinctive' or *zvláštní* – 'curious'.

3.2 Adherently Emotional Lexical Forms

Even though we explained the typical issues of inherently emotional lexical forms, we can still conclude that they are quite easy to understand both by humans and by machines, since they are usually single words with a relatively clear meaning. Adherently emotional lexical forms, on the contrary, tend to be much complicated, given the fact that they often gain their emotional meaning from the context in which they appear.

3.2.1 Expressions with Twofold Polarity

Although most words conveying emotional meaning have either positive or negative polarity, some of them can also have twofold polarity depending on their function in a sentence. This is true of adverbials like *pěkně* – 'pretty', *hrozně* – 'terribly', *strašně* – 'awfully', *příšerně* – 'horribly'. These expressions can, in certain contexts, work also as intensifiers and thus carry (or more precisely intensify) not just the same, but also the opposite polarity.

(18) a. Vypadala [pěkně].
'She looked [pretty].'
b. Choval se [pěkně] divně.
'He behaved [pretty] awkward.'

(19) a. Chovala jsem se [hrozně].
'I treated you [terribly].'
b. [Hrozně] se mi líbí.
'I find him [terribly] attractive.'

(20) a. Zacházeli s ní [strašně].
'She was treated [awfully].'
b. Jsou [strašně] milí.
'They are [awfully] nice.'

(21) a. [Příšerně] ji vyděsil.
'He scared her [horribly].'
b. [Příšerně] ji miluju.
'I'm [horribly] in love with her.'

Whereas in *a* examples the adverbials carry their original polarity, the overall polarity of *b* examples is exactly opposite.

Some of these words can be disambiguated based on their morphological context, e.g. taking into account whether they preceed yet another adverb or adjective or not (see Petkevič, 2006). For analysis regarding semantic development and speaker's perspective of intensifiers, see e.g. Athanasiadou (2007).

3.2.2 Euphemisms

Euphemisms are words or expressions replacing taboo words or words that may be found offensive or conveying inappropriate or negative meaning. They usually reduce or hide the negative connotation using substitution (e.g. *tlustý* – 'fat' → *plnoštíhlý* – 'plump'), see e.g. Kamenická and Rambousek (1995). Even though it is difficult to automatically recognize these expressions as conveying negative emotional meaning, they still should not be excluded from the analysis. Rather, they should be treated as somewhat weaker when performing more fine-grained sentiment analysis of the text, as in (22):

(22) hnusný > ošklivý > nehezký
'nasty > ugly > unsightly'

To complete the intensity scale, we could also add dysphemisms, i.e. expressions opposite to euphemisms (mentioned in previous section). Dysphemisms are offensive words or expressions replacing neutral words, e.g. *ústa* > *držka* – 'mouth > gob'.

Euphemisms tend to get lexicalized over time. They lose their emotional feature and thus have to be replaced by new expressions. Since they are the products of social changes and work more like a social construct within a given time, they are one of the most difficult evaluative items to be analyzed. This also holds for metaphors, i.e. speech figures using figurative meaning (*Byla šlehačkou na dortu jeho života.* – 'She was a whipped cream on the top of the cake of his life.').

3.2.3 Idioms and Phrasemes

An idiom (and phraseme) is an expression with figurative meaning. The meaning is non-compositional, i.e. it cannot be understood from the meaning of its parts. Any literal meaning is in the background, if it is considered at all. If somebody has *a ball is in his court* then he is expected to do the next turn and usually there are no court or ball involved. Similarly, when somebody *natáhne bačkory* (lit: pulls slippers) he dies, and any presence of slippers is irrelevant. According to Karlík et al. (2002), we use the term phraseme when performing formal analysis, whereas the term idiom is used when investigating semantic features.

Čermák (2007) describes the evaluative component as one of the basic components of phrasemes and idioms, closely tied up to its pragmatic component. According to him, the evaluative component "mediates the evaluative attitudes of the speaker on the scale good-bad" (p. 91). He divides evaluative idioms into contrastive categories: *sensory-intellectual* (*smrdí to jako bolavá noha* – 'it smells like an aching foot' v.s. *je chytrej jako opice* – 'he has a good brain') and *absolute-relative*, depending on whether it is a fixed idiom or whether the evaluative meaning is indirect.

Also, in his categorization, Čermák defines expressive and emotional class of phrasemes. The speaker uses the expressive phrasemes "to inform the listener of his interest in the situation, to express this interest, etc." (p. 110). The subclasses of expressive class (Ex) include especially the following categories (the examples of both expressive (Ex) and emotional (Em) classes were collected by the author of this book):

(a) congratulations – *Klobouk dolů!* – 'Well done!'
(b) thanks – *Tisíceré díky!* – 'Many thanks!'
(c) apologies – *Chybička se vloudí.* – 'Mistake may happen.'
(d) doubts – *hlas volajícího na poušti* – 'the voice of one crying in the wilderness'
(e) sympathies, recognition, dislike, repulsion, disgust, disdain – *má to své mouchy* – 'it has certain shortages'
(f) trust/mistrust – *mít tušení* – 'to smell a rat'
(g) accusation, reproach, condemnation – *být příjemný jako osina v zadku* – 'to be unpleasant'
(h) forgiveness, plea for forgiveness – *co bylo, bylo* – 'everything forgotten'

(i) joking – *šplouchá mu na maják* – 'a sandwich short of a picnic'
(j) helplessness, evasiveness – *tvářil se, jako by mu uletěly včely* – 'he looked unhappy'

Moreover, Čermák recognizes an emotional class of phrasemes, used when "the speaker conveys his emotional reaction to the listener, which is usually distinctly polarized evaluatively on the good-bad scale" (p. 111). Again, Čermák defines different subclasses of the emotional class (Em), namely:

(a) surprise, astonishment – *vyrazit dech* – 'to knock somebody's sock off'
(b) admiration, praise, contempt – *jít jako po másle* – 'to be done without any obstacles'
(c) pleasure, joy, refusal, displeasure – *být šťastný jako blecha* – 'to be very happy'
(d) envy – *mít se jako prase v žitě* – 'to have a good standard of living'
(e) satisfaction, disappointment, regret – *za málo peněz hodně muziky* – 'a lot of music for just a pittance', *stát za houby* – 'to be a drag'
(f) worry, fear, calm, confidence – *cítit se jako nahý v trní* – 'to feel uncomfortable'
(g) anger, vindictiveness – *být trnem v oku* – 'a thorn in one's side'
(h) derision, ridicule, sarcasm – *lepší než drátem do oka* – 'fairly good'

In our data, we found phrasemes of all these types conveying both positive and negative meaning. For the sentiment analysis purposes (see Chapter 9), it would be advantageous to split some of Čermák's subcategories like e.g. Em(e) into positive and negative and to create the lists (or tables, as suggested by Hnátková, 2002) of corresponding Czech positive and negative phrasemes. In this way, it could be easier to find these expressions even without using n-grams.

3.3 Alternative Ways to Express Opinion

In addition to lexical expressions, emotional meaning can also be expressed by other alternative opinion markers, namely emoticons and other graphical symbols. Emoticons are pictorial representations of facial expressions, using punctuation marks to express writer's emotions. As described by many authors (see e.g. Derks et al., 2007; Provine et al., 2007; Agarwal et al., 2011), they can express a whole scale of emotions from extremely positive (:-DDD) and positive (:-)) to neutral (:-|) to negative (:-() and extremely negative (>:-((). Emoticons can be put together using a whole range of punctuation marks and their use differs depending on the region. In sentiment analysis (see Chapter 9), they are one of the most important expressions to be identified (usually using words lists).

Other graphical forms indicating escalated expressivity can be also capital letters (23), increased quantity when using particular letters (24), repetitive use of punctuation (25) or other graphical means of emphasizing (26). Moreover, sometimes just adding an exclamation mark can add evaluative meaning into a sentence (27).

(23) Kolínka nadevšechno MILUJU!
'I LOVE elbow macaroni so much!'

(24) To je prostě božííííííííí!
'It is so greeeeeaaaaat!'

(25) Skvělý!!!!!!!!!
'Awesome!!!!!!!!!'

(26) Pepa je *kabrňák*!
'Pepa is a *great guy*!'

(27) a. Tento telefon stojí 6 420 Kč.
'This phone costs 6,420 CZK.'
b. Tento telefon stojí 6 420 Kč!
'This phone costs 6,420 CZK!'

3.4 Named Entites

Named entities, i.e. words or sequences of words which label names of things, can very often contain the above-mentioned lexical expressions typical for evaluative meaning. However, in this case they do not convey any evaluation. These terms can concern various fields. In the following phrases, the evaluative adjective *krásný* – 'beautiful' can be understood in the non-evaluative sense.

- presenter Jolka [Krásná]
- village [Krásná]
- Carpatian ridge [Krásná] polonina
- agricultural cooperative [Krásná] Hora nad Vltavou
- TV show [Krásný] ztráty
- magazine [Krásná] a zdravá
- lunar calendar [Krásné] paní
- song [Krásná] je Neapol
- poem [Krásná] Poldi
- novel Cizinec a [krásná] paní
- movie Život je [krásný]
- contest Věda je [krásná]

On the contrary, many named entities may carry either positive or negative evaluative connotations themselves, cf. *She was a real Joan of Arc* or *He was a true Hitler*. However, even if we feel some inherent evaluative meaning, it always depends on the context of the utterance.

In this chapter, we presented the basic characteristics of emotional meaning on the lexical level of linguistic description and pointed out some pitfalls which need to be taken into consideration e.g. when building the practical applications (see Chapter 9).

4
Morphosyntactic Aspects

The first part of this chapter is devoted to the ways in which emotional meaning can be conveyed on the morphosyntactic level of language, by which we mean parts of speech, inflectional morphology and structural relationships. Evaluative morphology communicated by derivational morphology is described in Chapter 3.

The second part of this chapter is focused on syntactic properties of emotional structures, starting from simple utterances up to the complex compound sentences and higher text units. Also, we examine negation as an important syntactic phenomenon substantially influencing emotional structures. Most of the observations in this part of the chapter come from the syntactically annotated Prague Dependency Treebank.[1]

4.1 Part of Speech Influence on Emotional Structure

Since the original aim of the research presented in this study was to build a simple but reliable sentiment classifier, we surveyed the possibilities of how to detect emotional meaning with as little effort as possible. Therefore, we first started with detecting lexical entries from subjectivity lexicon, hoping to apply a simple majority vote. However, it turned out that different parts of speech have different influence on the sentential polarity (which is closely connected to their syntactic positions in the given structures), as stated in Veselovská (2014b), a paper this section is partly based on.

As mentioned in the previous chapter, the core of the evaluation naturally consists of evaluative expressions. The most frequent parts of speech bearing evaluative information (according to their frequency in the Czech Subjectivity Lexicon, see Section 8.1) are the following, in descending order:

- nouns, *hulvát* – 'a boor'
- verbs, *ctít* – 'to honor'
- adjectives, *špatný* – 'bad'
- adverbs, *dobře* – 'rightly/well/correctly'
- particles, *bohužel* – 'unfortunately'

Of course, when assigning a positive or negative value, we need to be careful with disambiguation, since some of the words are evaluative only in some contexts, see e.g. (28):

[1] `http://ufal.mff.cuni.cz/pdt3.0/`

(28) a. [Dobře], já to udělám.
'[Well], I will do it.'
b. Zachoval se [dobře].
'He did it [well].'

In (28a), *dobře* – 'well' is a particle and expresses no evaluation, whereas in (28b) it is a positively evaluating adverb. Moreover, the evaluative expressions are frequently accompanied by intensifiers such as *strašně* – 'awfully' or *pěkně* – 'pretty', see (29):

(29) Ještě si pamatuju, že to kafe bylo [strašně] dobrý.
'I just remember that the coffee was [awfully] good.'

(SYN2000)

Intensifiers need to be treated carefully with respect to their collocations, as described in Chapter 3. Considering the structural properties of these expressions, we could apply a more fine-grained description of evaluative meaning. This means not only the binary opposition of positive and negative polarity, but also the employment of a scale classification (*good – bad* v.s. *good – pretty good – pretty bad – bad*).

It emerges from the data that the most influential part of speech (in terms of positive or negative orientation of the evaluation) is by far the verb. This holds not only because most of the verbs are in the position of the main predicate of the investigated sentences (verbs such as *milovat* – 'to love', *nesnášet* – 'to hate', *cenit si* – 'to appreciate', etc.), but also because there are a number of verbs which express individual opinion (e.g. , verbs such as *myslet si* – 'to think', *mínit* – 'to mean', *předpokládat* – 'to suppose', *považovat* – 'to consider', etc.). A more detailed semantic classification of evaluative verbs can be found in Chapter 5. Example (30) illustrate that verbs have higher indicative strength in terms of emotional meaning than e.g. nouns, which are more frequent in the lexicon. In all the following examples the plus or minus sign after the square brackets indicates the overall polarity of a phrase or a sentence. So far, we are not aware of any emotional meaning research taking into account the system of morphological labels and their influences on sentential polarity.

(30) a. [Toho hrdopýška všichni nesnášejí.]–
'[Everybody hates that braggart.]–'
b. [Toho hrdopýška všichni chválí.]+
'[Everybody praises that braggart.]+'

Although the negative noun *hrdopýšek* – 'braggart' appears in both sentences, the overall polarity is still conditioned by the verb. On the other hand, the fact that verbs in the Czech Subjectivity Lexicon are outnumbered by nouns, i.e. a part of speech with lower indicative strength, may be attributed to the fact that evaluative nouns frequently appear as part of the verbonominal predicate. Thus, they are incorporated in the typically verbal syntactic position, acquiring indicative strength in the construction as well. This phenomenon is described in Chapter 4 where we offer a thorough analysis of dependency data.

Another important part of speech which influences the orientation of evaluation in a given sentence are particles, or more specifically evaluative particles such as *bohudík* – 'fortunately', *bohužel* – 'unfortunately', and *chválabohu* – 'thank God' etc. As explained e.g. by Janečka (2014), these particles express a positive or negative stance of the speaker towards the whole proposition. Corpus data suggest that particles can switch the overall polarity of a given sentence just on their own, see the following example:

(31) [Bohudík toho hrdopýška všichni nesnášejí.]+
'[Fortunately, everybody hates that braggart.]+'

This example illustrates that even if the number of negative polarity items (*nesnášet* – 'to hate', *hrdopýšek* – 'a braggart') is higher than the number of positive polarity items (*bohudík* – 'fortunately'), the overall polarity of a sentence is still positive as a result of the evaluative particle influence. This also corresponds nicely to its syntactic position; a discourse particle modifies the whole sentence, and thus it gains the power to rule the overall polarity.

The evidence from the Czech National Corpus also indicates that evaluative nouns are somewhat weaker than evaluative adjectives:

(32) Byl to však [příjemný nepořádek]+, v němž se návštěvníci cítili uvolněně.
'However, it was a [pleasing mess]+, in which the guests felt good.'

(SYN2005)

After the corpus survey, we searched for similar structures in treebank data. The results described in detail further in this chapter demonstrate that an adjective modifying a noun is always more influential towards sentential polarity. Moreover, when an adjective happens to combine with a noun with an opposite polarity, we can sometimes find an ironic meaning in the sentence (if we accept the hypothesis that we can even talk about irony without prosody), as (33):

(33) Byl to hrdinný chlípník.
'He was a heroic lecher.'

For a discussion of the phenomena of irony and sarcasm, see Section 6.2.

4.2 Emotional Meaning Expressed by Particular Grammatical Categories

We have described different parts of speech that bear emotional meaning. However, emotional meaning can also be expressed at a more fine-grained level of particular grammatical categories, namely grammatical gender or number. Kučerová (2000) gives an example of agreement concerning third person masculine animate, which can be used either as specified or unspecified in terms of gender depending on the expressivity of a given structure. She claims that in more expressive sentences, the agreement of the grammatical gender can be weakened, giving the following example:

(34) Maminka, chudák malá, je na všechno sama.
Mom$_{\text{FEM}}$ poor-man$_{\text{MASC}}$ small$_{\text{FEM}}$ is on everything alone
'Mom, a poor thing, has to do everything herself.'

However, this type of mismatch concerns only a few words in Czech.

Moreover, we can assume that a usage of plural in connection with proper names can also indicate an increased level of expressivity:

(35) Chybí nám jednota! Co udělali ti noví mocipáni, všichni ti Havlové a Rumlové hned po tom svém puči? Rozbili jednotné odbory!!!! Každý tu říká, že ty odbory nebyly na nic, že jen organizovaly pionýrské tábory a dovolené. Ale ty tábory a dovolené byly skoro zadarmo – a kolik tam bylo legrace!!

'We miss unity! What have all the new big shots, all the Havels and Rumls done right after the putsch? The have destroyed the unions!!! Everyone says that the unions were good for nothing but organizing the pioneer camps and holidays. But the camps and holidays were almost for free – and what a big fun we had!!'

Concerning other morphological issues of emotional language, we need to take into account also a common Czech with the typical word endings like *-ej* instead of *-ý* etc., which might be in some contexts considered expressive:

(36) Slibuju, že budu řádně vykonávat práva a plnit povinnosti člena akademický vobce Univerzity Karlovy. Slibuju, že uchovám v úctě slavnou humanistickou a demokratickou tradici Univerzity Karlovy, budu dbát jejího dobrýho jména a budu studovat tak, aby moje činnost přinášela všestrannej užitek.

'I promise to respect the glorious humanistic and democratic tradition of Charles University, to heed its good reputation, and to study so that my activity brings universal benefit.'

(Matriculation and Graduation Code of Charles University)

Pattern			Example sentence
Subj_{target}	Verb_{copula}	PAdj	*Pavel je nevychovaný.* 'Paul is naughty.'
Subj_{target}	Verb_{copula}	PNoun	*Tahle dovolená je paráda!* 'This holiday is a dream!'
Subj_{target}	Verb	Adv_{eval}	*Pizza chutná výborně.* 'The pizza tastes so good.'
Attr_{eval}	Noun_{target}		*Skvělá cena.* 'Nice price.'
Subj_{target}	Verb_{eval}		*Modrý tým podvádí.* 'The blue team cheats.'
Subj_{source}	Verb_{eval}	Obj_{target}	*Péťa zbožňuje domino.* 'Peter likes dominoes.'
Subj_{target}	Obj_{source}	Pred_{eval}	*Nabídka vín mě potěšila.* 'The wine choice pleased me.'

Table 4.1: Syntactic patterns.

4.3 Syntactic Patterns of Basic Emotional Structures

As mentioned earlier in Chapter 2, there are three basic semantically-defined participants of emotional structures:

- the source, i.e. the person or entity that expresses or experiences evaluation – *<u>Susan</u> loves Peter*;
- the target that is evaluated – *Susan loves <u>Peter</u>*;
- evaluative element, i.e. word or phrase inherently bearing a positive or a negative value – *Susan <u>loves</u> Peter*.

In this section, we survey the ways these participants are realized at the syntactic layer of linguistic description. Table 4.1 captures basic patterns of simple prototypical evaluative constructions, using appropriate examples of the structure (37):

(37) Pavel je nevychovaný.
Pavel is naughty
SUBJ COPULA PADJ
'Pavel is naughty.'

Basic emotional structures can be employed in sentiment analysis tasks (see Chapter 9), since they can serve as patterns for more general rules like *if you find an evaluative adjective which is a part of a verbonominal predicate, its subject should be a target of evaluation*. These rules are applied e.g. when automatically detecting the target of evaluation (see Veselovská and Tamchyna, 2014).

As demonstrated in Table 4.1, both source and target can be present either in subject or in object position of the evaluative verb. Generally, the verb and verbal valency are important features of evaluative structures. Typical valency frames of emotional verbs are described in Chapter 5. Formal analyses of example emotional structures can be found in Chapter 7.

Apart from the positions proposed in Table 4.1, neither source nor target needs to be present in the surface syntactic structure of the text. This applies for several reasons:

1. Because Czech is a pro-drop language omitting inferable pronouns, as illustrated by the following example:

   (38) Miluje lanýže!
   loves$_{3SG}$ truffles
   'He/she loves truffles.'

2. Because both source and target of evaluation can be external:
   (a) The source can be the author who does not have to necessarily mention him/herself in the document:

   (39) Velmi příjemná oddychová hudební komedie. Výborné písničky, příběh i scénář, herecké výkony, kamera, režie.
   'Very nice relaxing music comedy. Great songs, story and screenplay, actors, camera, direction.'

   (CSFD.cz)

   Besides, the source can be the target him/herself:

   (40) Jsem blbec.
   'I am an idiot.'

   Neither the target has to be explicitly mentioned. However, the target is very often recognizable from the textual context (e.g. review of a particular movie).

   (41) Jedna velká křeč a plno klišé. Pointa žádná. Nuda.
   'One big cramp, plenty of clichés. No point. Boring.'

   (CSFD.cz)

   (b) Moreover, neither source nor target has to be expressed when reacting to situational context (*Do háje!* – 'Damn it!' when someone hurts him/herself). The situational context is further discussed in Chapter 6.

3. Very often, the target can be represented as a larger passage of the text. In the following example, most of the text is the target and only the very last sentence evaluates it.

   (42) Láska, která nenachází svoji odpověď. Mladý barový číšník Bob je zaslepen láskou k mladé prostituce, jejíž vychytralost ho připravuje o všechny úspory a odvádí na útěk před společností na loď. Jeho laskavá dobrota a naivní zamilovanost ho připravuje o všechny ideály a naděje, aby pak rezignoval na normální společenský život. Mladá prostitutka Jenny je příkladem toho, jak lehce lze klesnout na samotné dno lidské společnosti. Jediné zaváhání, mladická nerozvážnost a zaslepená důvěřivost k mužům ji přivádí k nejstaršímu řemeslu a morálnímu posunu k nečestnosti. Vždyť za vším stojí prostopášnost mužského rodu. Prostě blbost.

   'Love that finds no answers. Young bartender Bob is infatuated by young prostitute, whose craftiness deprives him of all his savings. She takes him away from the society and make him escape by the boat. His good nature and naive love makes him lose all his ideals and hopes and force him to resign to normal social life. Young prostitute Jenny is a good example of how easy is it to reach the bottom of human society. Only one small hesitation, young rushness and blind trustfulness towards men brings her to the oldest profession and a moral downfall to the dishonesty. Because it is an immorality of a man what stand behind all of this. Crap.'

   (CSFD.cz)

4.4 Compound Emotional Sentences

For a precise opinion target identification, it is usually useful to work with shorter syntactic chunks of text. However, when performing other sentiment analysis tasks, it can also be advantageous to take into account sentences with greater complexity. These are recognized mostly by their syntactic structure or lexical hints in certain context (e.g. *ale* – 'but' following a comma). For polarity classification, i.e. detection whether a given part of the sentence contains either positive or negative evaluation, one needs to be especially careful with the following structures.

4.4.1 But-clauses and And-clauses

But-clauses are main clauses coordinated by the conjunction *ale* (meaning *but* in Czech). *But* usually coordinates sentences (or constituents) with opposite polarity, as mentioned e.g. by Meena and Prabhakar (2007). The prototypical use can be found in the following example of adversative coordination:

(43) [Pláž byla hrozná,]– ale [v hotelu se nám líbilo.]+
'[The beach was awful,]– but [we liked the hotel.]+'

Contrary to *but-clauses*, *and-clauses* are main clauses coordinated by the conjunction *a* (meaning *and* in Czech). *And* usually coordinates sentences (or constituents) with the same polarity:

(44) [Pláž byla skvělá]+ a [hotel byl útulný.]+
'[The beach was great]+ and [the hotel was cosy.]+'

Although this phenomenon is rather a matter of semantics and it is discussed in detail in Chapter 5, we still have to take the syntactic hints into account, especially when solving some of the sentiment analysis subtasks. After a correct detection of adversative or conjunctive coordinations in parsed data, we can e.g. add the following consistency rules for opinion target identification (see Section 9.3):

(45) *If there are two opinion targets identified in the but-coordination, they should be marked with opposite polarities.*
If there are two opinion targets identified in the and-coordination, they should be marked with the same polarity.

More general rules like *if there are two parts of the sentence separated by "ale" following comma, these parts should have opposite polarities* can be employed in polarity classification, where they can produce features of probabilistic polarity classifier (see Section 9.2). More about this feature called *semantic consistency* can be found in Section 5.2.1.

Although correctly identified syntactic coordinations help to improve sentiment analysis, one still needs to be aware of the fact that conjunctions *but* and *and* can also express other relations, like e.g. gradation, as discussed in Chapter 5. In rare examples, they can even work the opposite way than usual, as in example (46):

(46) [Pláž byla skvělá]+, a [hotel byl hrozný.]+
'[The beach was great]+, and [the hotel was awful.]+'

4.4.2 Concessive Clauses

Concessive sentences, i.e. sentences beginning with *ačkoliv* – 'although', *jakkoliv* – 'even though', *byť* – 'albeit' etc. also play important role when expressing emotional meaning at the level of syntax, and thus have to be treated carefully. Concessive sentences still express evaluative meaning, but weakened in a way:

(47) Přestože baterie dlouho nevydrží, jsem spokojen.
'Although the battery life is not long, I am satisfied.'

The author is still satisfied with the product, but if it were not for the concessive clause, his enthusiasm could have probably been perceived as stronger. The concessive feature can be further used when applying a more fine-grained, scale-based classification of evaluation (see Wilson, 2008).

4.4.3 Conditional clauses

Even though conditional sentences can express emotional meaning as well, it is sometimes difficult to determine it, as observed e.g. by Narayanan et al. (2009). Some conditional sentences express no emotional meaning since they do not indicate whether the situation ever happened. Emotional words themselves do not distinguish an emotional sentence from a non-emotional one.

(48) Jestli uvaří, budu ho milovat.
'If he makes a meal, I will love him.'

In (48), we can find a positively evaluative word *milovat* – 'love'. However, we do not know whether the meal was ever prepared and thus we cannot affirm that this sentence evaluates the target *him*.

On the other hand, some conditional sentences still express some evaluation towards particular targets – see positive evaluation of *mascara* in (49).

(49) Pokud se vám nelíbí umělé řasy, pořiďte si tuhle úžasnou černou řasenku.
'If you do not like false lashes, buy this amazing black mascara.'

This sentence expresses no opinion about false lashes, but it still expresses positive evaluation of a black mascara. However, in this case it is a matter of lexical rather than syntactic features.

There is no clear agreement on how to treat conditional sentences in terms of emotional meaning detection, nevertheless, there are two basic approaches to depending on the data size. One can either withdraw them from classification at all (using the rule saying that whenever a sentence starts with *if*, it is not to be taken into account), or put them aside for further detailed analysis, as performed in Narayanan et al. (2009).

4.5 Text-level

On text level of syntactic analysis, we are dealing with connectors similar to these mentioned in Section 4.4. The only difference is that the information is split into several sentences.

(50) [Pláž byla hrozná.]– Ale [v hotelu se nám líbilo.]+
'[The beach was awful.]– But [we liked the hotel.]+'

Also, the source and target of evaluation are easier to find when taking into account a broader context. In the following example, it is easier to detect that it is Karel who loves truffles.

(51) Karel přišel na návštěvu. Miluje lanýže.
'Karel came for a visit. He loves truffles.'

However, this is again a matter of discourse (and even implicit discourse relations in this case), which is discussed in Section 5.2.

4.6 Negation and Evaluation

In connection with morphosyntactic properties of evaluative structures, we also have to mention syntactic negation (lexical negation e.g. *spokojený* – 'satisfied' v.s. *nespokojený* – 'dissatisfied' is a matter of expression-level of linguistic description and as such it would belong to Section 3.1.4). Syntactic negation in Czech is discussed in detail in Hajičová (1975).

For the purposes of this study, the most important feature of negation with respect to evaluation lies in the fact that negation often switches polarity (as explained e.g. by Wiegand et al., 2010):

(52) a. [Tenhle koláč je dobrý.]+
'[This cake is tasty.]+'
b. [Tenhle koláč [není] dobrý.]–
'[This cake is [not] tasty.]–'

The polarity switching feature of negation is considered to be language independent. However, we need to take into account the fact that it is still easier to detect syntactic negation in English where it is mostly expressed by the negative particle *not*, than in Czech where we use different ways to express negation, as described e.g. in Veselovská (2010).

Moreover, when dealing with emotional meaning (and in sentiment analysis applications as well), it is necessary to distinguish between sentential negation and constituent negation, since constituent negation only affects the target of evaluation. Unlike in English, sentential negation in Czech is a part of the verb and it switches polarity of the whole sentence. Constituent negation, on the contrary, negates just one constituent. To differentiate these two is not easy in Czech, since the negative prefix *ne* – 'not' as a part of the verb can negate either a whole sentence or a constituent.

(53) Petra neodsuzuji.
Petr_{Acc} $\text{not-condemn}_{\text{1SG}}$
'I do not condemn Peter.'

(54) Neodsuzuji Petra, ale Pavla.
not-condemn$_{1SG}$ Petr$_{Acc}$ but Pavel$_{Acc}$
'I do not condemn Peter, but Paul.'

In example (53), the whole presupposition (*I condemn Peter*) is negated. In the second part of the adversative paratactic structure (54), the predicate is omitted, but the presupposition (*I condemn somebody*) remains unchanged. The negative element negates only the object (*Peter*). This holds also for the following structure, which we consider semantically equivalent with (54):

(55) Odsuzuji ne Petra, ale Pavla.
condemn$_{1SG}$ not Petr$_{Acc}$ but Pavel$_{Acc}$
'I condemn not Peter, but Paul.'

The only difference is that in (55), the negative element is a separate word, whereas in (54) it is a part of the verb. Therefore, it can be assumed that sentences with a negated constituent should be treated as structures expressing constituent negation, regardless of whether the negative particle is a part of the verb or not.

The basic form of the constituent negation expressed by both a separate particle *ne* (or *nikoliv*) and a negative prefix *ne–* negates both complements and adjuncts. When expressed by a negative particle, it always precedes the negated constituent (and its extensions), i.e. it stands to the left of the given constituent. The constituent negation negates:

(56) a. subject:

Ne blázen, ale vlastenec to byl.
not fool but patriot it was

'He was not a freak, but a patriot.'

b. object:

Vzala si ne kriminálníka, ale hodného chlapce.
married REFL not criminal but nice chap

'She married no criminal, but a nice guy.'

c. attribute:

Ne chamtivý, ale milý a spořádaný nápadník se nakonec našel.
not greedy but nice and decent groom REFL finally found

'Finally, they found not a greedy, but a nice and decent groom.'

d. adverbial:

Dopadlo to nakonec ne špatně, ale dokonce výborně.
ended it finally not badly but event excellently

'It ended up not badly, but even excellently.'

As evident from the data and described in detail in Veselovská (2011a), there are several rules determining constituent negation in Czech (not only) evaluative structures:

- Constituent negation mostly requires adversative specification (on a sentence level or on a text level). The exceptions from these rules are defined below.
- Constituent negation can be represented not only as a separate negative particle *ne, nikoliv* – 'not, no', but also as a negative prefix on the verb.
- Constituent negation which is not a part of the verb always precedes the negated constituent (and its syntactic extensions).

The first rule accepts structures

- which express contrast, but there is an omitted verb

(57) Nefláká se Petr, ale Pavel.
not-loafs REFL Petr but Pavel
'Peter does not loaf, but Paul [does].'

- which represent common coordination of the two propositions in an adversative paratactic structure – i.e. the omitted verb is added into the structure

(58) Nefláká se Petr, ale fláká se Pavel.
not-loafs REFL Petr but loafs REFL Pavel
'Peter does not loaf, but Paul does.'

In the context of emotional meaning, we treat structures of the type (57) as coordinations of two sentences where the first one has an overall positive polarity, with verb being displayed on the surface, and the second one has negative polarity, with verb being omitted. This approach is supported by the *inverse polarity test*, i.e. test in which we switch positive and negative polarities of sentences. In case we preserve the position of the constituent negation before the constituent it negates (see example (59)), we can freely exchange the polarity (without changing the position of the negative particle – it still stands to the left of the constituent it negates).

(59) a. Fláká se Petr, ale ne Pavel.
loafs REFL Petr but not Pavel
'Peter loafs, but not Paul.'

b. Fláká se ne Petr, ale Pavel.
loafs REFL not Petr but Pavel
'Peter does not loaf, but Paul does.'

In case the negative particle follows the negated constituent, we lose this possibility (see example (60):

(60) a. Fláká se Petr, ale Pavel ne.
loafs REFL Petr but Pavel not
'Peter loafs, but Paul does not.'

b. * Fláká se Petr ne, ale Pavel.
loafs REFL Petr not but Pavel
'Peter loafs not, but Paul.'

In this example, we can observe that *ne* is not an ordinary negative particle which stands in front of the negated constituent. We take these sentences as the exceptions from the claim that negative particle always precedes the negated constituent.

Concerning the adversative coordination rule, it is also necessary to mention the fact that we found several cases where the structure did not require coordination. We can categorize these occurrences as follows:

We do not need to involve the coordination when the particle *ne*:

- precedes a quantifier

(61) Ne všechno je však skvělé a všeobecně přijímané.
'But not everything is great and generally agreed upon.'

- precedes adverbials

(62) Ne náhodou je tak protivná.
'Not accidentally is she so annoying.'

- is a part of a construction expressing comparison (*ani ne tak – jako* – 'not so – as')

(63) Od určitého okamžiku se projekt eura stal ani ne tak hospodářským projektem jako projektem politickým.
'At a certain point, the euro project became, primarily, not so much an economic project as a political one.'

(InterCorp)

- is a part of conditional constructions (*pokud ne – tak* – 'if not – then')

(64) Ráda bych dostala informace o těchto srovnávacích statistikách, pokud ne dnes, tak v nějaké budoucí odpovědi.
'I would be interested in having these comparative figures, if not today, then in a future reply .'

(InterCorp)

In this chapter, we described morphosyntactic features of emotional structures. However, to be able to detect emotions automatically, we still need to look into their semantic features, which is the subject of the following chapter.

5
Semantic Aspects

In this chapter, we discuss semantics of evaluative structures. Particularly, we focus on verbs bearing emotional information and participants in their valency frames within simple sentences. Also, we describe complex emotional structures linked up with various discourse connectives and their alternative realizations which have impact on sentential polarity. Moreover, we briefly analyze attitude markers and coreference relations and the role they play in emotional structures.

5.1 Emotional Verbs

As mentioned several times in this monograph, verbs play a crucial role in structures with emotional meaning. In this section, we propose an analysis of emotional verbs concerning their semantic and valency properties with respect to the degree of subjectivity and evaluativeness. The first part of the section concerning semantic patterns of emotional verbs is based on Šindlerová et al. (2014).

There are two main verb classes having impact on emotional structures which can be found in Czech SubLex (see Section 8.1):

1. good and bad news verbs
2. verbs propagating sentiments to their arguments

Whereas verbs from the first class express positive or negative content themselves (e.g. *jásat* – 'rejoice', *lamentovat* – 'lament'), verbs from the second class express sentiments towards arguments in their valency structure (e.g. *chválit někoho* – 'to praise someone'). Generally, there are several reasons why information about verbal valency is valuable. First, different valency frames are typically connected with different verb senses. It is a common phenomenon that individual senses of a verb differ with respect to the presence (or absence), degree and orientation of polarity. Disambiguating different senses of a verb allows us to identify sentiments more precisely. For example, in case of the verb *abdicate* in English, we are able to differentiate between the intransitive pattern *to leave a position* which does not constitute evaluative meaning directly, and the transitive pattern (*abdicate one's responsibilities*), meaning *to fail* and creating an evaluation stance with opinion target in the position of the Actor. There is a whole group of verbs in the original English lexicon sharing the non-evaluative semantics of *action under a physical disorder*, which in their second sense describe an evaluative stance (*omezovat* – 'hobble', *otřást* – 'jolt' etc.). This causes troubles e.g. when feeding them into the machine translation process, since without verb sense disambiguation we risk gaining a considerable number of inappropriate lexical units which may later spoil the polarity tracking results. In a broader perspective, such a disambiguation process represents a decision between real subjective sentiments and the so-called *good* or *bad news* (objectively presented positive and negative content). This task is recognized as important e.g. in the sentiment analysis of the news (see e.g. Balahur et al., 2010).

Valency is expected to be helpful in the task of the identification of the target of the evaluation as well. As already mentioned in Section 4.3, subjectivity analyses usually concerns three components of an evaluative private state that need to be distinguished:

- the source, i.e. the person or entity that expresses or experiences evaluation – *Susan loves Peter*;
- the target that is evaluated – *Susan loves Peter*;
- evaluative element, i.e. word or phrase inherently bearing a positive or a negative value – *Susan loves Peter*.

From the corpora of evaluative texts (see Chapter 8) we are able to extract and categorize typical abstract semantic patterns for expressing emotional meaning containing the three above mentioned components.

For the purpose of analysis, we also use verbal entries from SubLex 1.0 (see Section 8.1). There are 1,549 verbs in Czech SubLex. Within the analysis, each verbal item of the lexicon is considered independently in order to decide which valency argument of the verb corresponds to the target of the sentiment propagated by the verbal evaluative semantics. The current version of SubLex contains information about about lemma, polarity orientation and English translation.

5.1.1 Semantic Patterns of Emotional Verbs

As explained in Chapter 4, some verbs only serve as syntactic hints for evaluative words (evaluative nouns, adjectives, or adverbs). Typically, this is the case of copular verbs, psychological verbs (verbs describing mental action), communication eliciting verbs, or light verbs marking complex predication (phrasal verbs etc.).

Other verbs function as bearers of the evaluation themselves. These verbs are listed in a subjectivity lexicon. In a typical verb-centered evaluative stance, evaluation as such is carried by the verb, while the source and the target of the evaluation occupy the positions of verb arguments. The verbs in the lexicon then differ with respect to the question of sentiments propagation to individual arguments. Examples of semantic patterns for evaluative verbs can be found in Table 5.1. For all the tables in this chapter it holds that:

- ACT stands for the actor, i.e. the human or non-human originator of the event, the bearer of the event or a quality/property or the experiencer

Pattern	Example sentence
ACT_{src} PRED_{eval} PAT_{target}	*Líbí se mi to jméno.* 'I like the name.'
ACT_{target} PRED_{eval} PAT_{src} EFF	*Duchovní láska člověka obohacuje.* 'Spiritual love enriches the man.'
ACT_{target} PRED_{eval} PAT_{src}	*Nový ministr zdravotnictví dráždí novináře.* 'The new health minister irritates journalists.'
ACT_{src} PRED_{eval} PAT_{target}	*Novináři kritizují nového ministra zdravotnictví.* 'Journalists criticize the new health minister.'

Table 5.1: Semantic patterns of emotional verbs

- ADDR stands for the addressee, i.e. an argument with the cognitive role of the recipient of the event
- EFF stands for the effect, i.e. an argument with the cognitive role of the effect/result of the event, it is also assigned if the verb has at least three arguments
- ORIG stands for the origo, i.e. an argument with the cognitive role of the source of the event
- PAT stands for the patient, i.e. the affected object
- PRED stands for predicate

A number of verbs which appear in the lexicon do not propagate sentiments to any of its arguments. These are most probably candidates for what we mentioned earlier as *good/bad news* verbs. Generally, we describe good/bad news items as terms designating positive or negative situations or facts (like *válka* – 'war', *katastrofa* – 'disaster', *štěstí* – 'luck' etc.). The good/bad news verbs (in their primary meaning) do not evoke a positive or negative attitude to an entity/situation/fact occupying any of the valency positions. Rather, they function at the same time both as the polar word and the target of the sentiment. Examples of such verbal items are listed in Table 5.2.

In the sentence *The doctors had to amputate his foot* we see no propagation of the negative sentiment towards any of the verbal arguments. Yet the sentence still expresses negative news.

Due to the fact that none of the good news/bad news verbs propagate sentiment to any of its valency participants, it is necessary to mark them as a separate category in the lexicon. Still, it is beneficial to keep them in the lexicon because they provably, though indirectly, influence emotions of the reader and express overall polarity of the sentence.

Table 5.3 contains verbs propagating sentiments to the Actor position. They usually describe events of destruction (negative sentiments), or progress (positive sentiments), or events of direct experiencing emotional states. The interesting fact about verbs propagating sentiments to the Actor position is that they are usually verbs al-

SubLex Verb	Example sentence
abdikovat – 'abdicate'	*Císař Vilém II. abdikoval.* 'The emperor Wilhelm II abdicated.'
amputovat – 'amputate'	*Lékaři mu amputovali chodidlo.* 'The doctors amputated his foot.'
bědovat – 'bědovat'	*Maminka tiše bědovala.* 'Mum was quietly moaning.'
dovádět – 'frolic'	*Tanečnice na parketu dovádějí jako malé děti.* 'The dancers frolic on the dance floor like little children.'
hladovět – 'starve'	*Přiberu pět kilo, pak zase hladovím.* 'I put on five kilos, then I starve again.'

Table 5.2: Examples of verbs not propagating sentiments to any of their arguments

lowing the Abstract Cause-Subject alternation (Levin, 1993), i.e. an alternation of valency participants of the type *Mike distorted the wonderful moment with a scream* and *Mike's scream distorted the wonderful moment*. Different aspects of the semantic shift between the two alternations are widely discussed (Alexiadou and Schäfer, 2006) and the shift of the sentiment focus can be seen as significant in this respect.

As can be seen from Table 5.4, verbs propagating sentiments to an Addressee or Patient position usually describe events of taking and communicating a stance (both polarities), stopping or eliminating (negative), or praising (positive).

The data also give evidence for another pattern: the target of the evaluation is the centre of the evaluative stance. The way the source of evaluation is expressed is dependent on the verb's semantic choice of the target argument. If the target is expressed by a PAT argument, the source occupies the ACT position. If the target is selected at the ACT position, the source must be expressed external to the clausal structure (e.g. by means of *in my opinion* etc.).

The issue of propagating sentiments is more than complicated. There are of course more argument types than we have suggested so far to which sentiments can be propagated. The sentiments may be propagated to more than one argument in a structure. For example, in a sentence *John criticized Mary for her not coming*, the negative sentiment affects not only *Mary* as the patient, but also *her not coming* as the cause of critique. The same may apply to verbs allowing the Abstract Cause-Subject alternation, where the sentiments may affect secondarily not only the Actor position, but also the position of the Abstract Cause if present overtly.

Deciding the sentiment propagation direction is also a nontrivial question – which of the two affected arguments receives the sentiment primarily and which acquires it on the basis of some semantic transfer. To make the issue even more fuzzy, there is more than one kind of sentiment. We must distinguish between sentiments that

SubLex Verb	Pattern	Example sentence
bavit – ‘amuse’	ACT_{target} PRED_{eval} PAT_{src}	*Hoteliérství mě baví ze všeho nejvíc.* *I most enjoy being a hotel owner.*
děsit – ‘freak’	ACT_{target} PRED_{eval} PAT_{src}	*Nekonečná samota tě děsí.* ‘The neverending solitude freaks you out.’
kazit – ‘spoil’	ACT_{target} PRED_{eval} PAT	*Nedovolím ti kazit mi život.* ‘I won’t allow you spoil my life.’
narušit – ‘distort’	ACT_{target} PRED_{eval} PAT	*Nádhernou chvíli narušil výkřik.* ‘The wonderful moment was distorted by a scream.’
naštvat – ‘upset’	ACT_{target} PRED_{eval} PAT_{src}	*Rozhodčí naštval domácího borce.* ‘The referee upset the guy from a home team.’
ohrozit – ‘endanger’	ACT_{target} PRED_{eval} PAT	*Těžba ohrozí existenci jejich domovů.* ‘Mining will endanger existence of their homes.’
zachránit – ‘rescue’	ACT_{target} PRED_{eval} PAT EFF	*Dobré jméno vlády zachránil ministr Bursík.* ‘The Government’s credit was saved by minister Bursík.’
zlepšit se – ‘improve’	ACT_{target} PRED_{eval} ORIG PAT	*Zlepšila se jí pleť a rozjasnily oči.* ‘Her skin improved and her eyes brightened.’

Table 5.3: Example of verbs propagating sentiments to the Actor position

SubLex Verb	Syntactic Pattern	Example sentence
bát se – 'fear'	ACT$_{src}$ PRED$_{eval}$ PAT$_{target}$	*Bojím se, že přijdu o všechny své peníze.* 'I fear losing all my money.'
degradovat – 'degrade'	ACT$_{src}$ PRED$_{eval}$ PAT$_{target}$	*Tento přístup degraduje ženy na pouhé sexuální objekty.* 'This approach degrades women to mere sex objects.'
doporučit – 'recommend'	ACT$_{src}$ PRED$_{eval}$ ADDR PAT$_{target}$	*Studium lingvistiky bych doporučil každému studentovi.* 'I would recommend studying linguistics to any student.'
důvěřovat – 'trust'	ACT$_{src}$ PRED$_{eval}$ PAT$_{target}$	*Tvému úsudku plně důvěřuji.* 'I fully trust your opinion.'
eliminovat – 'eliminate'	ACT$_{src}$ PRED$_{eval}$ PAT$_{target}$	*Je potřeba eliminovat falešná doznání.* 'It is necessary to eliminate false confessions.'
kárat – 'reproach'	ACT$_{src}$ PRED$_{eval}$ PAT$_{target}$	*Vedoucí káral nevkusně oděného účetního.* 'The manager reproached the tastelessly dressed accountant.'
odmítnou – 'reject'	ACT$_{src}$ PRED$_{eval}$ PAT$_{target}$	*Odmítnul nabídku členství v KSČ.* 'He rejected the offer of becoming a member of communist party.'
oslavovat – 'praise'	ACT$_{src}$ PRED$_{eval}$ PAT$_{target}$	*Švýcaři oslavují nového šampiona ve sjezdovém lyžování.* 'The Swiss praise the new champion in alpine skiing.'
prosazovat – 'advocate'	ACT$_{src}$ PRED$_{eval}$ PAT$_{target}$ EFF	*Rychlé přijetí evropské měny prosazuje Jan Švejnar.* 'Jan Švejnar advacate prompt adoption of the euro.'

Table 5.4: Example of verbs propagating evaluation to the position of Addressee or Patient

affect the source of evaluation in the text and sentiments which affect the perceptor of the text. Thus, for example, in the sentence *John ignored Mary without reason*, Mary is the target of negative sentiments of John, but John may also be a target for negative sentiments held by the reader. Similar transfer of sentiments from the inner sentential structure (textual source of sentiments) to the external reader's perception appears with many verbs propagating sentiment to a non-actor position (see Table 5.4). In a valid lexicon for opinion target extraction, we must keep the information about which of the two sentiments (reader-oriented or source-oriented) we want to trace.

It can be seen from the analysis that verbs propagating sentiments to the same arguments usually belong to the same semantic classes, or at least share the same semantic components. The clusters of semantically similar verbs arising in the analysis are well traceable in common semantic class databases, such as FrameNet (see Ruppenhofer et al., 2006), or VerbNet (see Schuler, 2005). Semantic classes of emotional verbs are discussed in the following section.

5.1.2 Semantic Classes of Emotional Verbs

Looking at the 1,549 verbal entries in Czech Subjectivity Lexicon, we can define several semantic classes of emotional verbs. Inspired by Levin (1993), we distinguish more general classes and subclasses within these classes. Levin claims that the argument structure is derivable from meaning structure and that predictions about verb behavior are feasible because particular syntactic properties are associated with verbs of a certain semantic type. When it comes to verbs bearing emotional meaning, we can again generalize the basic participants of the structure as the source and the target of the emotion. However, we are aware of the fact that particular classes have their own more specific participants (as used by Levin), like e.g. in the following example:

(65) Marie obdivuje letce.
ADMIRER ADMIREE ADMIRED
'Mary admires pilots.'

Instead of admirer and admired, we still use only source and target in our classification.

Following Levin's classification, we recognize semantic classes of emotional verbs listed below (for every class, we only give several prototypical examples):

- Verbs of Positive Psychological State
  - Admire Verbs: *obdivovat* – 'admire', *milovat* – 'love', *zbožňovat* – 'adore'
  - Amuse Verbs: *mít požitek* – 'delight', *pobavit* – 'amuse', *těšit* – 'please'
  - Appeal Verbs: *být důležitý* – 'matter', *líbit se* – 'appeal', *mít rád* – 'like'
  - Marvel Verbs: *divit se* – 'wonder', *rozplývat se* – 'enthuse', *žasnout* – 'marvel'
- Verbs of Negative Psychological State
  - Bother Verbs: *obtěžovat* – 'irritate', *otravovat* – 'annoy', *štvát* – 'bug'

  - Detest Verbs: *mít odpor* – 'dislike', *nenávidět* – 'hate', *nesnášet* – 'detest'
  - Envy Verbs: *nevražit* – 'grudge', *závidět* – 'envy', *žárlit* – 'be jealous'
  - Fear Verbs: *bát se* – 'be scared', *hrozit se* – 'dread', *strachovat se* – 'be afraid'
- Verbs of Communication
  - Verbs of Manner of Speaking: *blábolit* – 'babble', *drmolit* – 'jabber', *odseknout* – 'retort'
  - Complain Verbs: *bědovat* – 'moan', *lamentovat* – 'lament', *reptat* – 'grumble'
  - Verbs Indicating Individual Opinion: *mínit* – 'mean', *myslet* – 'think', *předpokládat* – 'suppose'
- Verbs Involving the Body
  - Pain Verbs: *bolet* – 'hurt', *dusit* – 'suffocate', *sedřít* – 'graze'
  - View Verbs: *civět* – 'stare', *čumět* – 'gawk', *zírat* – 'gaze'
- Verbs of Evaluation
  - Verbs of Positive Evaluation: *cenit si* – 'appreciate', *chválit* – 'praise', *vážit si* – 'esteem'
  - Verbs of Negative Evaluation: *hanobit* – 'defame', *kritizovat* – 'criticize', *odsuzovat* – 'condemn'
- Verbs of Desire: *bažit* – 'crave', *toužit* – 'long', *dychtit* – 'desire'
- Verbs of Damage: *drtit* – 'crush', *ničit* – 'destroy', *poškozovat* – 'damage'
- Verbs of Fraud: *balamutit* – 'fool', *lhát* – 'lie', *vymýšlet si* – 'make things up'

These semantic categories are not exhaustive and they are somehow fuzzy. This means that one verb can belong to several classes at once (e.g. *nenávidět* – 'hate' expresses both psychological state and evaluation). However, the information about valency and semantic class characteristics can still be added to relevant entries. This can be possibly done by means of pointers to existing valency resources for Czech, such as VALLEX 2.5 (Lopatková et al., 2007) or PDT-Vallex (Urešová, 2011).

5.2 Emotional Meaning and Discourse Structure

In this section, we discuss semantic properties of more complex structures bearing emotional meaning. These can be either complex sentences or even larger chunks of text. In Czech linguistic tradition, the field of research concerning everything "beyond the sentence boundary" is usually covered by the terms *text linguistics*, which roughly corresponds to research on discourse structure.[1] Moreover, relations in text are widely surveyed within the field of stylistics (see e.g. Hausenblas, 1964 or Hausenblas, 1971). In this study, we work with the Functional Generative Description (FGD) approach to discourse research (see Sgall, 1967; Sgall et al., 1969, 1986). Whereas the broader interpretation of the term *discourse* is usually roughly equal to text (as in *discourse structure*, *discourse features* or *discourse coherence*, the FGD uses narrower

[1] Which has to be distinguished from the well known and rather interdisciplinary field of *critical discourse analysis*.

sense denoting semantic relations between propositions (as in *discourse relations*), see Poláková et al. (2013). This means that FGD sees discourse as a sequence of utterances expressed in a communication process and thus an interlinked system of syntactic, semantic and pragmatic relations (see e.g. Mladová, 2008).

5.2.1 Discourse Connectives

From the point of view of this study topic, we need to be especially careful about discourse relations expressed by discourse connectives connecting emotional sentences. The most common discourse connectives used in complex evaluative sentences are conjunctions *a* – 'and' and *ale* – 'but'. As described in detail in Hatzivassiloglou and McKeown (1997), these two connectives are more likely to conjoin structures with similar (case of *a* – 'and') or opposite (in case of *ale* – 'but') polarity orientation. We call this *Semantic Consistency Principle* and it is explained by the following examples (for illustration purposes simplified only to constituent coordinations):

(66) Marie je [příjemná]+ a [milá.]+
Mary is [nice]+ and [kind]+
'Mary is nice and kind.'

(67) Marie je [zlá]– a [sobecká.]–
Mary is [mean]– and [selfish]–
'Mary is mean and selfish.'

(68) # Marie je [příjemná]+ a [sobecká.]–
Mary is [nice]+ and [selfish]–
'Mary is nice and selfish.'

Whereas structures (66) and (67) are perfectly acceptable from the semantic point of view, structure (68) can be considered incorrect. The incorrectness holds for both positive/negative (*nice and selfish*) and negative/positive (*selfish and nice*) combinations.

For a conjunction *ale* – 'but' the semantic principle works the following way:

(69) Marie je [příjemná,]+ ale [sobecká.]–
Mary be [nice]+ but [selfish]–
'Mary is nice but selfish.'

(70) # Marie je [zlá,]– ale [sobecká.]–
Mary be [mean]– but [selfish]–
'Mary is mean but selfish.'

For the incorrect structures, same polarity orientation in both parts does not matter. It is inappropriate to combine either negative/negative (*mean but selfish*) or positive/positive (*nice but kind*). On the other hand, in case of the opposing polarities,

Same polarity orientation	Opposite polarity orientation
a – ‘and’	*ale* – ‘but’
ani – ani – ‘neither – nor’	*avšak* – ‘yet’
ba – ‘even’	*i když* – ‘even though’
i – ‘and’	*jenže* – ‘but’
jak – tak – ‘so – as’	*jenomže* – ‘but’
jakož i – ‘both – and’	*nýbrž* – ‘but’
jednak – jednak – ‘partly – partly’	*leč* – ‘but’
nejen – ale i – ‘not only – but also’	*ovšem* – ‘yet’
stejně jako – ‘as well as’	*přestože* – ‘although’

Table 5.5: Discourse connectives having impact on polarity orientation

it seems that the ordering of particular parts can have an impact on semantic meaning of the structure. When we combine positive/negative parts (*nice but selfish*), we can still consider Mary to be rather nice. But the combination of negative/positive (*selfish but nice*) does not give a feeling that Mary is rather selfish. More likely, the overall polarity of the evaluation would be positive. However, this kind of analysis is tightly connected to the lexico-semantic nature of the conjoint arguments and therefore, it can be rather individual. Also, we admit that there exist exceptions from the Semantic Consistency Principle, i.e. *but* in gradation:

(71) Byli jsme ne potěšení, ale přímo nadšení.
‘We were not just pleased, but even excited.’

Anyway, this principle is successfully used in automatic detection of synonyms and antonyms (see e.g. Izumi et al., 2014) and can also be employed in the form of a set of the rules in state-of-the-art sentiment classifiers (see Section 9.2). As for Czech, it has been explored by Přikrylová et al. (2016a) and partly in Přikrylová et al. (2016b).

Apart from *a* – ‘and’ and *ale* – ‘but’, there are other discourse connectives having impact on polarity orientation of evaluative structure. Table 5.5 shows the most common discourse connectors and their expected relations to the polarity of an utterance.

Again, we are aware of exceptions, e.g. *ani – ani* – ‘neither – nor’ in emotionally neutral utterances:

(72) Film nebyl ani dobrý, ani špatný.
‘The movie was neither good nor bad.’

Moreover, the same discourse relationships between emotional structures with the same or opposite polarity can be expressed by alternative lexicalizations of discourse connectives such as *navzdory tomu* – 'in spite of' or *v rozporu s tím* – 'in conflict with this' (see Rysová, 2012).

(73) [Navzdory tomu], že dobře vypadáte, jste pěkný posera.
'In spite of the fact that you look good, you are a sissy.'

(SYN2010)

5.2.2 Attitude Markers

Attitude markers are expressions or prosodic structures that speakers use to express their attitude towards the semantic content of their utterance. In Czech, the most common attitude markers are evaluative particles and stance particles like *naštěstí* – 'fortunately', *bohudík* – 'luckily', *bohužel* – 'unfortunately', *želbohu* – 'alas'. The issue of Czech attitude markers is described in detail in Janečka (2014). As explained in Chapter 4, attitude markers have crucial influence on sentence polarity and thus are very important for automatic detection of emotions (see Chapter 9).

5.2.3 Coreference Relations

Coreference is a referential agreement of two or more instances in a text. When analyzing structures expressing emotional meaning, coreference is important mainly when it concerns source or target of evaluation, see the following examples:

(74) Petr nesnášel učení. Třídní ho štvala.
'Peter didn't like studying. He hated his class teacher.'

As illustrated below, Czech is a pro-drop language, i.e. it does not always expresses pronouns at the surface layer of the text. Therefore, we need to be especially careful when detecting sources or targets of evaluation automatically.

(75) Nemám rád Pavla. (On) je trapnej.
'I don't like Paul. He is awkward.'

In this chapter, we described semantics of emotional structures, which is tightly connected to pragmatics, a matter of the following chapter.

6

Pragmatic Aspects

Pragmatics is defined as studying language meaning in context. In this chapter, we explore how emotional meaning is being constituted within pragmatic context. By *context*, we do not refer to the linguistic aspects of an utterance (i.e. its grammatical characteristics) which are described in other chapters of this monograph. Rather, we study situational context and communication means used both by the speaker and the hearer.

6.1 Context

Context means the manner, time, place, participants and generally the whole situation in which the utterance is produced. This includes also the genre of the utterance. In terms of emotional meaning survey, we need to keep in mind that some genres, such as a review or a judgment, are more likely to contain emotional meaning than others. For illustration of how the situational context influences emotional meaning and the polarity of emotion expressed, see the following example:

(76) Dneska je hezky.
'Today is a lovely day.'

This sentence can have three very different interpretations concerning its emotional meaning depending on a situational context.

- positive evaluation of the situation: the speaker sits on a bench in a park, taking a sunbath
- negative evaluation of the situation: the speaker lies in a hospital room and cannot go out
- irony: it is raining

Since we are usually not able to understand situational context just from a single written sentence, we say that all the language utterances are potentially ambiguous in this respect. One of the forms of such ambiguity is also known as irony (see the following section).

Situational context also includes context of the given conversation, meaning the speaker(s) and the hearer(s). This involves *pragmatic competence*, i.e. the ability of the hearer to understand speaker's intention so that the communication process could be considered successful, as described in detail in Section 6.3 and Section 6.4. In this respect, emotional meaning is one of the most important meanings to be transmitted. Also, emotive function is considered one of the basic functions of language, as described in detail in Jakobson (1960).

6.2 Irony and Sarcasm

The core of irony lies in the difference between what is written or said and what is actually meant. For the purposes of this study, we divide irony into two basic categories: *situational irony* and *verbal irony*.

Situational irony is a discrepancy between what is being described and what is really happening, as well illustrated by the above example (76). In terms of evaluation, situational irony usually switches the polarity of evaluation and can change it from negative/positive to neutral, depending on context. Therefore, the sentence *Go read the book!* can have a positive connotation when found in a book review, and a negative connotation when occurring in a movie review (as discussed in Pang and Lee, 2008). Also, when looking for a "rubbish hotel in Madrid", we can find both positive and negative reviews of a hotel made out of trash.[1]

In the so-called verbal irony (which, however, also depends on situational context), the meaning intended by the speaker is totally different (or even opposite) from the literal meaning of the utterance, like in the following sentence provided that Martin cannot cook at all.

(77) Martin výborně vaří.
'Martin is a great chef.'

From the point of view of emotional meaning, we can say that these kinds of ironic statements very often involve an explicit expression of one attitude or evaluation, but with indications in the overall speech-situation that the speaker intends a very different, and often opposite, attitude or evaluation. These indications can be e.g. intensifiers such as *pěkně* – 'pretty', *hrozně* – 'terribly', *strašně* – 'awfully', *příšerně* – 'horribly' (as described in detail in Section 3.2.1), particles like *teda* – 'really' or adverbials like *opravdu* – 'really', see the following examples:

(78) To je pěkně bezva.
'That's pretty smashing.'

(79) To je teda dobrý!
'That's really great!'

(80) To se ti opravdu povedlo!
'You really succeeded in this!'

[1] http://www.uniqhotels.com/corona-save-beach-hotel

The contrast between the lexical content and the real polarity of evaluation is often used also in idioms (e.g. *být příjemný jako osina v zadku* – 'to be as pleasant as a burr under somebody's saddle', i.e. to be unpleasant).

Although these lexical means of irony are domain-independent, we can generally assume that irony is a domain dependent phenomenon and thus it is difficult to be revealed not just by machines, but even by people. For instance in the following example, if we are not experts in pickaxes, it can be difficult for us to estimate whether this sentence is ironic or not:

(81) Tenhle krumpáč je výborný!
'This pickaxe is awesome!'

Since we do not have any corpora of sentences that were pronounced with different intention than the one which is expressed by their lexical structure, we cannot reliably determine the structure of ironic meaning automatically. However, there exists some research in this area, see e.g. Filatova (2012) or Reyes et al. (2012).

Similar means as described above are also used in sarcasm, which is often considered an intensified form of irony. Also, they can be used in the so-called *damning with faint praise*, meaning a praise which is too calm or marginal to be considered praise at all:

(82) Gratulujeme! Jste stejně úspěšný jako ostatní návštěvníci našeho webu!
'Congratulations! You are as successful as all the other visitors of our web page!'

Apart from the above mentioned features, ironic or non-ironic (sarcastic or non-sarcastic) meaning always depends on intonation and generally all the suprasegmental features of a given utterance. However, these features are beyond the topic of this study.

Aside from irony, situational context plays role in construing emotional meaning when the speaker or writer reacts emotionally to something happening in the real world. In these situations, people usually use the words like *Do háje!* – 'Damn it!', *Sakra!* – 'Shit!' etc. when reacting negatively or *Bezva!* – 'Cool!', *Senza!* – 'Great!' when reacting positively.

6.3 Speech Acts

Mostly, communication between a speaker and a hearer is a matter of the Speech Act Theory (see Austin, 1975). In Speech Act Theory, an utterance is not perceived as a demonstration of language as an abstract phenomenon but rather as a means of communication. Generally, speech acts include acts such as congratulating, appraisal, promising, warning, apology etc., i.e. highly emotional acts.

There are three basic levels of speech acts analysis:

- locutionary, concerning the actual performance of the utterance and its ostensible content
- illocutionary, meaning the illocutionary force or the real intention of the utterance
- perlocutionary, the actual effect of the utterance (persuading, scaring etc.)

For instance, the sentence *This carrot cake is to die for!* using a phrase about dying (locution) is intended to evaluate the carrot cake as very good (illocution) and may serve as an invitation for the hearer to have a piece (perlocution).

From the point of view of exploring emotions, the illocutionary act is the most important, containing the intention of emotion expression. The illocutional speech acts are further divided by Searle (1969) into several categories. Among others, Searle recognizes *expressive illocutional acts*, i.e. speech acts that express speaker's attitudes and emotions towards the proposition, e.g. congratulations, excuses and thanks.

Generally, illocutionary act or in other words speaker's or writer's intention can be expressed by several means, most of them described in another chapters of this book. These means are:

- performative formulae, i.e. an utterance that not only describes a given reality, but also changes it (*Hereby, I name her Susanne.*)
- lexical means
- modal verbs
- modal particles
- grammar: tenses, moods, negation, gender and other grammatical features
- suprasegmental features

We pay special attention to illocutionary acts since the intended emotional meaning can also be concealed in the context and thus difficult to get encoded. For instance, a sentence *If he's in, I'm not coming* can actually mean *I don't like him*.

When surveying emotional meaning on real data, we often encounter the problem of the difference between intention and interpretation. Let us illustrate this on real-life examples. We studied data from online news and reviews on a retail server, written by human reviewers. We let human annotators mark the data with positive or negative polarity (i.e. to guess writer's intention). Sometimes, it was very difficult for the annotators to agree on the type of emotion expressed. In case of the online news, they were dealing with a controversial political situation before Czech parliamentary elections. Therefore, it was difficult for the annotators (i.e. addressees) to abstract away from their political preferences. Whereas for one of them the sentence *Kalousek says it is possible* could have positive meaning, for another it can be totally negative. Since the annotators did not know the author's background and stances (i.e. situational context), it was not easy for them to guess his or her intention. For more details, see Chapter 8.

6.4 Conversational Maxims

When dealing with the data from the retail server Mall.cz, we experienced many problems concerning miscommunication between the speaker and the hearer. Mostly, these were cases in which the speaker (writer) took into account neither the hearer (addressee), nor the genre of the utterance etc. This phenomenon is widely described in pragmatics and is often mentioned when speaking about the cooperative principle which guarantees smooth communication (Grice, 1978). The cooperative principle can be divided into four maxims:

- Maxim of Quality – a speaker should not say what he believes to be a lie
- Maxim of Quantity – an utterance should be adequately informative
- Maxim of Relation – an utterance should be relevant
- Maxim of Manner – an utterance should not be awkward or ambiguous

Very often, more than one maxim is broken within a single utterance, but for illustration, each example below (taken from Veselovská and Hajič Jr., 2013) breaks a single maxim only.

Maxim of Quality is being broken when the reviewers at the retail server do not have enough evidence that the product is good or bad, but still try to review it within the required positive or negative category:

(83) Positive category:

Meteostanici mám jako dárek pro manžela, vyzkoušela jsem ji jen krátce při převzetí, ale myslím, že je super.

'I bought the meteostation as a present for my husband and I tried it out just quickly after I received it, but I think it is good.'

(Mall.cz)

Maxim of Quantity is usually being broken when the writer gives more information than needed (in the following example, Maxim of Relevance is also being broken, since the authors gives a review of dryers in general instead of the particular product):

(84) Positive category:

Nevím, jak jsem mohla bez sušičky být. Haní ji jen ten kdo ji nemá, nebo zhrzená manželka, když jí nechce manžel sušičku koupit. Úspora času, sice něco se musí žehlit, ale minimálně. Za sobotu jsem stihla usušit ložní prádlo, včetně obalů z matrací a lůžkovin (polštáře, deky) a ještě jsem měla spoustu času.

'I don't know how I could have lived without the dryer. Only those who don't have it defame it, or the turned down wives whose husbands don't want to buy it for them. It saves time, some things still need to be ironed, but very little.

I dried the bed linen during Saturday, including the mattress and bed linen cases (pillows, blankets) and I still had plenty of time.'

(Mall.cz)

Also, we often experience Maxim of Relation violation:

(85) Positive category:
Nemohu hodnotit, zboží jsem pro poškození vrátil.
'I cannot review this, I sent the goods back since it was damaged.'

(Mall.cz)

Maxim of Manner is usually violated when using vulgarisms or overreacting:

(86) Positive category:
Někdo píše SNAD dobrá značka???? Tato značka je mezi mraznicemi a ledniceni jednoznačná 1.
'Anyone said QUITE a good brand???? This brand is number one among freezers and fridges.'

(Mall.cz)

All maxims can be broken deliberately, e.g. in order to make a joke when everybody share the same context, or non-deliberately, i.e. with no recognized intention. This often happens when people do not share the conversational implicatures, i.e. the same assumptions for successful communication.

This chapter described the most problematic components of emotional meaning, namely its pragmatic aspects. We emphasized a role of a context and explained position of evaluative utterances in speech acts. Pragmatic aspects usually cause major problems when building practical applications. This topic is further discussed in Chapter 8.

7

Formal Representations of Emotional Structures

The aim of this chapter is to propose formal representations of the emotional structures described earlier in this monograph. We connect features from all the levels of linguistic description at which emotional meaning is being expressed, from lexical characteristics to pragmatic properties, to get more general picture of emotional expressing. Moreover, the exact frameworks can be further applied when investigating how individual means cooperate in context.

For this purpose, we choose two linguistic theories and their formal frameworks. First, we suggest annotation of emotional structures within the Prague Dependency Treebank, which is based on Functional Generative Description. Second, we capture emotional structures within the framework of Construction Grammar, proposing a new construction capturing subjective stance.

7.1 Prague Dependency Treebank

Prague Dependency Treebank (PDT, Hajič et al., 2006, Bejček et al., 2011, Bejček et al., 2013) is a corpus of Czech journalistic texts originally taken from the Czech National Corpus, with a multilayer annotation, including dependency trees and syntactic functions. PDT is built on the theoretical framework of Functional Generative Description, a dependency-based description proposed by Petr Sgall and his school in 1960's (FGD, Sgall, 1967, Sgall et al., 1969). FGD is based on the integration of the distinct layers of linguistic analysis, from phonetics and phonology to morphonology and morphology to surface syntax and tectogrammatics (deep syntax). The inclusion of a tectogrammatical level is one of the distinguishing characteristic of FGD. PDT uses (with some adaptations) an analytical and tectogrammatical layers of the description, in addition to two more structural layers (w-layer for words and m-layer for morphology). PDT thus consists of the following layers (from the more abstract – higher, to more concrete – lower):

- *tectogrammatical layer* (t-layer, 0.8 million tokens) – deep syntax capturing linguistic meaning, roughly corresponding to FGD's tectogrammatical layer
- *analytical layer* (a-layer, 1.5 million tokens) – surface syntactic annotation
- *morphological layer* (m-layer, 2 million tokens) – full morphological annotation
- *word layer* (w-layer) – layer of source texts, the tokenized plain text

As in FGD, modelling the relation of form and function, PDT also treats the lower levels as forms of the higher layers and the higher layers as functions of the lower layers.

All types of annotation mentioned above can be useful in terms of emotional meaning survey. When investigating emotional structures, we have to take into consideration that the most significant syntactic (and hypersyntactic) features important in identification of sentence polarity are negation, sentential modality marking, discourse relations, intersentential coreferential relations and depth of the polarity item in the tree. Having access to all this information at once would allow testing various hypotheses, for example whether the embeddedness of a polar node in a tree influences the polarity of a sentence.

However, when capturing the syntacticosemantic nature of emotional sentences, we mostly make use of the tectogrammatical layer. Despite the fact that the tectogrammatical layer seems already rather overburdened with linguistic annotation, it seems useful to keep the polarity identification at the same layer as the annotation of coreference and discourse relations, because these phenomena are closely related and are essential e.g. when searching for source and target of the emotional statement.

The idea of using a dependency treebank in emotional meaning research is not completely new (see e.g. Joshi and Penstein-Rosé, 2009). However, evaluative meaning is influenced by many layers of language, including morphology, surface, and deep syntax and treebanks annotated to such depth are rather rare.

7.1.1 Annotating Emotional Structures in Prague Dependency Treebank

In order to get the idea about how a generalized Czech emotional structure looks like, we first need to get the evidence of particular emotional structures from corpora of real language. To obtain a set of evaluative sentences in the PDT, we use the Czech Subjectivity Lexicon (see Section 8.1). All SubLex items found in the treebank data were marked there as potentially evaluative. The output was then manually refined following the steps described below. As newspapers, the main source of PDT texts, are not typically very evaluative, this step also served as a quick screen determining whether there are any emotional structures in PDT at all.

Using the Czech Subjectivity Lexicon, we have identified 71,440 evaluative tokens out of 833,193 tokens at the t-layer of the PDT, which means 33,066 potentially evaluative sentences. Since this ratio is relatively high, we have to verify whether the evaluative items are actually used in an evaluative context (the first very brief visual check indicates most of them are not). To review the treebank data manually, we built *PML_T_Sentiment*, a new extension for *TrEd*, a tree annotation editor (freely available at `http://ufal.mff.cuni.cz/tred/`). The extension provides a GUI supporting the entry and modification of sentiment related information, see Fig. 7.1.

All polarity items obtained from the subjectivity lexicon and found in the dependency data are highlighted, so that annotators could easily check one occurrence after

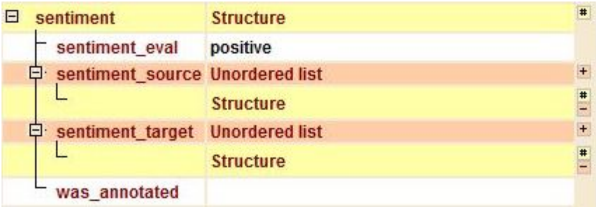


Figure 7.1: Sentiment annotation GUI

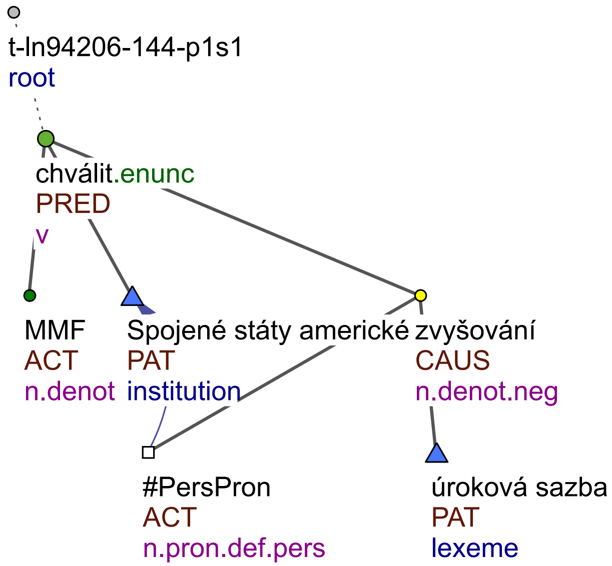
another. Also, the preliminary polarity from the lexicon is assigned to them (using two different colours, green for positive polarity and red for negative polarity). Moreover, the evaluative chunk of the text appearing above in the tree editor is marked with yellow. If the polarity is correct in the given context, the annotator simply confirms this fact. If the actual polarity does not correspond to the polarity from the lexicon, it can be altered manually by changing the value of the attribute `sentiment_eval` (attribute concerning the anchor of evaluation, i.e. evaluative expression). The annotator can choose from various options, depending on the polarity of the given evaluative item: `POS` for positive, `NEG` for negative or none when the item is not evaluative at all in this particular context. Once an item was checked/corrected, it is marked both visually and by setting the attribute `was_annotated` to the value of 1.

As for the `sentiment_source`, the assigned value can either be the identifier of the source node in the treebank, or is_external, when the source is e.g. the author of the text. This holds also for the `sentiment_target` attribute. The representation of an evaluative sentence in the treebank can be seen in Fig. 7.2.[1] The TrEd extension is described in detail in Appendix D.

Apart from the target and source description, both source and target nodes can be marked with arrows of different colours, which are interlinked with arrows for coreference (when the real source/target is situated outside of the sentence). This is a great advantage, since it allows us to find the original source and target, which would not be possible with plain text, effortlessly. Moreover, it is much easier to assign the target attributes, no matter how far they are from the governing word in the surface structure. In the treebank, one can see the whole dependency subtree immediately.

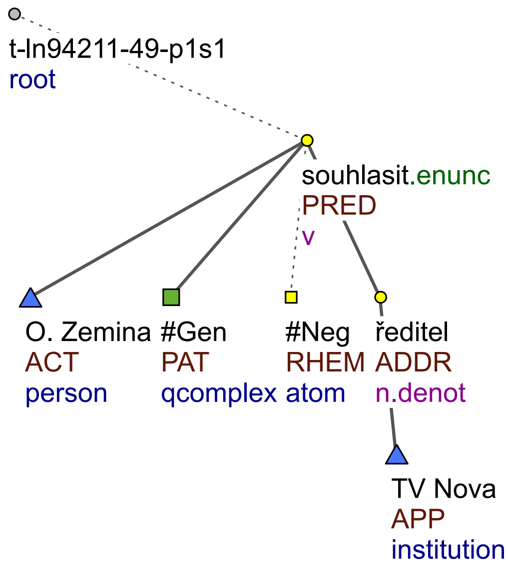
Another advantage of using syntactic structures could be an easy detection of negation. On plain text, both sentential and constituent negation in Czech is usually a part of the verb and thus it is difficult to distinguish between the two, i.e. to delimit the negative scope, as explained in Section 4.6.

[1] In Figures 7.2 and 7.3, we present the collapsed versions of the representations in which the multiword expressions are expressed as the single nodes and marked with the blue triangles. Functors ACT, PAT and ADDR stand for Agent, Patient and Addressee. Functor PRED stands for predicate. CAUS marks cause, APP refers to the person or thing something or someone belongs to and RHEM denotes rhematizer.



MMF chválí Spojené státy za zvyšování úrokových sazeb.
'MMF praises USA for increasing open market rates'

Figure 7.2: An evaluative sentence in PDT



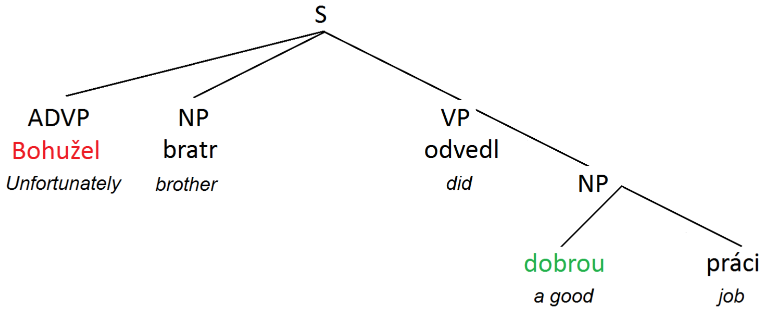
O. Zemina nesouhlasí s ředitelem TV Nova.
'O. Zemina disagress with the direcotr of TV Nova.'

Figure 7.3: Negation scope

This does not hold for the tectogrammatical dependency data, where the scope of negation is easily recognizable since it is represented by a separate node (see Fig. 7.3).[2] Therefore, we can detect negated items and, in consequence, switch their polarity (or the polarity of the whole sentence, depending on the negation type).

Also, we can easily observe the influence of the polarity node embeddedness on the overall polarity of a sentence. In Fig. 7.4, taken from Veselovská (2011b), we can see an example of a sentence *Bohužel, bratr odvedl dobrou práci.* – 'Unfortunately, brother did a good job.'. There are two polarity items in the structure, one positive (*dobrou* – 'good') and one negative (*bohužel* – 'unfortunately'), but its overall polarity is negative. Thus, we can assume that the higher a node appears, the stronger influence it has. However, this hypothesis should be a matter of a further research.

[2] For compact view reasons, the name O. Zemina is represented as a single node.



Bohužel, bratr odvedl dobrou práci.
'Unfortunately, brother did a good job.'

Figure 7.4: Embeddedness influence

7.2 Construction Grammar

As a second framework, we apply the Construction Grammar as introduced by Fillmore (1988) and developed till today (see e.g. Fried and Östman, 2005). As explained by Fried (in press) "Construction Grammar (CxG) is a theoretical approach in which generalizations about linguistic structure are formulated in terms of 'constructions', i.e. conventionalized clusters of features (syntactic, prosodic, pragmatic, semantic, textual, etc.) that recur as further indivisible associations between form and meaning." (p. 974) "It is a constraint-based, non-derivational, mono-stratal grammatical model that also seeks to incorporate the cognitive and interactional foundations of language" (p. 1005). CxG works on the assumption that the form is connected not only to the meaning, but also to the (mainly cognitive) function. The function is then accomplished by different means, no matter whether they belong to the core or periphery of language. CxG accounts for all types of linguistic expressions. It argues that they have the same informative value, since semantically unusual expressions can share syntactic properties of common structures and syntactically transparent structures can have semantic constraints. In terms of this study, this can be useful e.g. when exploring evaluative idioms.

The basic unit of the CxG theory is not the syntactic structure, but rather the grammatical construction, i.e. the conventionalized unit of language features which together form a structure. A grammatical construction has the form of attribute-value matrix, as indicated below in Fig. 7.5. Different domains are assigned different attribute-value pairs, such as *semantic role – agent* in semantic domain etc.

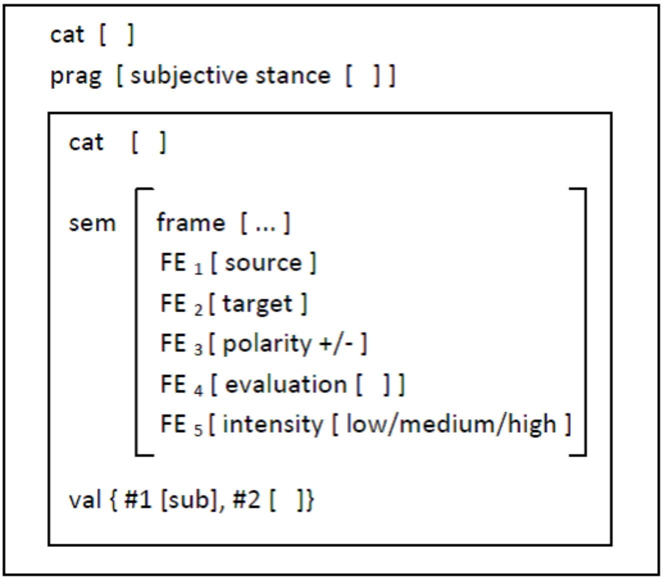


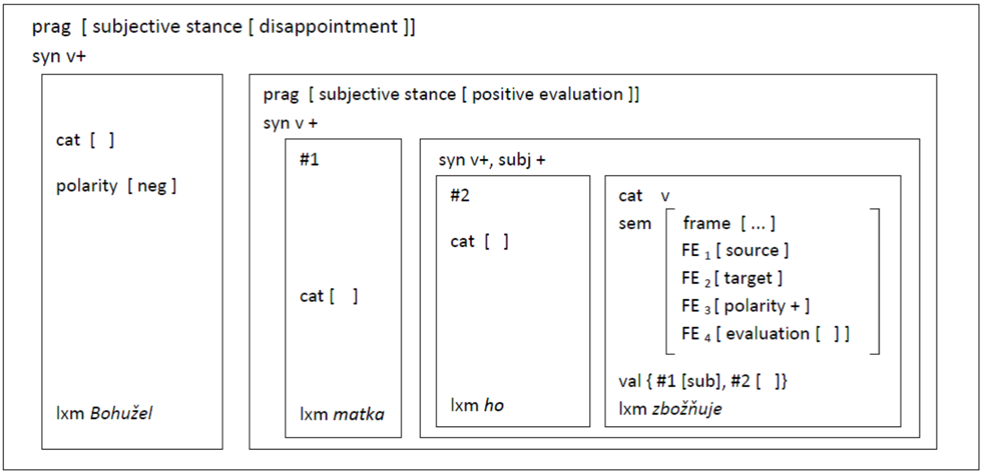
Figure 7.5: Subjective construction

In this part of the present study, based on Veselovská (2014b), we use CxG to capture the relationship between structure, meaning, and the use of subjective expressions. Also, we depict the morphological context of evaluative items, and represent possible connection between the syntactic structure and the polarity of the given sentence. Moreover, we use constructional framework for the description of the relationship between the form of a given sentence and its pragmatic content, which is one of the elementary pairing units in the CxG theory. This means that we describe emotional meaning from the contextual point of view, meaning not only sentential context, but also presumptive situational context.

For the purpose of the analysis of evaluative structures, we propose a new type of construction, the *Subjective construction* (see Fig. 7.5), and integrate it with the CxG representation, together with relevant attributes. A subjective frame also bears the form of a common attribute value matrix. The new matrix contains not only well-examined attributes, but also new attributes assigned especially for subjectivity.

Concerning the abbreviations used in Fig. 7.5:

- cat assigns the frame with morphological category
- prag defines the frame in terms of pragmatics
- sem is used to classify semantic features of the frame
- FE stands for frame element



Bohužel, matka ho zbožňuje.
'Unfortunately, mother adores him'

Figure 7.6: Evaluative Construction

- val means valency
- sub represents subject
- numbers #1 and #2 illustrate participants of the valency frame

In the schema, we introduce evaluative components of the construction: we suggest that pragmatically, the construction captures a subjective stance. The subjective stance attribute can have different values assigned, e.g. *approval*, *astonishment*, etc. Semantically, the frame consists of several frame elements: *the source*, *the target*, *positive or negative polarity*, *the type of evaluation*, and also the *intensity of the state that is expressed*. We establish a special slot for valency in case that the evaluative element is a verb. Different parts of a valency frame can be identical to the particular frame elements.

To be more specific, let us propose the constructional analysis of the evaluative structure *Bohužel, matka ho zbožňuje* – 'Unfortunately, mother adores him'; Fig. 7.6.

Syntactically, it is the finite sentence (syn v+). Lexemes (lxm) *matka* – 'mother' and *ho* – 'him' are participants of the valency frame of the verb *adore* – 'zbožňovat'.

From the point of view of emotional meaning, this structure contains one negative element *bohužel* – 'unfortunately' and one positive element *zbožňovat* – 'to adore'. However, the overall evaluation expressed by the construction is negative. The structure seems to suggest that the higher the frame is in the construction, the more influential it is. However, as in case of Fig. 7.4, this hypothesis should also be a matter of a further research.

II Computational Applications

8

Creating Evaluative Resources

This chapter is dedicated to the emotional language data resources we created and use in the present study. We use these sources both for examples of real utterances and for natural language processing tasks described in Chapter 9. Summarizing statistics about the sources' properties and annotation guidelines can be found in Appendix A and Appendix B. The other publicly available corpora used in the study are described in Section 2.3.

8.1 Czech Subjectivity Lexicon

In this section, based on Veselovská (2013), we describe particular stages of building the Czech Subjectivity Lexicon: translation from English, basic cleanup and advanced cleanup of the data.

In this work, we use the Czech SubLex as a source of domain-independent evaluative items surveyed e.g. in Chapter 3 and Chapter 4. Also, we use the lexicon as a lexical source for sentiment analysis tasks described in Chapter 9.

8.1.1 Building the Original Version

The Czech Subjectivity Lexicon (Czech SubLex, Veselovská, 2013; `http://ufal.mff.cuni.cz/seance/data`) is a list of Czech subjectivity clues, i.e. positive and negative evaluative items. The list contains 4,625 items (1,672 positive, 2,863 negative and 90 with both polarities assigned) together with their part of speech tags, polarity orientation and the original English expression.

Inspired e.g. by paper by Banea et al. (2008b), mentioned in Section 2.2.1, we obtained the core of the Czech Subjectivity Lexicon by automatic translation of a freely available English subjectivity lexicon, also known as the Pittsburgh subjectivity clues, introduced in Wilson et al. (2005b); `http://mpqa.cs.pitt.edu/lexicons`. The original lexicon, containing more than 8,000 polarity expressions, is a part of OpinionFinder (Wilson et al., 2005a; `http://mpqa.cs.pitt.edu/opinionfinder`), a system for subjectivity detection in English. First, the authors compiled a collection of seed words from a number of both manually and automatically identified sources (see Riloff and Wiebe, 2003). Second, the patterns and words were expanded iteratively. Afterwards, various scoring mechanisms were used to ensure the extracted words are in the same semantic category as the seed words.

For translating the data into Czech, we only used the CzEng 1.0 parallel corpus (Bojar and Žabokrtský, 2006) containing 15 million parallel sentences (233 million English and 206 million Czech tokens) from seven different types of sources automatically annotated at surface and deep layers of syntactic representation. By translation, we gained 7,228 potentially evaluative expressions.

8.1.2 Refining the Lexicon

However, some of the items or the assigned polarities appeared rather unreliable at first sight. For this reason, the lexicon has been proofread by an annotator, who excluded all obviously non-evaluative items such as *zelený* – 'green', which appeared in the corpora probably due to errors in semi-automatic translation. This resulted in the first applicable version of the lexicon with 4,947 evaluative expressions.

After this, the lexicon has been manually checked again for further incorrect entries by an experienced annotator. Below we mention the most significant types of inappropriate entries,

Items with rare evaluative meaning The most common problem was the inclusion of items that are evaluative only in a rare or infrequent meaning or in a specific semantic context whereas they usually represent non-evaluative expressions (e.g. *bouda* is in most cases used as a word for a 'shed', though it can as well mean 'dirty trick'). This concerns also the cases where the word is a part of a multi-word expression. The main criterion for marking the given item as evaluative was its universal usability in various contexts. Therefore, we excluded most of the domain-dependent items. The non-evaluativeness of the item was sometimes caused by a wrong translation of the original English expression. The correct translations were added manually, in case they had not been present in the lexicon yet.

Items with twofold polarity On the other hand, we found a lot of items with twofold polarity. Some of them were intensifiers like *neuvěřitelně* – 'incredibly' or quantifiers like *moc* – 'a lot' which should not be present in the lexicon, since their polarity depends on the polarity of the following word. However, we also detected intensifiers like *pěkně* – 'pretty' or *strašně* – 'terribly', which can in certain cases be used within the sentence as evaluative adverbs (see Section 3.2.1). This kind of intensifiers should be distinguished later on based on their context. Moreover, we found words which are frequently connected both with positive and negative meaning (e.g. *[dobré/špatné] svědomí* – '[clear/guilty] conscience'. Different polarities of such words should be recognized by recording them in the lexicon together with their prototypical collocations. There were also other instances falling under this category of dual polarity, such as ambiguous words which can be used both in positive and negative meaning – e.g. in *využít někoho* – 'to abuse somebody' (negative), and *využít příležitosti* – 'to take the opportunity' (positive). Such expressions seem to be crucial for more fine-grained sentiment analysis (see e.g. Benamara et al., 2007). As we indicated in Chapter 3, further research into their semantic features and corpus analysis of their collocations is needed.

Items with wrong polarity Another problem concerned words assigned an incorrect polarity value. For example, we observed diminutives marked with positive polarity although they are used in negative (mostly ironic) sense – e.g. *svatoušek* – 'goody-goody'. Another large group of items marked with wrong polarity consisted of incorrect translations of negated words like *nečestný* – 'not honest', *nemilosrdný* – 'unmerciful' etc. In this case, the translating system did not take into account the negative particle preceding the given word and assigned a positive polarity.

After the manual refinement, we obtained 4,625 evaluative items. The most frequent items in this set are nouns (e.g. *hulvát* – 'a boor', 1,851 occurences) followed by verbs (e.g. *mít rád* – 'to like', 1,549), adjectives (e.g. *špatný* – 'bad', 773), adverbs (e.g. *dobře* – 'rightly/well/correctly', 440) and particles (e.g. *bohužel* – 'unfortunately', 12). The final lexicon was evaluated within classification experiments with satisfactory results, as described further in this study.

8.2 Manually Annotated Data

Apart from the above mentioned corpora, the current linguistic analysis of emotional meaning in Czech is also based on the manually annotated evaluative data from different domains. These data have been prepared primarily for automated detection of evaluation in natural language texts in Czech (see Chapter 9). In this section, partly based on Veselovská et al. (2012) and Veselovská et al. (2014), we describe annotation process concerning evaluative resources.

8.2.1 Annotation Practice

There are three basic levels at which subjectivity can be annotated (the plus or minus sign after the square brackets indicate the overall polarity of a sentence):

- the expression level: *I [like]+ strawberries.*
- the sentence level: *[I like strawberries.]+*
- the document level: *[I like strawberries. They are sweet, juicy, succulent, and are a very nostalgic fruit for me. Freshly picked strawberries always remind me of summer... No school, just relaxation, and I only get them freshly picked like that where my grandma lives which reminds me of all the wonderful times I've had there at my favorite place in the whole world.]+*

In the subjectivity annotation project, we annotated the sentence level of plain text. As explained e.g. by Wiegand and Klakow (2009), sentence level annotation allows us to explore many useful linguistic features, such as part-of-speech information, clause types or depth of word constituents.

We distinguished three functional evaluative components that need to be identified, as mentioned earlier in this work:

- *the source*, i.e. the person or entity that expresses or experiences evaluation
- *the target* that is evaluated
- *evaluative element*, i.e. word or phrase inherently bearing a positive or a negative value

In contrast to Wilson (2008), who provides a detailed analysis e.g. concerning intensity of emotion, we restrict our survey to evaluative opinions only. On the other hand, we annotate some further aspects concerning a fine-grained subjectivity analysis, mainly expressions bordering the area of sentiment analysis, such are *good/bad news*, i.e. words expressing generally positive or negative situations or facts (like *válka* – 'a war'), or *elusive elements*, i.e. expressions bearing evaluative power, but such that cannot be described in terms of standard polarity values – e.g. words like *kontroverzní* – 'controversial', *průměrný* – 'average' or *nezvyklý* – 'unusual'.

The annotation practice lies in manual tagging of appropriate text spans, and is performed by two independent experienced annotators (later referred to as A and B). For the detailed guidelines and information about data formats see Appendix B.

8.2.2 Datasets

The primary motivation of the research presented in this publication was to create a tool for detecting the way news articles influence public opinion. Therefore, we initially used data from the Aktualne.cz news website. However, the analysis of such texts has proven to be a rather difficult task both in terms of manual annotation and automatic processing, because the authors of Aktualne.cz avoided strongly evaluative expressions and sometimes even any explicit evaluation. For this reason, we also decided to use reviews from the Czech movie database, CSFD.cz, since movie reviews have been successfully used in the area of sentiment analysis for many other languages, see e.g. Thet et al. (2009). As both sets of the manually annotated data were rather small, we also used auxiliary data, namely larger dataset of domestic appliance reviews from the Mall.cz retail server and other datasets described in Section 8.3.

8.2.2.1 Aktualne.cz

There are approximately 560,000 words in 1,661 articles obtained from the domestic news section of the Czech news website Aktualne.cz. In the first phase, we manually categorized some of the articles according to their subjectivity. We identified 175 articles (89,932 words) bearing subjective information, 188 articles (45,395 words) with no polarity, and we labelled 90 articles (77,918 words) as 'undecided'. There are 1,208 articles which have not been classified yet. Most of these data are not intended for manual processing but for various unsupervised machine learning methods in potential natural language processing applications.

The annotators annotated 410 segments[1] of text (6,868 words, 1,935 unique lemmas). These segments were gained from 12 randomly chosen opinion articles from Aktualne.cz. The segments are mostly sentences, but they also contain headlines and subtitles. In the sequel, we refer to annotation items as segments.

At the beginning, we tried to annotate all polar states that elicit a reaction from the reader. The primary instruction for the annotators was simple: *Should you like or dislike an entity occurring in a segment because of what that segment says, tag the entity accordingly.* This choice of annotator perspective was motivated by the desired application: if our goal is for the computer to simulate a reader and thus develop sympathies, then the training data should reflect this process. It would also enable us to bypass the issue of identifying sources and assigning some trust parameter to them. However, combined with requirement of impartiality towards the evaluated targets, this choice of perspective did impede the annotators' ability to make judgments about the presence of polarity in segments. The inter-annotator agreement was a little over 0.63 by Cohen's Kappa for all polarity classes, which can be considered a satisfactory result. The annotators tagged about 30% of all the segments in total.

In the examples below, we illustrate not only the polarity of expressions but we also indicate which annotator made the annotation. For example, in (87), both annotators, A and B, agree that *těžit na Šumavě kvůli kůrovci* – 'to mine in Šumava because of a bark beetle' is negative. The annotators were supposed to mark the whole subtree (phrase), if possible (see the annotation guidelines in Appendix B).

(87) Také další vědci si ale myslí, že [těžit na Šumavě kvůli kůrovci]$_{A-B-}$ je chyba.
'The other scientists also think that [to mine in Šumava because of a bark beetle]$_{A-B-}$ would be a mistake.'

We have experienced various problems during the annotation. The easily resolvable ones concerned insufficient clarity of annotators' instructions (e.g. regarding the question whether the annotators should tag the preposition as a part of the target or not), misinterpretation of the annotator instructions, or misinterpretation of some linguistic properties of the text. Due to the generality of the given task the boundary between the latter two phenomena is not very clear. In general it appeared quite difficult for the annotators to abstain from their personal sympathy or antipathy for the given target, especially because the texts deal with the controversial political situation before the Czech parliamentary elections in 2010.

[1] We use *sentence* and *segment* interchangeably in this study. Every sentence is a segment, but not every segment is a sentence as linguistics would have it, as there were items like news headlines or one-word exclamations in the data.

One of the specific problems of our annotation was the fact that all of the annotators had a linguistic background, so they might have tended to tag sentences with some supposedly linguistically interesting polarity item, even though the polarity lay in another expression or the sentence was not subjective at all. See (88):

(88) Vláda schválila [něco jiného]$_{B+}$, než co slibovala.

'The government approved [something else]$_{B+}$ than what it had promised.'

Here the target of the negative evaluation is actually the 'government'. However, annotator A considers the sentence to be neutral. For annotator B, the evaluated target is 'something else' (probably because it is governed by the verb 'approved' which is usually considered a positive term).

Further problems were caused by a vague interpretation of targets in polar sentences: in evaluative structures, there are different levels on which we can determine the targets, see (89):

(89) [Dům]$_{B-}$ [byl]$_{A-}$ před sedmi lety neúspěšně dražen, nyní je v zástavě banky.

'Seven years ago, [the house]$_{B-}$ [was]$_{A-}$ unsuccessfully auctioned; now it has been pledged to a bank.'

Annotator A apparently felt as negative the fact that the house had been offered in the auction, most likely because the auction was unsuccessful, whereas annotator B perceived the house itself as the evaluated entity because it failed in the auction. Here we prefer the second option, since with respect to the overall topic of the document in question and the fact that 'house' is in topic whereas 'auctioned' is in focus, we suppose that the reader will probably evaluate the house rather than the auction.

The above mentioned problems can also be caused by seeing the subjectivity structure *source – evaluation – target* as parallel to the syntacto-semantic structure *agent – predicate – patient*. Although these structures may be parallel (and they, as suggested by the real data, very often are), it is not always the case (see Chapter 5).

Moreover, we found many discrepancies between the local and global polarity – while a part of the sentence might be evaluative, the whole sentence appears rather neutral (or even its overall polarity is oriented in the opposite direction). For instance, in (90), the candidate is supported by the respondents, but only by 13% of them. Since the percentage is rather low, the overall polarity of the sentence can be considered neutral:

(90) V případě jeho kandidatury na tento post by [jej]$_{A+B+}$ podporovalo pouze 13% dotázaných, a to z řad voličů ČSSD a KSČM.

'In case of his candidacy for this post, [he]$_{A+B+}$ would be supported only by 13% respondents, mostly supporters of ČSSD and KSČM.'

In order to improve the annotation scheme, we found necessary to abandon the reader's perspective and to annotate only the explicitly mentioned evaluation instead.

Moreover, we decided to annotate not only targets, but also sources and expressions. Originally, we hoped that taking the reader's perspective could prove advantageous for the identification of those polar indicators which are most relevant for the readers. However, it turned out that it is hard to identify reader-oriented polarity (and its orientation) while keeping the sources and targets anonymous. Therefore we find it more useful to separate the task of identifying subjective structures and the assignment of relevance to the reader.

8.2.2.2 CSFD.cz

In the second phase of the research, we decided to use data more convenient for the evaluation detection task, namely the data from Czech movie database CSFD.cz. In comparison with the previous dataset, the language of these reviews was significantly more evaluative, even though it was much more domain-dependent. To compare the results, we again chose 405 segments and let the same two people annotate them. In this case, the results were slightly better, with Cohen's Kappa 0.66, as opposed to 0.63 for Aktualne.cz. However, we again experienced some problems.

Perhaps the most interesting and most disturbing issue we have encountered when annotating polarity was the annotator inconsistence and mutual disagreement in establishing the borderline between polarity target and polarity expression. A substantial part of inter-annotator disagreement in target identification lies in different perception of the extent of polarity expression with respect to the entity evaluated. This happens especially in copular sentences, both attributive and classifying, see (91):

(91) Tom Hanks je výborný herec.
'Tom Hanks is an excellent actor.'

Sentences in which qualification is not expressed by an adjective alone, but by a combination of adjective and predicative noun (*Tanja is pretty* v.s. *Tanja is a pretty girl*) are known as qualification by non-genuine (indirect) classification, see Mathesius (1975). In such sentences, annotators either tag 'Tom Hanks' or 'actor' or 'Tom Hanks; actor'[2] as targets of the polarity expression 'excellent'. The three alternative solutions illustrate three different, but equally relevant ways of polarity perception. Pragmatically, the real-world entity evaluated is Tom Hanks. Syntactically, it is the head 'actor' that is modified by the qualifying adjective 'excellent'. And semantically, it is the professional abilities of Tom Hanks as an actor which are being evaluated.

(92) Kate Winslet je špatně oblečená.
'Kate Winslet is poorly dressed.'

As in (91), the target of the negative evaluation in (92) is actually both Kate Winslet and the way she dresses herself. At the beginning we have tried to capture this prob-

[2] We use semicolon for appending discontinuous items.

lem by means of copying, i.e. we kept two separate instances of a polar state, one with 'Kate Winslet' as the target and 'poorly dressed' as the evaluation, the other as 'dressed' as the target and 'poorly' as the evaluation. Doubling the polar information though did not appear to be advantageous with respect to the annotators' time expenses. Moreover, the annotators did not succeed in capturing each single instance of the structure in question, therefore we withdrew them from such treatment in favour of the more complex variant of keeping the entity as the target and the attributed quality/ability/profession etc. in the evaluation category.

During the annotation of news articles the annotators expressed the need for a separate category capturing the *good* and *bad news*. It appeared useful to separate sentences involving events commonly and widely accepted as 'pleasant' or 'unpleasant', such as *triumph, wealth* or *death, injury, disease, natural disaster, political failure* etc., from individual subjective statements of sentiment. Interestingly enough, it appeared quite difficult for the annotators to identify a clear-cut borderline between subjective positive/negative opinion and good/bad news, perhaps because of generally widespread metaphorical uses of the '(un)pleasant'. With movie reviews, the situation was easier. First, due to the maximally subjective character of the texts, good/bad news did not appear significantly often, were easily identifiable and did not intervene much into the annotators' decision. Nevertheless, this type of disagreement did occur, e.g. in the sentence *Bůh je krutý.* – 'God is cruel.' or *Dialogy jsou nepřípadné.* – 'The dialogues are inappropriate.'

As expected, the inter-annotator agreement often fails in places where the subjectivity of the sentence is hidden and embedded in metaphorically complex expressions like

(93) Všichni herci si zapomněli mimické svaly někde doma.
'All the actors have forgotten their mimic muscles at home.'

(94) Slovo hrdina se pro něj opravdu nehodí.
'The word 'hero' does not really fit him.'

Moreover, sometimes the annotated polar expression serves the polarity task only within the given semantic domain. Thus, whereas expressions like *špatný herec* – 'bad (actor)' or *špatně (oblečená)* – 'poorly (dressed)' can function universally across different text genres and topics, the expressions like *psychologizující (postavy)* – 'psychologizing (characters)' or *jsou stříhnuty brzo* – 'are edited early' take the concrete polar value according to the presupposition whether we are dealing with a movie review or not. In a different text genre they could easily serve as neutral, non-subjective element, or even they could acquire a different polarity value.

During the annotation of CSFD.cz data we have decided to make two improvements in the original annotation scheme. First, we added two more polarity values, namely NONPOS and NONNEG, for capturing more fine-grained evaluation of the type 'not that good' or 'not that bad', respectively.

(95) Meryl není ani krásná ani výjimečná.
'Meryl is neither beautiful, nor exceptional.' NONPOS

(96) Ironický nadhled v první části vlastně nebyl tak zbytečný.
'The ironic detached view in the first part wasn't actually that pointless.' NONNEG

These additional labels do not equal simple 'bad' or 'good' values, but neither do they refer to a neutral state. Essentially, they describe a situation where the source's evaluation goes against a presupposed evaluation of the reader's. By adding additional values we risk a slight rise in the number of points of annotator's disagreement. On the other hand, we are able to capture more evaluative structures and get a more thorough picture of the evaluative information in the text.

The final set of values is shown in Table 8.1. NEG and POS stand for the extremely negative/positive values. The scale is completed by NONPOS, which is slightly negative, and NONNEG standing for slightly positive.

– –	–	0	+	++
NEG	NONPOS	NEUTRAL	NONNEG	POS

Table 8.1: Polarity level abbreviations

The second improvement was the addition of a special label TOPIC for cases where the evaluation is aimed at the overall topic of the document and there is no other coreferential item in the context to which the target label could be anchored.

(97) Skvěle obsazené, vtipné, brutální, zábavné, nápadité …
'Excellently casted, witty, brutal, funny, imaginative …'

As in the previous case, this label should help us capture more evaluative structures that would otherwise stay unidentified. We are aware of the fact that this label might be helpful only in domains with strong evaluative character (like product reviews), but maybe less useful in case of journalistic texts in general.

8.2.2.3 Mall.cz

In addition to the above mentioned data, we have obtained 10,177 domestic appliance reviews (158,955 words, 13,473 lemmas) from the Mall.cz retail server. These reviews are divided into positive (6,365) and negative (3,812) by their authors. We found these data much easier to work with, because they are primarily evaluative by their nature and contain no complicated syntactic or semantic structures. Unlike the data from Aktualne.cz, they also contain explicit polar expressions in a prototypical use. Furthermore, they do not need to be tagged for the gold-standard annotation.

The Mall.cz data, however, do present a different set of complications: grammatical mistakes or typing errors are frequent and cause noise in the form of non-existing lemmas. Some of the reviews are also categorized incorrectly. However, compared to the problems with news articles, these are only minor difficulties and can be easily solved. For this reason, the Mall.cz data are more suitable for the error analysis task, as performed in Veselovská and Hajič Jr. (2013) and partly described in Section 9.2.2.1.

The experiments described in Section 9.2 were carried out across the three datasets mentioned above. When merging the annotations, we used an "eager" approach: if one annotator has tagged a segment as polar and the other as neutral, we use polar classification. NONPOS and NONNEG are considered NEG and POS, respectively, and segments classified as BOTH and NEG (or POS) stay as BOTH. Varying the merging procedure had practically no effect on the classification.

8.3 Other

In addition to the data described above, annotated at the Institute of Formal and Applied Linguistics, Charles University in Prague, we also used the Czech Facebook dataset compiled at the University of Western Bohemia (see Habernal et al., 2013). This dataset contains 10,000 Facebook posts, of which 2,587 are positive, 5,174 neutral, 1,991 negative, and 248 'bipolar' posts (posts containing both polarities). The set comprises of 139,222 words and 15,206 distinct lemmas.

For experiments concerning opinion target identification in English (see Section 9.3.1) we used the restaurants and laptops reviews provided by the organizers of the SemEval2014 task (Pontiki et al., 2014) concerning aspect-based sentiment analysis. There were 3,041 train and 800 test sentences in the restaurant dataset and 3,045 and 800 test sentences in the laptops dataset.

For experiments concerning opinion target identification in Czech (see Section 9.3.2) we used 1,000 positive and 1,000 negative user reviews from a retail server Alza.cz, manually tagged with opinion targets.

For examples in Chapter 3, we used user reviews or discussion comments from the websites Heureka.cz, Nova.cz and Novinky.cz.

Furthermore, we also work with some non-digital resources like Bajger (1998) or Čermák (2009) to manually extract some other important lexical units expressing emotional meaning.

Concerning the non-Czech affective data for sentiment analysis, one of the most widely used manually annotated corpora is the MPQA corpus described in Wiebe et al. (2005). Another manually annotated corpus is the collection of newspaper headlines created during the SemEval 2007 task on affective text Strapparava and Mihalcea (2007) annotated with the six Eckman emotions (anger, disgust, fear, joy, sad-

	Aktualne.cz		CSFD.cz	
Annotator	1	2	1	2
Neutral	376	358	183	203
Polar	61	79	217	197
Negative	49	62	92	91
Positive	11	16	94	85
Both	1	1	31	21

Table 8.2: Annotator statistics

ness, surprise) and their polarity orientation (positive, negative). For English, there also exist Bing Liu's datasets for sentiment analysis `http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#datasets`, different Twitter datasets annotated with sentiment (`http://www.sananalytics.com/lab/twitter-sentiment/`) and many more freely available emotional data resources.

8.4 Annotation Evaluation and Results

In order to judge manual annotation quality and usefulness of the Mall.cz, Aktualne.cz and CSFD.cz data, we use annotator agreement. On segment level, we measure whether the annotators would agree on identifying polar segments (*unlabeled agreement*), polar segments and their orientation (*labeled agreement*) and whether they agree on orienting segments identified as polar (*orientation agreement*). Additionally, we measure text anchor overlap for sources, polar expressions and targets.

As evaluation measures, we use *Cohen's kappa* κ (Cohen, 1968) and *f-scores* on individual polarity classes for inter-annotator agreement (for measurement details, see Appendix C). We use f-score also for text anchor overlap for both polar expression and source and target of evaluation. The f-score is denoted *f-ntr* for neutral segments, *f-plr* for polar segments, *f-neg* for negative segments, *f-pos* for positive segments and *f-both* for both. Orientation was evaluated as BOTH when an annotator found both a positively and negatively oriented polar state in one segment.

For the Aktualne.cz data, the annotators tagged 437 segments and for CSFD.cz, they annotated 400 segments altogether, as shown in Table 8.2. The agreement on different segments is explained in Table 8.3 and the agreement on Aktualne.cz and CSFD.cz data is demonstrated in Tables 8.4 and 8.5.

	Aktualne.cz	CSFD.cz
Source	0.484	0.750
Polar expr.	0.601	0.580
Target	0.562	0.706

Table 8.3: Overlap measured in f-score.

Agreement	κ	f-ntr	f-plr	f-neg	f-pos	f-both
Unlabeled	0.659	0.944	0.714	-	-	-
Labeled	0.649	0.944	-	0.708	0.593	0
Orientation	0.818	-	-	0.975	0.889	0

Table 8.4: Agreement on Aktualne.cz data

Agreement	κ	f-ntr	f-plr	f-neg	f-pos	f-both
Unlabeled	0.659	0.809	0.850	-	-	-
Labeled	0.638	-	0.806	0.752	0.757	0.371
Orientation	0.702	-	-	0.873	0.876	0.425

Table 8.5: Agreement on CSFD.cz data

On Aktualne.cz, the annotators have better agreement on neutral segments and worse on polar ones than on CSFD.cz. A similar κ would suggest that this can be attributed to chance, though, because the higher prevalence of polar segments in the CSFD.cz data makes it easier to randomly agree on them. However, the text anchor overlap suggests that as far as expression-level identification goes, the annotators were much more certain on the CSFD.cz data in what to "blame" for polarity in a given segment.

Comparing the above mentioned attempts of annotating subjectivity, we realized that the success in inter-annotator agreement is dependent on the annotated text type. Unlike newspaper articles, where opinions are presented as a superstructure over informative value, and personal likes and dislikes are restricted, CSFD.cz reviews were written with the primary intention to express subjective opinions, likes, dislikes and evaluation.

However, the fact that we can train a good-performance classifier (see Section 9.2) indicates the usability of the annotation.

9

Sentiment Analysis

This chapter is dedicated to a practical application of emotional meaning research, namely *sentiment analysis*. The goal of sentiment analysis is to automatically extract subjective information from text and to determine the attitude of the speaker. This is applicable e.g. in opinion polls, business intelligence and customer insight or generally in market trends prediction or prediction of consumer behaviours and intentions.

Sentiment analysis is also referred to as *opinion mining* and the terms are often used interchangeably, see e.g. Liu (2012). Historically, the term opinion mining first appeared in the community associated with Web search, whereas the term sentiment analysis was used by researchers focused on natural language processing (see Pang and Lee, 2008). While the issue of subjective texts recognition has been discussed by linguists since early 80's and 90's, a substantial progress in the area has started with the rise of the semantically defined Web 2.0, which is based on user-generated content, e.g. social networks and weblogs (see Ruppenhofer et al., 2008). For more details on related work, see Section 9.1.

There are two basic types of text classification in sentiment analysis: *subjectivity detection* and *polarity detection*. In subjectivity detection the task is to determine whether a given text represents an opinion or a fact – or more precisely whether given information is factual or nonfactual. Since this study is dedicated to emotional texts and thus we use only evaluative training data, we are not concerned with subjectivity detection task.

The aim of polarity detection is to define whether the opinion expressed in a text is positive or negative. As well as subjectivity, polarity can also be detected at various levels, for example:

- the expression level, i.e. individual words or phrases
- the sentence level, i.e. individual sentences or other shorts segments like newspaper headlines etc.
- the document level, i.e. whole articles, reviews etc.

Polarity of the higher span is usually derived from polarity at the lower levels, either by using simple methods as the majority vote or by employing different probabilistic classifiers. Mostly, polarity is indicated by *polar elements*, i.e. words and expressions containing positive or negative polarity (e.g. *nice*, *awful* etc.), or by syntactic or morphological devices, as described earlier in this study. Polarity items are subject to influences of sentence or larger text span context (e.g. negation or changes in aspect in both Czech and English) and thus can be profitably explored in a syntactic treebank (as described in Chapter 7).

An important subtask in polarity detection is an *opinion source/target identification*, i.e. the task in which the opinion holders/evaluated entities need to be identified in evaluative texts (see e.g. Kim and Hovy, 2006).

9.1 Related Work

In the present study, we mostly refer to the relevant papers in the given chapters, so this section just sums up a basic sentiment analysis literature. The overview of traditional approaches to sentiment analysis and evolution of the area are described in detail in Wiebe et al. (2004), Pang and Lee (2008) and also in Bing Liu's monography (Liu, 2012). A nice summary of the field is also given by Westerski (2007) and Cambria et al. (2015).

Early papers on sentiment analysis by Turney (2002) and Pang et al. (2002) deal with polarity classification on product reviews and movie reviews. The issue of subjective text annotation is widely described e.g. in Wiebe et al. (2004). The domain dependent nature of sentiment analysis tasks is explored by Lee et al. (2009). The detailed description of subjectivity lexicon generation is given e.g. by Banea et al. (2008a) (for more details, see Section 2.2.1). Papers concerning opinion target identification are described in Section 9.3.

Regarding sentiment analysis in Czech, this work mostly draws on Veselovská (2012), Veselovská et al. (2012), Veselovská et al. (2014), Šindlerová and Veselovská (2013), Veselovská and Hajič Jr. (2013), Veselovská and Tamchyna (2014) and Tamchyna and Veselovská (2016).

Other work studying sentiment analysis in Czech mostly focuses on building polarity classifiers based on supervised machine learning techniques. Červenec (2011) uses support vector machines (SVM) for polarity classification. Habernal et al. (2013) also employ SVM along with a Maximum Entropy classifier. Moreover, they compare the influence of different feature settings when performing the classification on a large manually annotated social media dataset created for the purposes of their experiments. Steinberger et al. (2011) uses a multilingual parallel news corpus annotated with opinions towards particular entities, projecting sentiment annotation from one language to the others. The corpus contains Czech as one of the seven languages it consists of. Burget et al. (2010) also deal with Czech news data, but they primarily focus on Czech news headlines. Ptáček et al. (2014) use a cross-linguistic comparison when applying Twitter sentiment analysis to Czech and English data. The authors also present a manually annotated corpus of Czech tweets. They use both SVM and MaxEnt to perform a pilot study for sarcasm detection in both languages. In Habernal and Brychcín (2013), the authors use semi-supervised methods in document-level sentiment analysis, adding the word cluster features into standard supervised classification. As already mentioned earlier in this work, Lenc and Hercig (2016) build neural network for sentiment analysis in Czech with rather satisfactory results. Concerning the document level, Žižka and Dařena (2010) build a system for identifying Czech and English opinionated documents based on the similar patterns they share.

Many studies deal with the topic of gathering subjective expressions. In Steinberger et al. (2012), the authors suggest a semiautomatic 'triangulation' approach for creating sentiment dictionaries for many languages, again including Czech. Smrž (2006a) uses parallel corpora to create algorithms for collecting patterns that can extract subjectivity clues from the texts. The same author also experiments with using WordNet extension for the same purpose to be able to automatically identify opinionated texts (see Smrž, 2006b).

Some authors also discuss the opinion target and its particular aspects. Brychcín and Habernal (2013) apply document-level sentiment analysis incorporating global context of the sentiment target. Employing unsupervised methods, they improve the standard classifiers performance. Steinberger et al. (2014) focus on aspect-based sentiment analysis in Czech. They identify targets and the given aspects in restaurant reviews using supervised machine learning methods on manually annotated reviews corpus.

9.2 Sentence Level Polarity Detection in Czech

In this section, elaborating on Veselovská (2012), Veselovská et al. (2012) and Veselovská et al. (2014), we explore polarity at the sentence level, using both polarity expressions derived from expression level and probabilistic models trained on the manually annotated polar sentences. Apart from the document level where we have to deal with many subjectivity clues and the overall polarity tends to be less evident, the polarity at the sentence level is easier to get identified unambiguously. According to Wiegand and Klakow (2009), at the document level, text classification relies very much on redundancy and there are so many cues suggesting positive polarity more likely than negative polarity. Additionally, polarity is usually not uniformly distributed across a document, so the frequency analysis further used e.g. in text summarization is not enough without knowledge of influence of particular polar expressions at the sentence level and the position of a given sentence in the document. We are aware of the fact that even at the document level we should still take into account features from the sentence level. As mentioned earlier, we can always derive the overall polarity of the document based the polarity of particular sentences.

The sentence level enables us to explore many useful linguistic features in the analysis, such as part-of-speech information or features derived directly from the sentence structure such as clause types, as described e.g. in Wiegand and Klakow (2009). Every word needs to be interpreted in its sentential context (e.g. in English one needs to determine whether *like* is a verb and hence a positive polar expression or a preposition. In Czech, we need to distinguish between particular senses of semantically ambiguous adjectives like *hrubá* – 'rough/huge/coarse' v.s. *hrubá* – 'rude' etc., see

Chapter 3). Also, sentence-level polarity detection is useful for opinion target identification, which is at the first phase also carried out at the sentence level (see Section 9.3).

The main task in polarity classification is the detection of the items indicating positive and negative polarity, since all sentences to be classified are assumed to be subjective and carrying either positive or negative evaluation. For this purpose, we can either use the already existing domain-independent subjectivity lexicons, possibly translated from other languages, or we can find the set of subjectivity clues which work for a given domain. To determine these expressions, mostly supervised learning techniques are used, although there are also some unsupervised methods (see e.g. Turney, 2002).

In polarity classification, there are three main supervised classifiers usually employed: Naive Bayes classifier, maximum-entropy-based classifiers and Support Vector Machines. As indicated by Pang et al. (2002), for polarity detection, all of the classifiers give more or less the same performance. Therefore, in our experiments we simply chose one of these classifiers that was convenient to use, namely Naive Bayes, and focused on other aspects of the task, such as feature selection. Moreover, we built two other classifiers for comparison.

All in all, we use three classifiers in our experiments:

- Naive Bayes classifier (NB), using lemmas as features
- Lexicon-based classifier (LB), using lemmas as features employing different filters
- SubLex-based classifier (SB), using clues from Czech Subjectivity Lexicon as features

Moreover, we combine the SubLex classifier with probabilistic model based on Maximum Entropy, as described below.

There are several steps leading to an effective polarity classifier. In case of NB and LB classifiers, during the pre-processing phase, all the data first undergo lemmatization, using the tagger of Hajič et al. (2007). From the tagger output, not only do we retain the lemma but also the part of speech and negation morphological tags. Then, we automatically generate a polarity lexicon from the training data and compute the measurement of how reliable a given lexicon item works as a polarity indicator. From our data, we first need to estimate the probability that a given lemma, encountered in the data, is a part of a polar segment. Assuming we have that probability for each lemma we encounter in a given segment, we can easily decide whether to classify the given segment as polar by means of some aggregation, for instance a simple sum. Then we can analogously determine its orientation. The desired properties of an indicative strength function are satisfied by lemma precision (see Wiebe et al., 2004). Then we need to compute a baseline for our lexicon, i.e. the probability that a randomly chosen word implicates the given polarity. For details, see Appendix C.

9.2.1 Naive Bayes Classifier

In this section, we adopt Naive Bayes classifier with Laplacian smoothing (see Appendix C), using bag of lemmas as the features. In English, one can use word-forms directly as features. However, since Czech is an inflectionally rich language, this would lead to data sparsity problems. Therefore, we reduce the dimensionality of the feature space by considering lemmas instead of word-forms. In addition, we annotate each lemma with its part of speech. This does not influence the feature space in any way, but allows us to further filter the lemmas, e.g. to ignore all numbers etc. However, we need to be careful with filtering, so that we would not disregard e.g. the negation, as described in Section 9.2.2 in more detail.

In the very first classification scenario, we attempt to classify individual units of annotations – segments. By segments we consider sentences and other short segments like headlines etc. In this study, we use the terms *sentence* and *segment* interchangeably. The aim of the experiments is not really to build state–of–the–art sentiment analysis application, but rather to evaluate whether the data coming from the annotations are actually useful, where are their limits and how to possibly change the annotation guidelines to provide higher-quality data. For this purpose, we built the classifier the way described in Appendix C.

The experiments are carried out across the three datasets described in Chapter 8: the small richly annotated Aktualne.cz and CSFD.cz datasets and the larger Mall.cz data (which are not annotated below segment level). Annotation guidelines can be found in Appendix B.

9.2.1.1 Classifier performance

The baseline for all classifier experiments assigns the most frequent class to all segments. For all classifier experiments, we report f-score and improvement over baseline. The reported f-score (see Table 9.1) is computed as an average over f-scores of individual classes weighed by their frequencies in the true data.

20-fold cross-validation (see Appendix C) was performed, with the train/test split close to 4:1. The split was done randomly, i.e. a segment had a 0.2 chance of being put into test data. No heldout data were necessary as the smoothing parameter α was set manually to 0.005. Changing it does not significantly alter results.

On Aktualne.cz data, the classifier is not able to perform better than the baseline. This may be caused by a large proportion of the neutral segments in the data. Therefore, we repeat the experiment only with the first 100 neutral segments.

The CSFD.cz data consist of reviews which were assigned a score from -1 to 5 by their authors. In order to see whether the simple unigram model can be scaled up, we tried – using the model trained on manually annotated segments – to predict this document-level classification. Scores of -1 to 2 were considered negative, 3 neutral and 4-5 positive (the score-class mappings were chosen manually to adequately represent the reviewers' opinions, there were: 20 negative, 11 neutral and 32 positive

	Mall.cz		Aktualne.cz		Aktualne.cz 100N		CSFD.cz 1	
	Acc	F	Acc	F	Acc	F	Acc	F
Baseline	0.630	0.286	0.787	0.694	0.304	0.142	0.341	0.173
Naive Bayes	0.827	0.781	0.787	0.694	0.778	0.531	0.766	0.754

Table 9.1: Performance of the Naive Bayes classifer

segments in our data). The classifier considered the whole review as one segment. However, generalizing sentence-level annotation to document-level annotation performed badly, probably due to the ambiguous meaning of the documents and the fact that they contained many opinion targets.

Possibly the most important finding of the classifier experiments is that the very simple Naive Bayes polarity classifier can be trained with decent performance (at least on the film review data) with only a very modest amount of annotated data (see Tables in Chapter 8). The time to annotate the dataset did not exceed five to six hours per annotator, although both annotators reported their confidence deteriorated after roughly an hour of uninterrupted work.

9.2.2 Lexicon-based Classifier

After finishing the initial phase of our research during which we built a standard unigram-based Naive Bayes classifier, we also built a lexicon-based classifier on the same data for comparison, as described in Veselovská and Hajič Jr. (2013), which serves as a basis for this section. Similarly as in the NB classifier, we consider negation-marked lemmas. The classifier uses a standard unigram bag-of-words model, simply summing the indicator strength measurements over all lemmas in a given segment. Then it selects the polarity class with the highest accumulated value in the relevant measure. Apart from NB classifier, we also employ a number of simple filters and other methods in order to improve the automatic annotation: filtering by frequency, weighed filtering by frequency (where the threshold for accepting a lemma as a feature is weighed by the baselines so that smaller polarity classes do not get discriminated), statistical significance filtering (where we accept a lemma if we can exclude the hypothesis that it is evenly distributed across polarity classes at a given level – 0.999, 0.95 and 0.8) or filtering by part of speech. Also, we deal with sentence-level negation: first, if a segment contains a negative verb, the values for positive and negative polarity would be reversed for the segment, and a less crude method where we would specify which parts of speech to the right of a negative verb we would like to reverse. The results for particular filtering methods can be found in Hajič Jr. (2011).

As shown in Table 9.2, the lexicon-based classifier consistently outperforms the Naive Bayes classifier. Moreover, on the less complicated data, it performs comparably to state-of-the-art, see Cui et al. (2006). We assume this is due to the linguistic filtering taken into account. Acc, R, P and F stand for accuracy, recall, precision and f-score, respectively.

Model	Acc	Negative			Positive			Evaluative		
		R	P	F	R	P	F	R	P	F
baseline	0.630	0	0	0	1	0.630	0.773	0.370	0.233	0.286
LB, train	0.960	0.964	0.935	0.949	0.958	0.977	0.967	0.960	0.961	0.960
LB, test	0.889	0.907	0.821	0.862	0.878	0.939	0.908	0.889	0.894	0.890
NB, train	0.864	0.717	0.901	0.798	0.955	0.849	0.899	0.803	0.879	0.833
NB, test	0.827	0.630	0.872	0.730	0.947	0.811	0.874	0.745	0.847	0.781

Table 9.2: Baseline, comparing performance on training and test data

9.2.2.1 Classifier Error Analysis

Having trained the classifier, we took a closer look into the list of incorrectly detected instances. We found a number of functional words assigned with a wrong polarity. Unfortunately, the first-aid filtering methods have proven rather useless – even those which appeared promising. When we remove such items from the classification, the overall results do not improve. Moreover, when we start to eliminate content words, the results get even worse. In order to reveal the main cause of the mistakes, we have to go back into the data once again.

We found two main sources of errors: errors caused by human annotators and system errors. Typical errors caused by human annotators are described in Section 6.4. Below we describe typical system errors.

We discovered various reasons of the system errors which can be divided into following categories. Statistically, the significant source of errors are still the short segments like *Nic.* – 'Nothing.', *Cena* – 'Price.' or *Nevím.* – 'I don't know.' which appear in both positive and negative reviews. These can by classified by the simple majority vote. If the vote is equal, the lemma classification is based on the baseline.

Also, some of these short segments have pretty high indicative strength for one polarity, but they often appear in the reviews expressing opposite evaluation (so filtering by frequency does not help):

(98) Negative category:
Kvalita.
'Quality.'

(Mall.cz)

In these cases the system always assigns the incorrect value. The solution to these problems could be elimination of all one-word answers or assigning the polarity of these items according to the polarity they have in subjectivity lexicon for Czech (see 8.1).

One of the most frequent wrongly detected short phrases was *vysoká cena* – 'high price' tagged by the classifier with a positive instead of negative value. Besides, the classifier sometimes could not detect the domain-dependent evaluation, like *dlouhé prací programy* – 'long washing programs'. These cases could be solved by using n-grams instead of just unigrams. Using n-grams could also hold for incorrectly detected evaluative idioms like *je to sázka na jistotu* – 'it is a safe bet' etc., which are not listed in the Czech Subjectivity Lexicon or which are domain-dependent.

Furthermore, it could be advantageous to apply a coefficient for the initial and terminal position of words in a given segment. According to the reviews, it seems that the words occurring at the beginning or in the final parts of the text are more predictive towards the overall polarity:

(99) Positive category:

Je to výkonný a kvalitní vysavač, vím to, protože jsem ho měla víc jak deset let, ale bohužel se častým používáním porouchal a nechtěla jsem ho nechat opravovat, tak jsem si koupila nový. Ten starý vysavač funguje pořád jako vysavač, nejdou s ním čistit koberce. Půjčovala a půjčuje si ho celá rodina i příbuzný, je fakt dobrý, mohu ho doporučit.

'It is a high-performance and quality vacuum cleaner, I am sure, because I had it for more than ten years, but unfortunately it got destroyed by the frequent use and I did not want to have it fixed, so I bought a new one. I still use the old one, but it is not possible to clean the carpets with it. The whole family borrows it constantly, it is really good and I can only recommend it.'

(Mall.cz)

Moreover, the system is at the moment not able to treat emoticons: it considers every part of the smiley to be a separate word. Identifying positive and negative emoticons could help to detect given sentiment much better, as outlined in Read (2005).

There are also errors that can be corrected using some simple linguistic features. For sentential negation, we use the rule roughly saying that all the negated verbs switch the overall polarity of the given sentence. But there are still plenty of additional rules which could be implemented. Mostly, this concerns syntactic features. We found many incorrectly detected adversative constructions like:

(100) Positive category:
Není to žádný luxusní model, ale na chalupu stačí.
'It is not a luxurious model, but for the cottage it will do.'

(Mall.cz)

The *but-sentences* can be as well solved by the rule concerning *Semantic Consistency Principle*, as described in detail in Section 5.2.1. Also, there were many incorrectly evaluated concessive or conditional sentences in the data:

(101) Positive category: Přestože neplní hlavní funkci kvůli které jsem ho kupoval (uklidit jednu místnost po druhé během naší nepřítomnosti), tak se jedná o jednoho z nejlepších robotů v nabídce na našem trhu.
'Although it is not suitable for the function I bought it for (to clean the rooms one by one when we are not at home), it is still one of the best available robots.'

(Mall.cz)

These problems might be eliminated by compiling a stop-words list of items signalling non-evaluative part of the sentence (like e.g. *pokud* – 'unless'). Another stop-words list could contain words which can be misleading in terms of classification, since they often appear in segments of either of polarities (e.g. *nic* – 'nothing' or *cena* – 'price').

9.2.3 SubLex-based Classifier

In the following phase, we employ the Czech Subjectivity Lexicon (see Section 8.1) into a polarity classification. The experiment is described in Veselovská et al. (2014), a paper this section is based on.

First, we measure the raw performance, based exclusively on lexicon features. Second, we combine the lexicon with a probabilistic model. However, before doing so, there are several questions we have to ask about the lexicon quality. First, we need to check the coverage of the lexicon and find out whether lexicon entries appear in the data at all. Also, it is useful to know how often does a lexicon entry occur in the data and how many distinct lexicon entries appear in the data. This gives us a very loose upper bound on lexicon density in the given data: even if every negative/positive hit comes from a text span of the given orientation, the proportion of lexicon items in the evaluative text would be the number of hits divided by the size of the data with the given orientation. Table 9.3 summarizes how many times a lexicon word occurred in various datasets. We refer to the occurrence of a lexicon entry in the data as a lexicon hit. *Neg. words* is the total word count over all items tagged as negative in the dataset, *neg. hits* is the total count of words in the data that were found in the lexicon with the negative orientation (negative hits) and *dist. neg. hits* is the amount of distinct negative lexicon entries found in the dataset. Analogously for positive items and lexicon entries. Besides the three datasets mentioned in previous sections, we also employ a Facebook dataset (for details, see Chapter 8).

Dataset	Words		Hits			
	Neg	Pos	Neg	Dist. neg.	Pos.	Dist. pos.
Aktualne	1,003	358	119	53	102	59
CSFD	4,739	6,231	254	68	301	65
Mall	60,652	98,303	1,676	154	4,174	146
Facebook	33,091	30,361	1,166	186	2,661	182

Table 9.3: Lexicon coverage

Since many lexicon hits are not in the text span of the corresponding polarity, we need to proceed to testing how good the lexicon is as a predictor. For this purpose, we use a series of primitive, "raw" binary classifiers. Note that these classifiers are just helper constructs for measuring the relationship between lexicon hits and data item (expression) orientations.

We define lexicon features: the counts of positive and the count of negative items from the lexicon in the text span. We call the features POS and NEG. If a lexicon item permits both polarities, it contributes both to POS and NEG counts. If the text span did not include any lexicon item, it was given a technical neutral (NTR) feature with count 1.

We then derive lexicon indicator variables from the lexicon features: if a lexicon feature is greater or equal to some threshold frequency (denoted $threshold_{\mathrm{LI}}$, by default 1) for a data item, the indicator variable value for the given data item is 1; otherwise it is 0. We denote these features as LI_{POS}, LI_{NEG} and LI_{NTR} (*LI* stands for Lexicon Indicator).

The raw negative classifier then labels all items with negative hits – those with a LI_{NEG} value of 1 – as negative and all the others as non-negative. These binary "predictions" are then evaluated against the binarized "true classes" – all negative data items receive a 1, all non-negative a 0. Analogously for positive items. Note that under this scheme, one data item may receive a 1 for multiple lexicon indicator features – if it contains both a negative and a positive lexicon hit. This would be a concern if we were building a classifier for all classes at once. However, it only has one true orientation, so it can only contribute once to a correct classification.

The raw neutral classifier labels as neutral items without more than $threshold_{\mathrm{LI}}$ lexicon hits. The "both" class is not predicted. For each raw classifier on each dataset, we report its precision, recall and support (the true number of data items with the given polarity label) for the label of interest (NEG for the raw negative classifiers, etc.). Recall is the ratio of text spans of the given polarity "found" by the lexicon to the total amount of data items labelled with this polarity, precision is the proportion of correctly identified data items in the set. A recall of 0.5 for the label NEG and negative polarity data items means that in half of the negative data items, a negative lexicon entry appeared. A precision 0.5 means that half the data items in which a negative lexicon entry appeared are actually items labelled as negative in the data.

Given that we build a separate raw classifier for each class, the baseline performance is also computed for each class separately. The baseline classification assigns a 1 to the *LI* feature for each data item. This simulates the situation of a lexicon which tags at least one word in every item with the given orientation. Baseline recall is thus 1.0 and so recall ceases to be of interest. Our focus is precision, which tells us how well the lexicon hits are able to signal that an item actually has the orientation they indicate. At the same time, we watch recall to see a more detailed overview of lexicon coverage.

Recall and precision the raw classifiers achieved is captured in Table 9.4.

Dataset	Target label	Recall	Precision	Baseline p.	Support
Aktualne	POS	0.294	0.054	0.040	17
	NEG	0.324	0.230	0.166	71
	NTR	0.598	0.792	0.792	338
CSFD	POS	0.454	0.451	0.345	183
	NEG	0.377	0.333	0.284	151
	NTR	0.579	0.467	0.371	197
Mall	POS	0.354	0.744	0.639	6,500
	NEG	0.204	0.551	0.361	3,677
	NTR	0.000	0.000	0.000	0
Facebook	POS	0.278	0.320	0.259	2,587
	NEG	0.162	0.298	0.199	1,991
	NTR	0.741	0.554	0.517	5,174

Table 9.4: Lexicon feature raw performance

The most important finding from Table 9.4 is that raw classifier precision tends to follow the baseline for the given label (i.e. the proportion of text spans of that class in the data). This means that the presence or absence of lexicon words itself gives us no additional information. If a lexicon words were present in every data item, we would have the same precision.

Setting $threshold_{\mathrm{LI}}$ to 2 very predictably slightly improves precision (at most on the order of 0.1) while drastically reducing recall (to between 0.03 and 0.1). Setting the threshold to 3 suggests that no neutral item contained 3 or more lexicon hits and very few non-neutral items does. While precision can be improved by using more sophisticated classification methods, recall is more limiting – if only 65% of positive items contain a positive lexicon item, unless we are able to generalize from the lexicon to unseen words, we simply cannot improve recall over 0.65 unless we expand the lexicon.

Again, note that feature performance as measured above is not the performance of 'real' classifiers using the lexicon features. The raw classifiers are among the most unsophisticated classification methods based on the lexicon. However, they set a lower bound on what should definitely be achievable with the lexicon, based on how lexicon words occur in or outside items with corresponding orientations.

9.2.3.1 Classification against Annotated Evaluative Expressions

Since the Aktualne.cz and CSFD.cz datasets are annotated at the expression level including explicitly tagged polar expressions (i.e. parts of data items that make the annotator believe the item contains an evaluation), we can measure how much the lexicon hits correlate with these expressions. In this polar expression data, there are naturally only positive and negative data items, since only in them the polar expressions were annotated. Again, we measure precision, which in this case is the proportion of hits that occur inside polar expressions to the total amount of hits, and recall, which is the proportion of polar expressions with lexicon hits to the number of all polar expressions. The results are reported in Table 9.5. In this case, support is the number of polar expressions annotated with the given orientation by the given annotator. Since the polar expressions were tagged by two annotators with both significant overlap and significant differences, we report precision and recall for annotators separately (annotator 1/annotator 2).

Dataset	Orientation	Recall	Precision	Support
Aktualne	POS	0.15/0.24	0.50/0.67	13/17
	NEG	0.26/0.26	1.00/0.94	58/66
CSFD	POS	0.09/0.14	0.72/0.87	194/143
	NEG	0.09/0.10	0.78/0.82	152/138

Table 9.5: Precision and Recall against annotated polar expressions

While recall is still low, if the lexicon identifies something, it does tend to lie in expressions of the corresponding orientation. This again suggests that a disambiguation stage is in order. Once we know the lexicon hit lies in an evaluative statement, the hit orientation can be relied upon.

9.2.3.2 Evaluation within Classification Experiments

A further way of testing the lexicon is using lexicon features directly in a classification task, comparing them to automatically extracted features (word and n-gram counts) and evaluating also the combination of automatic and lexicon features. Contrary to the precision/recall scores reported above, the results reported here are for 'real' classifiers that classify items by orientation, so that the NEG, NTR, POS and BOTH labels are generated at once.

Class	Lex features	Recall	Precision	F-score	Support
NEG	+	0.12	0.50	0.20	71
	−	0.01	0.20	0.03	
NTR	+	0.94	0.82	0.87	338
	−	1.00	0.79	0.88	
POS	+	0.47	1.00	0.62	17
	−	0.00	0.00	0.00	
BOTH	+	0.00	0.00	0.00	2
	−	0.00	0.00	0.00	

Table 9.6: Aktualne.cz dataset. Classification with/without lexicon features and using only lexicon features

Class	Lex features	Recall	Precision	F-score	Support
NEG	+	0.60	0.71	0.60	151
	−	0.32	0.54	0.40	
NTR	+	0.88	0.68	0.76	197
	−	0.75	0.57	0.65	
POS	+	0.53	0.71	0.60	183
	−	0.64	0.63	0.63	

Table 9.7: CSFD.cz dataset. Classification with/without lexicon features and using only lexicon features

Automatic features used in classification are simply word counts. The value of feature f in a text span represents how many times the lemma corresponding to feature f was present. All classification experiments report 5-fold cross-validation averages. This time, we use a Maximum Entropy classifier. The Maximum Entropy classifier (Berger et al., 1996) is a discriminative classifier. It is called maximum en-

Class	Lex features	Recall	Precision	F-score	Support
NEG	+	0.94	0.94	0.94	3,677
	−	0.40	0.73	0.52	
POS	+	0.89	0.89	0.89	6,500
	−	0.91	0.73	0.81	

Table 9.8: Reviews dataset, classification with/without lexicon features and using only lexicon features

Class	Lex features	Recall	Precision	F-score	Support
NEG	+	0.43	0.61	0.51	1,991
	−	0.06	0.46	0.10	
NTR	+	0.85	0.71	0.77	5,174
	−	0.88	0.56	0.68	
POS	+	0.70	0.77	0.73	2,587
	−	0.30	0.48	0.37	
BOTH	+	0.05	0.36	0.08	248
	−	0.00	0.00	0.00	

Table 9.9: Facebook dataset, classification with/without lexicon features and using only lexicon features

tropy, because it makes as few assumptions as possible: from all the models that fit the training data, it selects the one which has the largest entropy. In other words, it prefers distributions that are as uniform as possible. These minimal assumptions also mean the classifier is fairly robust and it can be used when we have little or no knowledge of the prior distribution. Unlike Naive Bayes, a MaxEnt classifier does not assume that the features are conditionally independent of each other. The classifier is implemented as Logistic Regression in the scikit-learn Python library (available at `http://scikit-learn.org/stable`). The regularization parameter was set to 1.0 with the exception of the Aktualne.cz dataset, where setting it to values of several thousand significantly improves the performance on the positive text spans.

We report results for the individual classes. It is more informative, especially for datasets with large imbalances of classes, than to report the averaged performance. Since the classifier performance was never significantly changed by including the lexicon features, the results are reported for classification with automatic and combined lexicon/automatic features in the same table. Table 9.6 demonstrates the results on the Aktualne.cz dataset. Note that given the small size and heavily imbalanced nature of the dataset, the results for the negative and positive classes are very unstable. The positives f-score varying by as much as 0.2 in consecutive cross-validation runs.

Table 9.7 shows the CSFD.cz dataset (while small, the dataset proved much more stable, varying within 0.05 in consecutive runs). Note that using only the lexicon features improves recall on positive items. In Tables 9.8 and 9.9 we present the results for the Mall dataset and Facebook dataset.

9.2.3.3 Identifying Problematic Lexicon Entries

By looking at the lexicon entries which appear in items of opposite or neutral polarity, we can try to detect problematic patterns – those left over from the translation phase that have slipped through the refining process, or problems connected to the usage of lexicon entries in Czech. We report the top ten words for each problem category, the English lexicon entries they were translated from, their frequencies in the opposite data and in their home data and notes on the prevailing nature of the error after manually inspecting error sites. Tables 9.10 and 9.11 show problems with orientations, whereas Tables 9.12 and 9.13 illustrate problems with detecting evaluations v.s. neutrality.

We see that the most frequent causes of misclassification are domain mismatches, where a word that is a priori – or in the source domain – oriented one way is oriented differently (manipulation, comedy) in another domain. Other frequent problems arise from translation by the CzEng parallel corpus (for details, see Section 8.1). Either a "lost in translation" phenomenon, where what is an originally subjective and evaluative word becomes a more or less neutral word, or a word that is evaluative only weakly or in a very specific context (and thus escaped manual cleansing), or a straight mistranslation. The statistical machine translation system can also translate rare words as more frequent ones due to the target-side language model. Some other problems suggested by our inspection are the use of words frequently negated in a domain (*error* in *hasn't got a single error*), words that are translated as colloquial phrases with only one part of the phrase included in the lexicon, and the occasional use of frequent and strong evaluative words ironically (*super*).

We used the same approach to see which negative and positive words most often appear in neutral segments (Tables 9.12 and 9.13). Aside from legitimate language use reasons (regular non-evaluative usage), the discovery of which is again a task for disambiguating whether an entry is used as an evaluative word, the most frequent problems stemmed from translation.

9.2.3.4 Automated Lexicon Pruning

The number of incorrect hits caused by negative polarity indicators found in positive segments (and vice versa) decreases roughly exponentially. It means that only a few expressions can have a significant impact on accuracy of the whole system. Therefore

Negative hits, positive data	**Original English entry**	**Frequency pos.**	**Frequency neg.**	**Note**
manipulace – 'manipulation'	manipulation, tamper	178	27	domain-specific (household apps.)
chyba – 'error'	error, mistake, flaw	65	56	negation mismatch ('no flaw at all')
nastavit – 'set'	plot	32	35	mistranslated
vypnout – 'turn off'	disable	24	41	mistranslated/lost in translation
manipulovat – 'manipulate'	manipulate, manipulation	18	3	analogous to *manipulace*
komedie – 'comedy'	comedy, farce	18	1	domain mismatch (film reviews)
hluk – 'noise'	din, clamor	17	28	domain+negation mismatch ('little noise')
odpad – 'waste'	waste, drain	13	20	domain mismatch (household apps.)
zkusit – 'try'	try	9	12	homonymy: 'try the car' v.s. 'a trying test'
skvrna – 'stain'	stain, blemish	9	7	domain+neg. mismatch (household apps.)

Table 9.10: Negative entries occurring most often in positive segments

Positive hits, negative data	**Original English Entry**	**Frequency neg.**	**Frequency pos.**	**Note**
dost – 'enough'	pretty, plenty	135	58	lost in trans.: positive to neutral intensifier
smlouva – 'agreement, contract'	agreement, covenant	30	1	domain mismatch (phone operator problems)
informace – 'information'	intelligence	28	28	mistranslated
cena-2 – 'price'	worth	24	12	lemmatization disambiguation error
dodat – 'provide'	embolden	22	16	split phraseme: embolden=provide+courage
lehce – 'easily'	easily	20	56	lost in trans.: positive to neutral modifier
vypadat – 'look'	minister	19	35	mistranslation: rare Eng. to common Cz.
energie – 'energy'	energize	19	158	lost in trans. + mistrans.: wrong POS
super – 'super'	super	17	127	irony/sarcasm + adversative constructions
snadno – 'easily'	easily, ease, attractively	16	69	analogous to *lehce*

Table 9.11: Positive entries occurring most often in negative segments

Negative hits, neutral data	Original English Entry	Frequency ntr.	neg.	Note
zkusit – ‘try’	try, difficult	48	12	homonymy: *try the car* v.s. *a trying test*
chyba – ‘error’	error, mistake, failure, flaw...	46	56	regular non-evaluative usage
situace – ‘situation’	crisis, predicament, plight...	17	7	lost in translation
nastavit – ‘set’	plot	17	2	mistranslated
chybit – ‘miss’	miss	16	2	regular non-evaluative usage
ztratit – ‘lose’	lose, vanish, doom, dishearten	12	1	regular non-evaluative usage
smrt – ‘death’	death, martyrdom, dying	11	2	regular non-evaluative usage
zmizet – ‘disappear’	vanish, abscond, swagger	9	5	lost in translation
vypnout – ‘turn off’	disable	9	41	lost in translation
sranda – ‘fun’	fun, goof	9	7	orientation error in lexicon refinement

Table 9.12: Negative entries occurring most often in neutral segments

Positive hits, neutral data	Original English Entry	Frequency ntr.	pos.	Note
cena – ‘price’	worth	40	12	lemmatization disambiguation error
doufat – ‘hope’	hope, hopefully, hopefulness	36	32	lost in translation: neutral colloquial usage
vypadat – ‘look’	minister	30	35	mistranslation: rare Eng. to common Cz.
informace – ‘information’	intelligence	29	28	mistranslation: rare Eng. to common Cz.
dost – ‘enough’	pretty, plenty	28	56	lost in trans.: positive to neutral modifier
dobrý – ‘good’	good	27	42	phrase *dobrý den* (greeting)
souhlasit – ‘agree’	agree, consent, concur...	21	15	regular non-evaluative usage of ‘agree’
smlouva – ‘agreement’	agreement, covenant	20	1	domain mismatch (cell phone operators)
radost – ‘joy’	joy, pleasure, delight, happiness...	15	33	non-eval. usage, misannotated items
chystat – ‘to prepare’	solace	14	6	mistranslated

Table 9.13: Positive entries ocuurring most often in neutral segments

we hypothesise that we could improve the lexicon indicator precision by pruning. To see how much we could gain by removing misleading lexicon entries, we combine half of the Facebook and Mall.cz data to find lexicon entries that impede classification. We then compute the recall and precision statistics of the lexicon indicator features and coverage statistics on the second halves of the data (see Fig. 9.1).

An entry is classified as misleading if we cannot reject the hypothesis that its occurrences are evenly distributed across items of its class v.s. items of all other classes combined, or if we can reject this hypothesis and it occurs less frequently in items of its class than in other items. We use the binomial exact test since lexicon hits are often low-frequency words and we thus cannot accurately use the chi-square test.

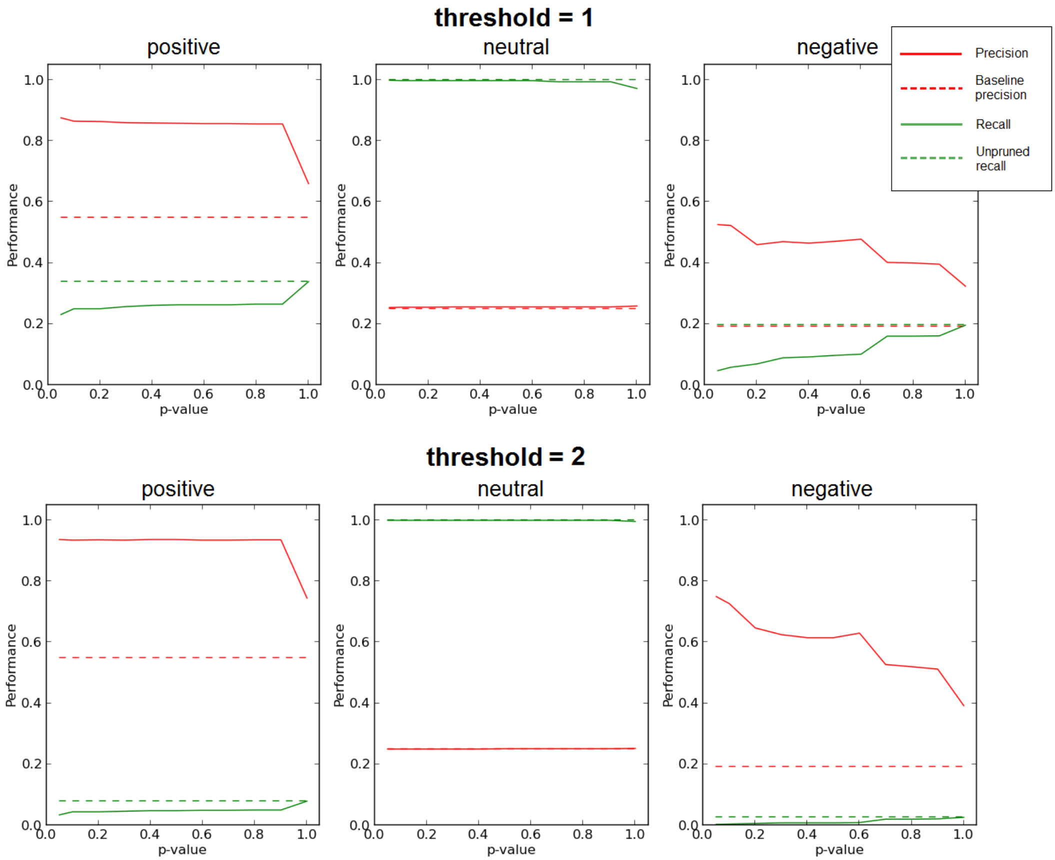
We try pruning at various levels of the test, to find a good tradeoff between gaining precision and not losing too much recall, so that the pruning isn't too severe. The results are reported in Fig. 9.1. The rightmost data point ($p = 1.0$, $\alpha = 0.0$) is for the lexicon before pruning, so the large skip between $p = 0.9$ and 1.0 is caused by removing words which appear more frequently in items of other orientations than their own orientation. We also use both $threshold_{\mathrm{LI}} = 1$ and 2 (setting the indicator threshold to 3 is mostly useless, since very few items contain 3 lexicon hits).

The very low recall for some classes means that less than 10 items actually contains a lexicon hit of their polarity. However, after such automated pruning, the lexicon may be suitable for building a high-precision classifier such as in Riloff and Wiebe (2003). On the Aktualne.cz dataset, the pruned lexicon never achieves higher precision than the unpruned version. However, on the CSFD.cz dataset, for $p = 0.05$ and, $threshold_{\mathrm{LI}} = 2$, the precision for LI_{POS} defeats the unpruned (0.793 v.s. 0.543) with precision for the other indicators not significantly different from the unpruned lexicon scores.

From the experiment with lexicon feature recall and precision, we believe that a disambiguation stage, where the occurrence of a lexicon item is assigned some confidence that the occurrence actually is polar, could be highly beneficial, since words from the lexicon frequently appear in text spans of opposite polarities or neutral text spans.

Adding the lexicon features to sentiment classifiers does not significantly improve the results in any of the above mentioned experiments, with the exception of positive text spans in the CSFD.cz dataset. Using the lexicon features alone, which is an option in a scenario where manually annotated data is not available, might produce decent results on the datasets with preeminently evaluative user-generated content: Aktualne.cz and CSFD.cz. However, to confirm this claim it would be useful to repeat the experiments using other classifiers.

As for the general usefulness of the lexicon, it is apparent that the lexicon by itself – at least by using lexicon features in the manner described above – cannot compete with statistical methods on a representative in-domain annotated dataset such as Mall.cz, and even when the automatic features are combined with the lexicon features, classifier performance does not improve. However, the lexicon does not hurt classification either, and it remains to be seen whether it can help in classifying previously unseen



Red lines indicate precision, green lines recall;
dotted lines are baseline precision and pre-pruning recall.

Figure 9.1: Pruned SubLex Performance

domains (the Aktualne.cz and CSFD.cz datasets are not large enough for conclusive testing). Nevertheless, the prevalence of domain mismatch among frequent causes of entry/data item orientation mismatch suggests that this at least requires a more sophisticated method.

In order to improve the automatic polarity classification, it could also be useful to enhance the subjectivity lexicon by several methods. Firstly, we could use the dictionary-based approach as described by Hu and Liu (2004) and Kim and Hovy (2004) and grow the basic set of words by searching for their synonyms in Czech WordNet (Pala and Ševeček, 1999).

Secondly, we could employ the corpus-based approach based on syntactic or co-occurrence patterns as described in Hatzivassiloglou and McKeown (1997). Also, we can extend the lexicon manually by Czech evaluative idioms and other common evaluative phrases. Moreover, it would be useful to add back some special domain-dependent modules for the different areas of evaluation. To improve the lexicon itself by automatic means besides pruning by statistical significance, we can ablate the lexicon: try removing features and see how much the removal hurts (or helps) classification in various scenarios both already implemented and new.

9.3 Opinion Target Identification

In addition to the polarity classification task where we try to identify polarity of the evaluation, it is important to be able to perform automatic target identification, i.e. the task in which the evaluated entities need to be identified. This task can be divided into two subtasks: identification of the high-level opinion target (object of evaluation) and identification of target aspects, like battery in a notebook etc. The latter subtask is applied since in a typical opinionated text, the author usually evaluates both positive and negative aspects of the object, even though the overall evaluation on the object can be positive or negative. Identification of aspects is often called *aspect-based sentiment analysis* (see Pontiki et al., 2014). Since we assume that there can also be different targets in one document, we call targets all the evaluated entities, i.e. also particular aspects. To be able to identify a general domain, we assign target categories. This section describes how we automatically assigned opinion targets and target categories in various languages.

9.3.1 Target Identification in English

This part of the study is based on Veselovská and Tamchyna (2014), but the issue of opinion target identification in sentiment analysis is discussed in many articles proposing different methods, mainly tested on product review datasets (see e.g. Popescu and Etzioni, 2007; Mei et al., 2007; Scaffidi et al., 2007). Some of the authors take into consideration also product aspects, defined as product components or product attributes (Liu, 2007). Hu and Liu (2004) take as the aspect candidates all noun phrases found in the text. Stoyanov and Cardie (2008) see the problem of target identification as part of a topic modelling problem, similarly to Mei et al. (2007). In this study, we follow the work of Qiu et al. (2011) who learn syntactic relations from dependency trees. We use rule-based classification applied primarily on English data provided by SemEval2014 task (Pontiki et al., 2014). Our approach is related to polarity detection based on the subjectivity lexicons, generally described e.g. in Taboada et al. (2011). The English ones we use are minutely described in Wiebe et al. (2005) and several papers by Bing Liu, starting with Hu and Liu (2004). Inspired by Kobayashi et al. (2007), who make use of evaluative expressions when learning syntactic patterns obtained via pattern mining to extract target-evaluation pairs, we use the opinion words to detect evaluative structures in parsed data.

In opinion target identification, it is advantageous to use training data manually annotated with targets. However, in a real-life scenario, we usually do not have any golden targets at our disposal. Therefore, it is practical to be able to extract both targets and their polarities at once. In our experiments, we first parse the data, bearing in mind that it is very difficult to detect targets on plain text corpora. This holds especially for pro-drop languages like Czech, but the proposed method is still language-independent to some extent. Secondly, we detect polarity items in the parsed text using existing subjectivity lexicons. Afterwards, we extract target terms in the dependency structures containing polarity expressions. In this task, we employ several hand-crafted rules detecting targets based on syntactic features of the evaluative sentences. Finally, we identify target term categories with the help of the English WordNet and derive their polarities based on the polarities of individual targets, so that we would get the idea about evaluation assigned to particular topics.

9.3.1.1 Data

For this experiment, we use the training and trial data provided by the organizers of SemEval2014 task (Pontiki et al., 2014). The training and test data sizes are described in Table 9.14. There are two data domains, restaurants data and laptop data. The restaurants training data contains 3,041 English sentences and it is a subset of the dataset from Ganu et al. (2009). These data include annotations for target categories and overall sentence polarities. The SemEval organizers added annotations for target terms and target term polarities and also for target category polarities. Additionally, they manually annotated the test data (800 sentences) in the same manner. The laptops dataset consists of 3,845 English sentences extracted from laptop customer reviews. The target terms and their polarities were annotated by human annotators. 3,045 sentences were used for training and 800 for testing. During system development, we used the trial section as a held-out set. In the final phase, both datasets are utilized in training.

Domain	Train	Test	Total
Restaurants	3,041	800	3,841
Laptops	3,045	800	3,845
Total	6,086	1,600	7,686

Table 9.14: Sizes (sentences) of the datasets

9.3.1.2 Pipeline

We first pre-process the data, then mark all targets seen in the training data (still on plain text). The rest of the pipeline is implemented in Treex (Popel and Žabokrtský, 2010) and consists of linguistic analysis (tagging, dependency parsing), identification of evaluative words, and application of syntactic rules to find the evaluated targets. Finally, for restaurants, we also identify target categories and their polarity. Our workflow is illustrated in Fig. 9.2.

9.3.1.3 Pre-processing

The main phase of pre-processing (apart from parsing the input files and other simple tasks) is running a spell-checker. As data for this task come from real-world reviews, it contains various typos and other small errors. We therefore implemented a statistical spell-checker which works in two stages:

1. Run Aspell (`http://aspell.net/`) to detect typos and obtain suggestions for them.
2. Select the appropriate suggestions using a language model (LM).

We trained a trigram LM from the English side of CzEng 1.0 (see Bojar et al., 2012) using SRILM (see Stolcke et al., 2002). We binarized the LM and use the Lazy decoder (Heafield et al., 2013) for selecting the suggestions that best fit the current context. Our script is freely available for download at `https://redmine.ms.mff.cuni.cz/projects/staspell`.

We created a list of exceptions (domain-specific words, such as *netbook*, are unknown to Aspell's dictionary) which should not be corrected and also skip named entities in spell-checking.

9.3.1.4 Marking Known Targets

Before any linguistic processing, we mark all words (and multiword expressions) which are marked as targets in the training data. For the final phase, the list also includes targets from the provided development sets.

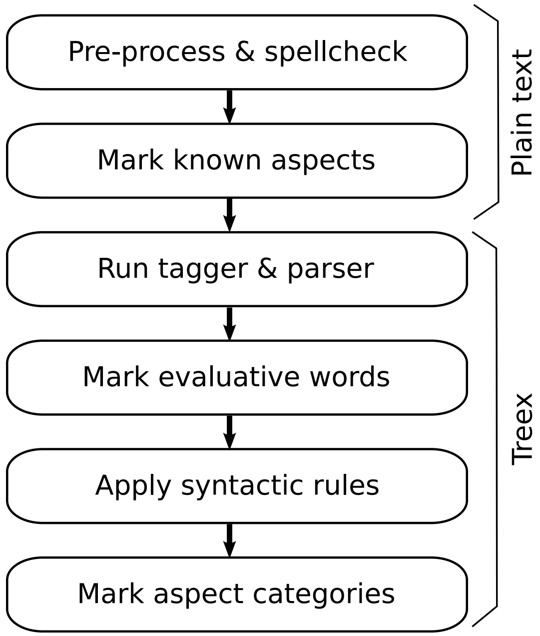


Figure 9.2: Overall schema of our approach

9.3.1.5 Morphological Analysis and Parsing

Further, we lemmatize the data and parse it using Treex (Popel and Žabokrtský, 2010), a modular framework for natural language processing (NLP). Treex is focused primarily on dependency syntax and includes blocks (wrappers) for taggers, parsers and other NLP tools. Within Treex, we used the Morče tagger (Hajič et al., 2007) and the MST dependency parser (McDonald et al., 2005).

9.3.1.6 Finding Evaluative Words

In the obtained dependency data, we detect polarity items using MPQA subjectivity lexicon (see Wiebe et al., 2005) and Bing Liu's subjectivity clues available at `http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon`. We lemmatize both lexicons and look first for matching surface forms, then for matching lemmas. (English lemmas as output by Morče are sometimes too coarse, eliminating e.g. negation – we can mostly avoid their matching by looking at surface forms first.)

Pattern			Example sentence
Subj_{target}	Pred_{copula}	PAdj	The food was great.
Subj_{target}	Pred_{copula}	PNoun	The coconut juice is the MUST!
Subj_{target}	Pred Adv_{eval}		The pizza tastes so good.
Attr_{eval}	Noun_{target}		Nice value.
Subj_{target}	Pred_{eval}		Their wine sucks.
Subj_{source}	Pred_{eval}	Obj_{target}	I liked the beer selection.

Table 9.15: Syntactic rules.

9.3.1.7 Syntactic Rules

Further, we manually created six basic rules for finding targets in sentences containing evaluative items from the lexicons. The rules were created based on the manual control of the parsed data and were conceived e.g. the following way: *if you find an adjective which is a part of a verbonominal predicate, the subject of its governing verb should be a target*, see Table 9.15. Situational functions are marked with subscript, PAdj and PNoun stand for adjectival and nominal predicative expressions.

Moreover, we applied three more rules concerning coordinations (some based on Semantic Consistency Principle explained in Section 5.2.1).

We assume that if we find a target, every member of a given coordination must be a target too.

(102) The excellent mussels, puff pastry, goat cheese and salad.

Concerning but-clauses, we expect that if there is no other target in the second part of the sentence, we assign the conflict value to the identified target.

(103) The food was pretty good, but a little flavorless.

If there are two targets identified in the but-coordination, they should be marked with opposite polarity.

(104) The place is cramped, but the food is fantastic!

9.3.1.8 Target Categories

We collect a list of targets from the training data and find all their hypernyms in WordNet (Miller and Fellbaum, 1998). We hand-craft a list of typical hypernyms for each category (such as *cooking* or *consumption* for the category *food*). Moreover, we look at the most frequent targets in the training data and add as exceptions those for which our list would fail.

	Task 1 target identification			Task 2 target polarity	Task 3 category detection			Task 4 category polarity
	P	R	F	Acc	P	R	F	Acc
UFAL	0.50	0.72	0.59	0.67	0.57	0.74	0.65	0.63
best	0.91	0.82	0.84	0.81	0.91	0.86	0.88	0.83

Table 9.16: Results of our system on the Restaurants dataset as evaluated by the task organizers.

	Task 1: target identification			Task 2: target polarity
	P	R	F	Acc
UFAL	0.39	0.66	0.49	0.57
best	0.85	0.67	0.75	0.70

Table 9.17: Results of our system on the Laptops dataset as evaluated by the task organizers.

We rely on the output of target identification for this subtask. For each aspect marked in the sentence, we look up all its hypernyms in WordNet and compare them to our list. When we find a known hypernym, we assign its category to the target. Otherwise, we put the target in the *anecdotes/miscellaneous* category. For category polarity assignment, we combine the polarities of all targets in that category in the following way:

- all positive → positive
- all negative → negative
- all neutral → neutral
- otherwise → conflict

9.3.1.9 Results and Discussion

Tables 9.16 and 9.17 summarize the results of our experiment and indicate that we tend to do better in terms of recall than precision. This effect is mainly caused by our decision to also automatically mark all targets seen in the training data.

9.3.1.10 Effect of the Spell-checker

We evaluated the performance of our system with and without the spell-checker. Overall, the impact is very small (f-score stays within 2-decimal rounding error). In some cases its corrections are useful (*convienent* → *convenient parking*), sometimes its limited vocabulary harms our system (*fettucino alfredo* → *fitting Alfred*). This issue could be mitigated by providing a custom lexicon to Aspell.

9.3.1.11 Sources of Errors

As we always extract targets that were observed in the training data, our system often marks them in non-evaluative contexts, leading to a considerable number of false positives. However, using this approach improves our f-score due to the limited recall of the syntactic rules.

The usefulness of our rules is mainly limited by the (i) sentiment lexicons and (ii) parsing errors.

(i) Since we used the lexicons directly without domain adaptation, many domain-specific terms are missed (*flavorless*, *crowded*) and some are matched incorrectly.

(ii) Parsing errors often confuse the rules and negatively impact both recall and precision. Often, they prevented the system from taking negation into account, so some of the negated polarity items were assigned incorrectly.

The "conflict" polarity value was rarely correct – all targets and their polarity values need to be correctly discovered to assign this value. However, this type of polarity is infrequent in the data, so the overall impact is small.

9.3.1.12 Target Identification in English: Conclusion

Our system can be readily deployed as a complete solution which covers the whole process from plain text to targets and target categories annotated with polarity. Considering the number of tasks covered and the fact that our system is entirely rule-based, the achieved results seem satisfactory.

I this experiment, we developed a purely rule-based system for opinion target identification which can both detect target terms (and categories) and assign polarity values to them. We have illustrated that even such a simple approach can achieve relatively good results.

9.3.2 Target Identification in Czech

For opinion target identification in Czech, we adapted the pipeline developed for English, but we also employed machine learning. The syntactic rules required only minor changes – we adjusted the morphological tagset and translated several lemmas (such as *to be* – 'být'). Instead of using the rules directly, we utilized them as additional features in a machine-learning setting. In this task, we employed linear-chain conditional random fields (CRF), a statistical modelling method which takes into account also context. To build the CRF classifier, we used the CRF++ toolkit available at `http://taku910.github.io/crfpp/`. The task was performed on the Alza.cz dataset – 1,000 positive and 1,000 negative user reviews manually tagged with opinion targets.

In order to evaluate various feature sets and hyperparameter settings, we used 5-fold cross validation. We tuned the cost parameter and the maximum number of training iterations using grid search.

We only evaluate aspect identification and leave aspect polarity detection for future work; the training data was rather noisy and polarity annotations proved unreliable.

Baseline features which only look at the surface word forms of the current word and of words within a small (linear) context window achieve an f-score of 0.492. When we add also the lemma, morphological tag and analytical function (obtained from the dependency parse) of the current and nearby words, the f-score jumps to 0.630. When we also mark words which appear in the Czech SubLex, the score further improves to 0.634. Finally, adding the output of morphosyntactic rules as features improves the f-score to 0.641. The results of the system are described in Table 9.18.

Features	**P**	**R**	**F**
baseline	0.855	0.346	0.492
lemma, POS-tag, afun	0.796	0.521	0.630
+ SubLex item	0.825	0.514	0.634
+ synt. rules	0.806	0.533	0.641

Table 9.18: Results of the system on the Alza.cz dataset

9.3.3 Target Identification Using Recurrent Neural Networks

This part of the study is based on Tamchyna and Veselovská (2016) describing methods we used when taking parts in SemEval 2016 (Pontiki et al., 2016) task focused on opinion target identification in multiple languages and domains. Namely, we participated in Subtask 1 (sentence-level opinion target identification), focusing specifically on target category detection. Data overview is provided in Table 9.3.3. The data was annnotated on opinion target, opinion target category and sentiment. Details about annotation metodology can be found in Pontiki et al. (2016).

The addition of multiple languages made language-independent approaches attractive. Therefore, we proposed to utilize neural networks which should be capable of discovering linguistic patterns in the data automatically, thereby reducing the need for language-specific tools and feature engineering.

The task concerns sentence-level opinion target identification. The aim is to identify all the opinion tuples at the sentence level, taking into account the context of the whole review. In this case, we try to detect particular target term categories (e.g. FOOD for PIZZA) and their attribute labels (e.g. QUALITY for FOOD). Both the categories and the attribute labels were assigned based on the predefined inventories.

We apply our method on several languages covering the following domains:

Language	Domain	Train Texts	Train Sent.	Train Tuples	Test Texts	Test Sent.	Test Tuples
EN	**REST**	350	2000	2507	90	676	859
EN	**LAPT**	450	2500	2909	80	808	801
AR	**HOTE**	1839	4802	10509	452	1227	2604
CH	**PHNS**	140	6330	1333	60	3191	529
CH	**CAME**	140	5784	1259	60	2256	481
DU	**REST**	300	1711	1860	100	575	613
DU	**PHNS**	200	1389	1393	70	308	396
FR	**REST**	335	1733	2530	120	696	954
RU	**REST**	302	3490	4022	103	1209	1300
ES	**REST**	627	2070	2720	286	881	1072
TU	**REST**	300	1104	1535	39	144	159
TU	**TELC**	-	3000	4082	-	310	336

Table 9.19: Datasets provided for SemEval 2016

- Arabic: hotels
- Dutch: restaurants
- English: consumer electronics and restaurants
- French: restaurants
- Russian: restaurants
- Spanish: restaurants
- Turkish: restaurants and telecom

Our submission was the best system for Russian and Turkish but did not achieve noteworthy results in other domains/languages.

9.3.3.1 Related Work

So far, most of the researcher in the field have been focused on the traditional machine learning approaches, such are different probabilistic methods (see Agarwal and Mittal, 2016), or employed deterministic methods, e.g. subjectivity lexicons (Taboada et al., 2011).

Neural networks have been used for sentiment analysis, as meantioned in Section 2.2.3. Particularly, in 2015 SemEval Twitter sentiment classification task, several submissions applied convolutional networks (Toh and Su, 2015 and Ebert et al., 2015). In our work, we use recurrent networks instead. Our motivation for this decision is that for target identification, syntactic relationships and long-distance dependencies may play a significant role and that such phenomena may be better modeled with a recurrent network. Furthermore, our recurrent network could easily be adapted to perform sequence labeling instead of sentence level classification – this would allow us to identify the exact position in the sentences where the target was mentioned.

9.3.3.2 System Description

Our system addresses sentence-level opinion target identification. Within the subtask, we focus on target category detection. For each sentence, our goal is to identify all target categories which are mentioned. Each category is composed of an entity E (e.g. FOOD or SERVICE) and its attribute A (e.g. QUALITY or PRICE). We do not decompose this definition and treat each category independently, effectively reducing the task to many binary classification subtasks (one for each E#A pair).

Each classifier in our system is a deep recurrent neural network with Long Short-Term memory cells (LSTM, Hochreiter and Schmidhuber, 1997). LSTMs have been designed to overcome the vanishing gradient problem present in standard recurrent neural networks. Their ability to remember information over many time steps should enable them to capture long-term dependencies in the data.

Our network encodes the input sentence word by word and at the end produces a binary classification decision based on the representation of the full sentence.

We represent each word (token) on the input by its pre-trained word embedding (and we do not further optimize the embeddings when training the network).

For each language, we use the current dump of Wikipedia[1] as training data for the word embeddings. We use WikiExtractor[2] to extract plain text from the dump. We run a sentence splitter and we tokenize the sentences. Table 9.20 shows statistics of the data for each language.

Language	Sentences (M)	Tokens (M)
English	97.0	2103
French	26.2	633
Spanish	20.5	505
Russian	18.9	347
Dutch	15.4	252
Turkish	3.9	60
Arabic	3.1	63

Table 9.20: Sizes of Wikipedia dumps for each language.

[1] `http://wikipedia.org/`, we retrieved the current dumps on January 6, 2016.

[2] `http://medialab.di.unipi.it/wiki/Wikipedia_Extractor`

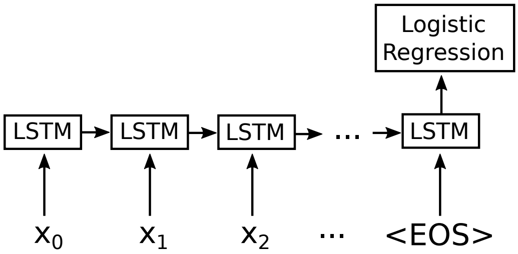


Figure 9.3: Architecture of our network (a single binary classifier).

We train the representations using `word2vec`[3] with continuous bag of words as the underlying model. We set the embedding size to 200.

Our network is essentially a deep LSTM encoder followed by a logistic regression layer with two output classes (neurons). For each sentence, we go over the input word by word and provide the embeddings of the tokens to the input layer. The hidden LSTM layers maintain a state at each step which encodes the (partial) sentence. After the final token, we input an artificial end-of-sentence token which signals the network to output a classification decision on the final layer. Depending on the activation of the two output neurons, we either classify the instance as positive (i.e. containing the given E#A pair) or negative.

Figure 9.3 shows an illustration of the architecture. We initially experimented with a single LSTM layer but stacking several layers lead to improved accuracy. All of the networks used in the final submission share an identical architecture consisting of an input layer (word2vec embeddings, size 200) followed by three LSTM layers (64, 32 and 32 cells) and the final layer with two neurons. We did not experiment with tuning a decision threshold; we simply assign the class (positive or negative) which has the higher probability according to the network. We implement the networks in Chainer,[4] an open-source framework for neural networks.

We evaluated several optimization algorithms and found that Adam (Kingma and Ba, 2014) lead to the fastest convergence. We also use gradient clipping as described in Pascanu et al. (2012): when the L2 norm of the gradients increases over a given threshold (which we set to 1), gradients are rescaled to fit within that norm. We also found dropout Srivastava et al. (2014) to improve the results and we utilize it in all LSTM layers with the probability set to 0.5.

When training, we use cross-validation and measure held-out accuracy to detect overfitting. We found that even though training error (almost) monotonically decreases, held-out performance tends to be rather unstable. Moreover, many classes are

---

[3] `https://code.google.com/archive/p/word2vec/`

[4] `http://chainer.org/`

Domain, Language	English	Spanish	French	Dutch	Russian	Turkish	Hotels Arabic	Laptops English
	Restaurants							
Off. Baseline	59.93	54.69	52.61	42.82	55.88	58.90	40.34	37.48
Baseline	58.05	**62.17**	**54.81**	54.71	60.75	34.41	49.43	35.08
Submitted	**59.30**	58.81	49.93	53.88	**64.83**	**61.03**	47.30	26.98
Optimized	58.40	58.54	50.84	**55.03**	60.19	56.54	**52.59**	**38.26**
Best	73.03	70.59	61.21	60.15	64.83*	61.03*	52.11	51.94

Table 9.21: F-measure of the baselines, the submitted system, the fully optimized system and the winning system for all domains and languages.

rare and the held-out set may therefore only contain a handful of positive instances. We were not able to mitigate this instability which made it difficult to choose the final model.

Nonetheless, we decide how many training iterations to run based on cross-validation; we measure the average f-measure over the folds for each iteration and then choose the iteration with the highest average.

9.3.3.3 Results

Official Baseline. We report the results of the official baseline (provided by the task organizers). The official baseline is an SVM classifier with bag-of-words features which assigns a positive label to a sentence if the predicted probability is above a certain threshold.

Baseline. We also implement our own baseline. We use a simple logistic regression model and, similarly to the submitted system, we train a binary classifier for each category. We only use bag of words as features; we do not include any other information (such as morphological analysis or subjectivity lexicon features). We use L2 regularization with weight 1 for all models. Our motivation for including this baseline is to provide a direct comparison with a simpler model trained in a similar way as the neural network.

Submitted. The system as submitted for the official evaluation. Due to the number of networks that we had to train, we were not able to fully optimize all of them before the submission deadline. In some cases, we therefore use models trained only on a handful of iterations over the training data.

Optimized. We report the results of the fully optimized system separately. These results were obtained after the deadline. In this case, all models were selected based on cross-validation as described in Section 9.3.3.2.

Best. To put the performance of our system into context of the state of the art, we also report the scores of the best system for each language and domain.

We evaluate our system using the toolkit provided by the organizers of the task. Table 9.21 shows all our results. The best result for each data set (excluding the winning system) is marked in bold. Asterisks mark the cases where our system won.

Concerning baselines, we observe that our implementation is often substantially better than the official baseline. This does not hold for the restaurants domain in English and Turkish, and for laptops. For Turkish, we suspect that the size of the test set (144 sentences) plays a role – the scores can be unstable when test data is small. For laptops, we believe this can be attributed to the large number of possible categories in this domain; our classifiers do not use a tuned threshold.

Our submitted systems do not always outperform the baselines. While we did win in Russian and Turkish, results in other data sets are less promising, with scores similar to the (stronger) baseline or even lower. This is a discouraging finding, especially considering the amount of additional data used in network training – word embeddings were trained on millions of sentences from Wikipedia.

However, we do observe some generalization in the outputs of the deep-learning models which is beyond the capabilities of the baseline models. For instance, our model correctly identifies the category FOOD#QUALITY in the sentence "Green Tea creme brulee is a must!", even though this dish is not mentioned in the training data.

Our *optimized* networks do not always perform better than the submitted systems. Considering that we trained the submitted networks with fewer iterations (due to time constraints), we suspect that even with our model selection based on cross validation, overfitting may still have affected the results.

Overall, there are several possible explanations for the weak performance. Due to significant domain mismatch between Wikipedia and the data sets for the task, the trained word embeddings may not be suitable.

Overfitting, and more generally, suboptimal setting of model (hyper)parameters, also most likely play a role.

The system design may also be problematic: we build a relatively large number of completely independent models (each doing binary classification for a single category) even though it seems clear that some parameter sharing should be possible. This problem is particularly prominent in data sets with a large number of possible classes, such as the laptop domain. In these cases, many E#A pairs are very rare. Because we do not decompose the entity and attribute and we train separate models, our classifiers only observe a handful of positive training instances, which results in very unreliable models.

9.3.3.4 Target Identification Using Neural Networks: Conclusion

Our system won in two categories but overall does not outperform a simple baseline solution. We believe that more careful training of the networks is required and that we may need to revise the system design. On the other hand, the deep-learning model does show some generalization power, so this direction seems promising.

III Conclusion

10

Summary

As stated in the introduction, this monograph has two main goals. The first one is to give a compact description of basic emotional language means in Czech. The second one is to employ the findings concerning emotional language in computational applications.

The first goal was primarily addressed in the first part of the publication. We described emotional meaning in Czech at different levels of linguistic description: lexical, morphosyntactic, semantic and pragmatic, excluding prosody.

Concerning the lexical level, we documented that besides the clearly interpretable lexical forms like vulgarisms, there are some less ambiguous subjectivity clues like euphemisms. However, even the words which seem unambiguous in terms of emotional meaning at first sight can be tricky in some utterances. Therefore, when dealing with automatic detection of emotional meaning in particular utterances, we always need to take into account the context of the whole sentence.

Besides, there are many strategies how to encode emotional meaning in the lexical units, which makes it difficult to get processed by the machines. This concerns the use of non-alphabetical signs or idioms. Since idioms are generally semantically specific, it is difficult to detect them automatically in the text, as has been verified in several computational experiments. The solution of this problem in emotional meaning detection issues could be adding a list of evaluative idioms or using of n-grams.

At the level of morphosyntax, we suggested that different parts of speech have different importance when assembling emotional meaning. Therefore, it is always advantageous to employ part of speech analysis when detecting emotional meaning automatically. Another important morphosyntactic feature which has to be taken into account is negation, since negation switches polarity of evaluation. Therefore, it has to be handled with a special care, especially in languages with a negative concord, such as Czech.

Regarding semantic aspects of emotional structures, special attention has to be given to their valency structures, since verbs play very often a central role in evaluation. In this respect, it is advantageous to perform semantic labelling on emotional structures when processing them automatically, especially when performing opinion target identification. Also, we surveyed discourse connectives and the Semantic Consistency Principle, which can be easily employed in automatic processing of emotional meaning.

Pragmatics is generally the most limiting aspect of emotional meaning research. We will probably never be able to know the whole context when assessing particular text. Although we proposed a definition of irony or sarcasm, we gave up on a definite identification of these phenomena in practical applications of emotional meaning research.

In addition to the above mentioned issues, we had to deal with the problem of perspective, meaning that the same sentence can be perceived differently by different participants of the communication situation. For instance the sentence *Liberal party won the elections* can be perceived as a good news by liberals and their voters, but as a bad news for the other parties. Therefore, it could be beneficial to employ a knowledge base in emotionality detection systems.

Finally, we suggested formal representation of emotional structures within the frameworks of Prague Dependency Treebank and Construction Grammar.

In the second part, we applied the above conclusions to automatic detection of emotional meaning in Czech. We introduced several newly created datasets, including Czech Subjectivity Lexicon, and tested their reliability in the main sentiment analysis tasks. Also, we built three sentence-level polarity classifiers reaching rather satisfactory performance. Moreover, we prepared an opinion target identification system and tested its performance on various language data, again with quite satisfactory results.

All the above mentioned phenomena cause different issues in automatic processing of emotions. Current methods are especially limited by the pragmatic aspects of emotional meaning (this is similar to many other areas of natural language processing). This also has to do with the need of the domain adaptation of the language resources created within the scope of this work. When working with these sources, one has to be very careful when using classifiers trained on one domain to process data from other domain, even if the domain seems to be very specific, cf. e.g. *tichý vysavač* – 'silent vacuum cleaner' vs. *tichý budík* – 'silent alarm clock'.

11

Future Work

Apart from further work on the particular topics investigated in this study, we would like to continue in several other areas connected to emotional language research.

- We would survey prosody and information structure of evaluative utterances, since it turned out that it has a crucial impact on emotional impression the utterance gives, especially in case of irony and sarcasm.
- We would be interested in the influence of the text domain on the choice of language means and pragmatics in general, since evaluative language is a type of meaning expression which applies across different genres, from everyday conversation up to scientific discourse.
- To test our findings in the domain of lexical semantics and semantics, we would like to employ psycholinguistic experiments to find out e.g. whether there is a scale in emotional expressions intensity or how particular experiment participants differ in perceiving emotions expressed in Czech.
- Generally, we would experiment more with neural networks in sentiment analysis task. Namely, we would employ convolutional neural networks and corpus-dependent training of word embeddings when predicting the intensity of attitudes in Czech texts. Also, we would use linguistically annotated data (e.g. Universal Dependencies, available from `http://universaldependencies.org/`) for the networks training.
- We would like to improve opinion target identification by employing machine-learning methods other than CRF and to perform aspect-based sentiment analysis (i.e. analysis taking into account also particular aspects of evaluated targets) on the data.
- We would focus more on opinion mining, combining polarity detection with clustering opinions based on their semantic similarity.
- We would like to study multi-modal resources of emotional data (like e.g. the Czech part of the Visual History Archive (VHA) of the USC Shoah Foundation, available from `http://sfi.usc.edu/`) to connect our findings with other non-language means of emotion expressing, as indicated e.g. in Veselovská (2014a). Results of such a research could be applicable e.g. when improving emotional expressing of the dialogue systems.
- We would like to perform a cross-linguistic comparison of emotional expressions e.g. to determine whether similar languages use similar means to express emotions or how particular languages differ in this respect. Thus, we would

focus on the comparison of our results with the current results from the field of modelling evaluative language for other languages.

All of these topics will be very exciting yet challenging subjects of further investigation within the endlessly inspiring domain of emotions in language.

IV Appendices

A

Data Overview

Many of the examples used in this work are real utterances taken from Czech National Corpus (CNC, `http://korpus.cz/`), Prague Dependency Treebank (PDT, `http://ufal.mff.cuni.cz/pdt3.0`) or manually annotated sources generated by Internet users. Table A.1 provides an overview of the resources we used, together with their basic statistics.

The CNC consists of various corpora. SYN corpora is a non-referential union of all referential synchronic written corpora of the SYN series (including SYN2000, SYN2005, SYN2006PUB, SYN2009PUB, SYN2010 a SYN2013PUB). InterCorp is a large parallel synchronic corpus covering a number of languages (version 7 contains 38 languages). In this study, InterCorp was mostly used for comparative purposes, i.e. to get a right referential translation to English. All the corpora can be searched via the KonText tool (`http://korpus.cz/kontext`).

PDT consists of the data from two daily newspapers, a business weekly and a popular scientific journal. The data is annotated on three layers: the morphological layer, analytical layer (surface dependency syntax), and tectogramatical layer, as described in detail in Section 7.1. The corpus can be searched via the PMLTQ tool (`http://ufal.mff.cuni.cz/pmltq/`).

Czech Subjectivity Lexicon (Czech SubLex) is a dictionary of evaluative items in Czech.

Concerning the manually annotated data, Aktualne.cz are journalistic texts concerning domestic news. The data from the CSFD.cz are the collection movie reviews written by movie database users and assigned with 1–5 asterisks (1 is the worst, 5 is the best). Mall.cz dataset consists of appliance reviews classified as positive or negative depending in which field they were filled. Data from the social network Facebook.com are the Facebook posts manually categorized into positive, negative, neutral and bipolar. SemEval2014 are data provided by the organizers of SemEval task focused on aspect-based sentiment analysis and thus annotated on sentiment aspects and their categories. The dataset contains restaurants and laptops reviews. Alza.cz data are positive and negative laptop reviews annotated on opinion targets.

Corpus	Size	Topic	Source		Period	Annotation
SYN2000	100 MW	balanced	written	15% fiction	20th cent	morph automatic
				25% non-fiction		
				60% news	1990-99	
SYN2005	100 MW	balanced	written	40% fiction	20th cent	morph automatic
				27% non-fiction	1990-2004	
				33% news	2000-04	
SYN2006pub	300 MW	news	written		1989-2004	morph automatic
SYN2010	100 MW	balanced	written	40% fiction	20th cent	morph, syntax,
				27% non-fiction	1989-2010	automatic
				33% news	2005-2009	
InterCorp	1,390 MW	mostly fiction	written		20th cent	morph automatic
PDT	2 MW	news, scientific journal	written			morph, syntax, manual
Czech SubLex	4,625 W	evaluative items	MPQA translation			polarity, manual
Aktualne.cz	560,000 W	domestic news	written, online		2010	eval, manual
CSFD.cz	6,868 W	movie reviews	written, online		2011	eval, manual
Mall.cz	158,955 W	appliance reviews	written, online		2011	eval, manual
Facebook.com	139,222 W	Facebook posts	written, online		2012	polarity, manual
SemEval2014	7,686 W	rest. & laptop reviews	written, online		2014	targets, manual
Alza.cz	2,000 W	hardware reviews	written, online		2015	targets, manual
SemEval2016	N/A	various	written, online		2016	targets, manual

(W stands for word, MW stands for million of words)

Table A.1: Corpora used as example sources

B

Annotation Guidelines

This appendix is to explain the annotation process for Aktualne.cz and CSFD.cz data and to briefly describe what kind of annotation was performed on the other manually annotated online resources used in our experiments. The resources are ordered with respect to the order in which they appear in the monograph. For data statistics, annotation statistics, inter-annotator agreement and evaluation, see Chapter 8.

When preparing the annotation, it turned out that it is very difficult for the annotators to agree on what is actually emotional and what is not, based on their individual preferences and generally on the whole pragmatic situation. To smooth the annotation process as much as possible, we had to prepare the guidelines clearly stating what to annotate and what not to annotate, i.e. what is and what is not the evaluative part of the text. Below we provide annotation guidelines for the two rounds of Aktualne.cz annotation and the CSFD.cz annotation. Also, we explain how the Mall.cz, Facebook.com and Alza.cz data were annotated by the users.

B.1 Aktualne.cz

B.1.1 First Round

When annotating Aktualne.cz, the annotators were asked to respect the following guidelines in the first round. The first round was to prove that is it even possible to mark the sentences on evaluation with a reasonable agreement (which was confirmed by the annotation results). Aktualne.cz annotation was performed in the simple .doc format.

Annotation In the Word document, please mark the targets of the evaluation (i.e. entities towards which the evaluation is being expressed):

- with **bold** when positive
- with *italic* when negative
- with the word EXTERN in an according bold/italic format at the end of the sentence when the target is not expressed on the surface structure
  e.g. *Blbec.* – 'Dummy.' – *EXTERN*
- when the target happens to be a dependent clause, mark the root of it, e.g.

Pan Novák si stěžoval, že se bude kácet alej.
Mr. Smith complained about the fact that they plan to saw down the alley.

mark *že* – 'that'

Segmentation

- sentence segmentation
- split also clauses inside direct speech (Wall Street Journal style)
- keep headlines and other non-sentential fragments
- get rid of the time entries in minute-by-minute reports from the Chamber of Deputies
- get rid of the links, if still there for some reason
- get rid of the introductory figures (but these should be removed by the cleaning script)
- get rid of the remaining figure descriptions

Miscellaneous

- The problematic constructions are supposed to be set aside for the moment. Please copy them to the "Problem sentences folder" any time you are not sure about the right solution. They will be surveyed later on.
- forget about the source of evaluation, e.g. when finding a sentence like e.g.

  Václav Klaus zveřejnil seznam osob, které podle něj škodí státu.
  Václav Klaus published a list of people who according to him are harmful to the country.

  do not mark the word "people" with a positive polarity
- keep the html formatting
- mark *(non)polar_sentences_merged*
- make notes
- measure time
- do not think too much, follow the first impression

B.1.2 Second Round

After we finished the first round, it turned out there were some shortages in the guidelines. Therefore, we improved the guidelines for the second round of the annotation (performed on a different part of the data).

Concerning the target notation, we decided to conduct two types of annotation: annotation of any kind of entities and annotation of selected targets, namely Czech politicians. For the first task, the rules remained the same as in the first round, except in a dependency clauses annotation, we decided to mark the whole evaluated part instead of just a root. Also, we had to take into account coreference. We decided to mark the referential noun or pronoun which is closest to the evaluative expression. One of the problems discovered during the first round was a multiple target of the sentence

(*Prohrát je špatné a vyhrát je ještě horší* – 'To lose is bad and to win is even worse' In these cases, we agreed to mark all the targets with the given polarities. Moreover, we had to newly deal with the "bad news" category (e.g. *Šumava není v pořádku, došlo tady ke katastrofě* – 'Šumava is not fine, it experienced a calamity'). However, we decided not to mark it for the moment and we put the problematic sentences aside.

In annotation of selected targets, the annotators were supposed to mark the sentence as evaluative even if the target was not mentioned there at all, but they felt the content was either good or bad for the person. However, we were not able to get rid of the *hallo effect*, i.e. a situation in which the annotator marked the sentence based on his or her previous sympathies (or antipathies) towards the target. To avoid this, we masked the data with the false names. However, the original targets were still easily recognizable based on the context.

As for the segmentation, we decided to change the format to one sentence per one line for clarity reasons. Also, we had to ask the annotators not to add punctuation to the segments which were not considered sentences (segments like headlines etc.), since the tended to do "improve" them.

Section Miscellaneous remained the same.

B.2 CSFD.cz

The CSFD.cz data annotation guidelines got more complex in comparison to annotation of Aktualne.cz data, since we decided to mark particular participants (the source, the target and the evaluative element) of the evaluative structures individually. CSFD.cz annotation was performed in the .xls format.

Annotation
In CSFD.cz data, we decided to mark two types of evaluative expressions: evaluative states and evaluative elements.

- **Evaluative states/events** – parts of the text where the speaker (even mediated) or writer expresses evaluation towards any entity. Evaluative states consist of SOURCE, TARGET and EVALUATIVE ELEMENT.
- **Evaluative elements** – word or phrases which inherently express evaluation, but they are not evaluative in the given context
  - **Expressive subjective elements**, i.e. explicitly evaluative expressive items, which can work either alone or as the parts of evaluative states in which they are also marked:

    Skvělé! – 'Great!' v.s. *Podle Pavla jsou ty šaty skvělé.* – 'Pavel thinks this dress is great.'

  - **Good/bad news** – items in which we feel evaluative meaning and we are able to describe them in terms of positive or negative polarity. When they

are a part of evaluative state, they are often a target and evaluation at the same time (e.g. *katastrofa* – 'disaster', *problém* – 'problem' etc.).
  - **Elusive elements**, i.e. items in which we feel evaluative meaning, but we are not able to describe them in terms of positive or negative polarity without using a complicated inference, e.g. *kontroverze* – 'controversy'

In connection with evaluative elements, we usually do not annotate source and target. A potential source is usually the author of the text.

Particular columns in annotation excel table are:

- **Source** – evaluating entity, can also get values EXTERN or AUTHOR. AUTHOR is the final, least embedded source of the evaluation, mostly useful as a "general source". AUTHOR can be annotated e.g. when all the other direct sources are missing.

  Courtney Love (47) vypadá příšerně. – 'Courney Love (47) looks terribly.'

  In this sentence, SOURCE is the AUTHOR, EVALUATIVE ELEMENT is looks terribly and TARGET is Courtney Love.
  !!WARNING!! EXTERN and AUTHOR are not the same entities. EXTERN is just a source, which is situated outside the sentence, but it is recognizable from the context.
  If there are more evaluative structures in the sentence, one should copy the original sentence and describe the second evaluative structure on the copied sentence, like e.g. in:
  Pankrác nemá rád, když si Servác stěžuje na Bonifáce, který miluje Žofii.
  Pankrác does not like when Servác complains about Bonifác, who loves Žofie.

  - SOURCE Pankrác, ELEMENT does not like, TARGET when Servác complains about Bonifác, who loves Žofie
  - SOURCE Servác, ELEMENT complians, TARGET Bonifác, who loves Žofie
  - SOURCE Bonifác, ELEMENT loves, TARGET Žofie
- **Evaluative element** – the part of the text which contains the evaluation itself
- **Target** – evaluated entity (can get the value EXTERN)
- **Orientation** – orientation of evaluation polarity, which can be
  - NEG
  - POS
  - NONNEG, e.g. not bad – not the same as POS!
  - NONPOS, e.g. not nice – not the same as NEG!
- **Good/bad news**
- **Elusive element**
- **Expressive subjective element**

- **False polarity** – column for annotation of parasite verbs (a verbal form of good/bad news, *schválit zákon* – 'to pass a law', *ocenit snahu* – 'to appreciate an effort')
- **Note** – notes, comments etc.

Segmentation
One sentence per line (this holds also for xml tags).

Technical

- all the columns should be filled with the original word forms not lemmas
- follow maximalistic approach: annotate the whole target, the whole source and all the evaluative elements
- keep the "significant-and-particular" principle: if you are not sure whether something is or is not evaluative, do not annotate it
- if there are more parts of evaluative element, write them into one field separated by a semicolon

B.3 Mall.cz

The Mall.cz data were annotated by the users, who simply filled the text into positive/negative column.

B.4 Facebook.com

The Facebook posts were manually annotated as positive, negative, neutral and bipolar in the .csv format.

B.5 SemEval2014

There were two data domains in the SemEval2014 dataset, namely restaurants data and laptop data. The restaurants data were annotated for target terms, target categories, target category polarities and overall sentence polarities. The laptops data extracted from customer reviews and annotated with target terms and their polarities. Both datasets were available in the .xml format.

B.6 Alza.cz

The data in the Alza.cz dataset were divided into positive and negative entries, depending on whether they were filled in the positive or negative form. Both datasets were manually annotated with opinion targets. The data were stored in the .xml format.

B.7 SemEval2016

There were several data domains in the SemEval2016 dataset, namely hotels, restaurants, consumer electronics and telecom data. The data was annnotated on opinion target, opinion target category and sentiment. All datasets were available in the .xml format.

C

Computational Basics

This appendix summarizes the basics of computational metrics used in this study. The metrics are ordered with respect to the order in which they appear in the book.

C.1 Cohen's kappa

Cohen's kappa measures the agreement between two annotators each of whom classifies X items into Y mutually exclusive categories.

$$\kappa = \frac{P(a) - P(e)}{1 - P(e)} \tag{C.1}$$

where P*(a)* is the relative observed agreement among annotators, and P*(e)* is the probability of random agreement, using the observed data to calculate the probabilities of each annotator randomly saying each category.

C.2 Accuracy, Precision and Recall

We compute accuracy, precision and recall using the following confusion matrix:

		Predicted	
		Negative	Positive
Actual	Negative	True Negative	Negative Positive
	Positive	False Negative	True Positive

- Accuracy is the proportion of the total number of predictions that were correct.

$$AC = \frac{TN + TP}{TN + FN + FP + TP} \tag{C.2}$$

- Precision P is the number of correct positive results divided by the number of all positive results.

$$P = \frac{TP}{FP + TP} \tag{C.3}$$

- Recall R is the number of correct positive results divided by the number of positive results that should have been returned.

$$R = \frac{TP}{FN + TP} \tag{C.4}$$

C.3 F-score

F-score is a measure of a test's accuracy. It takes into account both the *precision P* and the *recall R* of the test to compute the score.

$$F = 2 \times \frac{P \times R}{P + R} \tag{C.5}$$

C.4 Naive Bayes Classifier

For polarity classification in Czech, we use Naive Bayes classifier. Naive Bayes is a discriminative model which makes strong independence assumptions about its features, i.e. properties of phenomenon which is being surveyed. Generally, in polarity classification, features are words and they are assigned with the likelihood that they express either positive or negative polarity. To feed the classifier, one first needs to create a large manually annotated training data. The datasets used in all our experiments are described in detail in Chapter 8. Sample of the annotated data can be found in Appendix X.

In Chapter 9, we build the Naive Bayes classifier the following way:

Let $\mathcal{C}$ denote a set of polarity classes $C_1, C_2 ... C_{|\mathcal{C}|}$. The classified unit is a segment, denoted s_j from a set of segments $\mathcal{D}$. A segment s_j is composed of n lemmas $s_{j,1}, s_{j,2} \dots s_{j,n}$. Each lemma actually has three factors: the "real" lemma itself, its part of speech (N, V, A etc.) and the negation tag, as described below within experimental settings. However, for the purposes of the classifier, it is important to keep negation with the real lemma, as disposing of it would make e.g. *flattering* and *unflattering* indistinguishable. The lexicon is then the set of all lemmas in $\mathcal{D}$ and is denoted as $\mathcal{L}$. The size of the lexicon, i.e. the number of distinct lemmas in the lexicon, is M. The classification features $F_i, i = 1 \dots M$ are then the presence of the i-th lemma l_i in the classified segment.

Given that the probability of classifying a segment as belonging to C is

$$p(C|F_1, F_2, \dots F_M) \propto p(C)p(F_1, F_2, \dots F_M|C) \tag{C.6}$$

by the Chain Rule ($p_C p(F_1, F_2, \dots F_M|C) = p(C, F_1, F_2, \dots F_M)$), and by assuming conditional independence of features $F_1 \dots F_M$ on each other it yields the following formulas:

$$p(C|F_1, F_2 \dots F_M) \propto p(C) \prod_{i=1 \dots M} p(F_i|C) \tag{C.7}$$

$$\propto \log p(C) + \sum_{i=1 \dots M} \log p(F_i|C) \tag{C.8}$$

Maximization follows by simply $\mathtt{argmax}$-ing over both sides. The model parameters – conditional probabilities of seeing the lemma l_i in each of the classes $C_1 \dots C_{|\mathcal{C}|}$ – can be estimated as Maximum Likelihood Estimations $p_T(F_i|C)$ on some training data $T = w_1 \dots w_{|T|}$ with Laplacian smoothing of strength α, computed as

$$p_T(F_i|C) = \frac{\mathtt{freq}(i, C) + \alpha}{\mathtt{freq}(C) + \alpha|\mathcal{C}|} \tag{C.9}$$

where $\mathtt{freq}(i, C)$ is the number of times lemma l_i was seen in a segment s_j labeled as belonging to the class C.

A special *UNSEEN* lemma for words that are not in the lexicon and thus cannot be lemmatized was also added to the model, with parameters $p(C|UNSEEN)$ estimated as the marginal probabilities $p(C)$ – the probability of something generating a polarity class should be the general probability of seeing that class anyway.

C.5 Maximum Entropy Classifier

The Maximum Entropy classifier is a discriminative classifier. It is called maximum entropy, because it makes as few assumptions as possible: from all the models that fit the training data, it selects the one which has the largest entropy. In other words, it prefers distributions that are as uniform as possible. These minimal assumptions also mean the classifier is fairly robust and it can be used when we have little or no knowledge of the prior distribution. Unlike Naive Bayes, a MaxEnt classifier does not assume that the features are conditionally independent of each other.

C.6 Conditional Random Fields

Conditional Random Fields are a category of statistical modelling methods often applied in machine learning for prediction purposes. Since CRF take into account a context in the form of a linear chain, it can predict sequences of labels for sequences of input data.

C.7 Cross-validation

Cross-validation is a method for evaluation a model estimating how the results of a statistical analysis could generalize to an independent dataset.

D

TrEd Extension

This Appendix describes a structure of tree editor TrEd extension for annotating emotional structures in Prague Dependency Treebank. The extension is called `PML_T_Sentiment` and it brings the new attribute `sentiment`, which is empty by default. The attribute `sentiment` consists of four parts:

- `sent_source` – this field is filled with the number of node to which points the referential arrow for source or can have a value EXTERN when the source is not present anywhere in the text
- `sent_eval` – this field is marked with the `pol_pos` or `pol_neg` value
- `sent_target` – this field is filled with the number of node to which points the referential arrow for target or can have a value EXTERN when the target is not present anywhere in the text
- `was_annotated` – this field is filled with the 0/1 values to make sure the item was checked by the annotator

All the attributes are shown in Fig. D.1. The are part of the node representing evaluative expression. For more fine-grained analysis, the sub-attributes like `sent_eval` status or `sent_eval` with the values as judgement, success and failure etc. can be added.

Everything is annotated in the annotation mode `PML_T_Sentiment`. Annotators jump on the highlighted evaluative nodes marked as positive or negative. All potentially evaluative items are marked with the polarity based on the lexicon. The node is marked with a green colour when positive and with red when negative. Besides, the evaluative part of the above text is marked with yellow. The colour can be either switched or the changed to none when the expression is not evaluative in the given

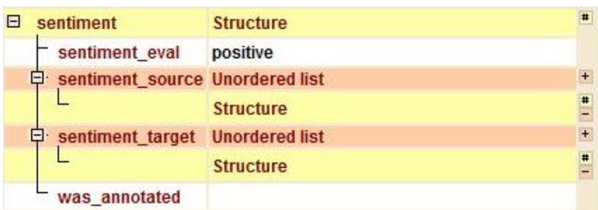


Figure D.1: Sentiment annotation GUI in TrEd context. The colour is changed using “p” as a switch for positive polarity and “n” as a switch for negative polarity. When the initial polarity is confirmed, the light green turns to dark green and the light red turns to dark red. Also, the status was annotated is switched to 1. After pressing the space bar, all values of the attribute sentiment are deleted.

Attributes `sent_source` and `sent_target` are marked by the arrow pointing from the evaluative item. There is a macro for this action: an evaluative node is moved to the selected node when pressing a particular key (the same principle as used for coreference annotation in TrEd, but the arrows for sentiment are of different colours).

List of Figures

List of Tables

List of Abbreviations

ACT	actor
ADDR	addressee
Adv	adverbial
APP	appurtenance
Attr	attribute
cat	category
CAUS	cause
CDA	Critical Discourse Analysis
CNC	Czech National Corpus
CPM	Component Process Model
CRF	conditional random fields
CxG	Construction Grammar
EFF	effect
Em	emotional class
Ex	expressive class
FE	frame element
FGD	Functional Generative Description
GUI	graphical user interface
LB	lexicon-based
LI	lexicon indicator
LM	language model
MPQA	multi-perspective question answering
NB	Naive Bayes
NEG/POS	negative/positive
NLP	natural language processing
NOM	nominative
NONNEG	non-negative
NONPOS	non-positive
NTR	neutral

Obj	object
ORIG	origo
PAdj	predicative adjective
PAT	patient
PDiT	Prague Discourse Treebank
PDT	Prague Dependency Treebank
PMLTQ	Prague Markup Language Tree Query
PNoun	predicative noun
POS	Part of Speech
prag	pragmatics
Pred	predicate
PRED	predicate
REFL	reflexive
RHEM	rhematizer
SB	SubLex-based
sem	semantics
SFL	Systemic Functional Linguistics
SRILM	Stanford Research Institute language modelling
sub	subject
Subj	subject
SubLex	subjectivity lexicon
SVM	Support Vector Machine
TrEd	Tree Editor
val	valency

Bibliography

Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. Sentiment analysis of twitter data. In *Proceedings of the Workshop on Languages in Social Media*, pages 30–38. Association for Computational Linguistics, 2011.

Basant Agarwal and Namita Mittal. Machine learning approach for sentiment analysis. In *Prominent Feature Extraction for Sentiment Analysis*, pages 21–45. Springer, 2016.

Artemis Alexiadou and Florian Schäfer. Instrument subjects are agents or causers. In *Proceedings of the 25th West Coast Conference on Formal Linguistics*, pages 40–48, 2006.

Angeliki Athanasiadou. On the subjectivity of intensifiers. *Language Sciences*, 29(4):554–565, 2007.

John Langshaw Austin. *How to do things with words*, volume 367. Oxford University Press, 1975.

Fabienne Baider and Georgeta Cislaru. *Linguistic Approaches to Emotions in Context*, volume 241. John Benjamins Publishing Company, 2014.

Kryštof Bajger. *Šestijazyčný slovník vulgarizmů*. IMPEX, 1998.

Akshat Bakliwal, Piyush Arora, and Vasudeva Varma. Hindi subjective lexicon: A lexical resource for Hindi polarity classification. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC)*, pages 1189–1196, 2012.

Alexandra Balahur, Ralf Steinberger, Mijail Kabadjov, Vanni Zavarella, Erik van der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva. Sentiment analysis in the news. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'2010)*, pages 2216–2220, 2010.

Carmen Banea, Rada Mihalcea, and Janyce Wiebe. A bootstrapping method for building subjectivity lexicons for languages with scarce resources. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pages 2764–2767, 2008a. ISBN 2-9517408-4-0.

Carmen Banea, Rada Mihalcea, Janyce Wiebe, and Samer Hassan. Multilingual subjectivity analysis using machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 127–135. Association for Computational Linguistics, 2008b.

Laurie Bauer. Evaluative morphology: in search of universals. *Studies in Language*, 21(3):533–575, 1997.

Paul J Beauvais. A speech act theory of metadiscourse. *Written Communication*, 6(1):11–30, 1989.

J. V. Bečka. Subjektivnost a polarita výrazu. *Naše řeč*, 58:1–12, 1975.

Eduard Bejček, Jarmila Panevová, Jan Popelka, Lenka Smejkalová, Pavel Straňák, Magda Ševčíková, Jan Štěpánek, Josef Toman, Zdeněk Žabokrtský, and Jan Hajič. Prague dependency treebank 2.5, 2011.

Eduard Bejček, Eva Hajičová, Jan Hajič, Pavlína Jínová, Václava Kettnerová, Veronika Kolářová, Marie Mikulová, Jiří Mírovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Magda Ševčíková, Jan Štěpánek, and Šárka Zikánová. Prague dependency treebank 3.0, 2013.

Adriana Belletti and Luigi Rizzi. Psych-verbs and θ-theory. *Natural Language & Linguistic Theory*, 6(3):291–352, 1988.

Farah Benamara, Carmine Cesarano, Antonio Picariello, Diego Reforgiato Recupero, and Venkatramana S Subrahmanian. Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In *International Conference on Weblogs and Social Media (ICWSM)*, pages 1–7, 2007.

Adam L. Berger, Vincent J. Della Pietra, and Stephen A. Della Pietra. A maximum entropy approach to natural language processing. *Computational linguistics*, 22(1):39–71, 1996.

Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan, and Randolph Quirk. *Longman grammar of spoken and written English*. Longman, Harlow, Essex, 1999.

Ondřej Bojar and Zdeněk Žabokrtský. CzEng: Czech-English Parallel Corpus release version 0.5. *Prague Bulletin of Mathematical Linguistics*, 86:59–62, 2006.

Ondřej Bojar, Zdeněk Žabokrtský, Ondřej Dušek, Petra Galuščáková, Martin Majliš, David Mareček, Jiří Maršík, Michal Novák, Martin Popel, and Aleš Tamchyna. The Joy of Parallelism with CzEng 1.0. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 3921–3928, 2012.

Lenka Brdečková. Die bewertenden Sprachmittel der deutschen Sprache. Master's thesis, Masarykova univerzita, 2006.

Tomáš Brychcín and Ivan Habernal. Unsupervised improving of sentiment analysis using global target context. In *Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP)*, pages 122–128, 2013.

Radim Burget, J. Karásek, and Z. Smékal. Classification and Detection of Emotions in Czech News Headlines. In *The 33rd International Conference on Telecommunication and Signal Processing (TSP)*, volume 2010, pages 1–5, 2010.

Erik Cambria, Soujanya Poria, Federica Bisio, Rajiv Bajpai, and Iti Chaturvedi. The clsa model: A novel framework for concept-level sentiment analysis. In *Computational Linguistics and Intelligent Text Processing*, pages 3–22. Springer, 2015.

Jan Čáp. *Psychologie pro učitele*. SPN, Praha, 1980.

Marie Čechová, Jan Chloupek, Eva Minářová, et al. *Stylistika současné češtiny*. Institut sociálních vztahů – nakladatelství, Praha, 1997.

František Čermák. *Frazeologie a idiomatika česká a obecná: Czech and general phraseology*. Univerzita Karlova, Praha, 2007.

František Čermák. *Slovník české frazeologie a idiomatiky: Výrazy větné*, volume 4. Leda, Praha, 2009.

Radek Červenec. *Rozpoznávání emocí v česky psaných textech*. PhD thesis, Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií, 2011.

Yejin Choi and Claire Cardie. Learning with compositional semantics as structural inference for subsentential sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 793–801. Association for Computational Linguistics, 2008.

Světla Čmejrková and Jana Hoffmannová. *Jazyk, média, politika*. Academia, Praha, 2003.

Jacob Cohen. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*, 70(4):213, 1968.

Susan Conrad and Douglas Biber. Adverbial marking of stance in speech and writing. In *Evaluation in text: Authorial stance and the construction of discourse*, pages 56–73. Oxford University Press, 2000.

Avon Crismore and Rodney Farnsworth. Metadiscourse in popular and professional science discourse. In *The Writing Scholar: Studies in the Language and Conventions of Academic Discourse*, pages 45–68. Sage, 1990.

Hang Cui, Vibhu Mittal, and Mayur Datar. Comparative experiments on sentiment classification for online product reviews. In *Association for the Advancement of Artificial Intelligence (AAAI)*, volume 6, pages 1265–1270, 2006.

František Daneš. Kultura mluvených projevů (její základní předpoklady a aktuální problémy). *Naše Řeč*, 52:95–109, 1969.

Tom De Smedt and Walter Daelemans. "Vreselijk mooi!"(terribly beautiful): A Subjectivity Lexicon for Dutch Adjectives. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 3568–3572, 2012.

Daantje Derks, Arjan ER Bos, and Jasper von Grumbkow. Emoticons and social interaction on the internet: the importance of social context. *Computers in Human Behavior*, 23(1):842–849, 2007.

Miloš Dokulil, K. Horálek, J. Hůrková, M. Knappová, and J. Petr. *Mluvnice češtiny. 1. Fonetika, fonologie, morfonologie a morfemika, tvoření slov*. Academia, Praha, 1986.

Cícero Nogueira Dos Santos and Maira Gatti. Deep convolutional neural networks for sentiment analysis of short texts. In *COLING*, pages 69–78, 2014.

John W. Du Bois. The stance triangle. In *Stancetaking in discourse: Subjectivity, evaluation, interaction*, pages 139–182. John Benjamins Publishing, Amsterdam, 2007.

Sebastian Ebert, Ngoc Thang Vu, and Hinrich Schütze. Cis-positive: A combination of convolutional neural networks and support vector machines for sentiment analysis in twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 527–532, Denver, Colorado, June 2015. Association for Computational Linguistics.

Susana A. Eisenchlas. On-line interactions as a resource to raise pragmatic awareness. *Journal of Pragmatics*, 43(1):51–61, 2011.

Robert Englebretson. *Stancetaking in discourse: Subjectivity, evaluation, interaction*, volume 164. John Benjamins Publishing, Amsterdam, 2007.

Andrea Esuli and Fabrizio Sebastiani. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of the Fifth Conference on Language Resources and Evaluation (LREC'O6)*, pages 417–422, 2006.

Norman Fairclough. *Critical discourse analysis. The critical study of language. Language in social life series*. Longman, London, 1995.

Elena Filatova. Irony and sarcasm: Corpus generation and analysis using crowdsourcing. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 392–398, 2012.

Charles J. Fillmore. The mechanisms of "construction grammar". In *Annual Meeting of the Berkeley Linguistics Society*, volume 14, 1988.

Johnny R.J. Fontaine, Klaus R. Scherer, and Cristina Soriano. *Components of emotional meaning: A sourcebook*. Oxford University Press, 2013.

Mirjam Fried. Construction grammar. *Handbook of syntax (2nd ed.)*, pages 974–1003, In press.

Mirjam Fried and Jan-Ola Östman. *Construction grammar in a cross-language perspective*, volume 2. John Benjamins Publishing, Amsterdam, 2004.

Mirjam Fried and Jan-Ola Östman. Construction grammar and spoken language: The case of pragmatic particles. *Journal of pragmatics*, 37(11):1752–1778, 2005.

Gayatree Ganu, Noemie Elhadad, and Amélie Marian. Beyond the stars: Improving rating predictions using review text content. In *WebDB*, volume 9, pages 1–6, 2009.

Alec Go, Richa Bhayani, and Lei Huang. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1(2009):12, 2009.

Philip W. Graham. Critical discourse analysis and evaluative meaning: Interdisciplinarity as a critical turn. In *Critical discourse analysis: Theory and interdisciplinarity*, pages 110–129. Palgrave Macmillan, New York, 2003.

Miroslav Grepl. *Emocionálně motivované aktualizace v syntaktické struktuře výpovědi*, volume 113. Univerzita J. E. Purkyně, 1967.

Herbert P. Grice. Some further notes on logic and conversation. *Syntax and Semantics*, 9:113–127, 1978.

Ivan Habernal and Tomáš Brychcín. Semantic spaces for sentiment analysis. In *Text, Speech, and Dialogue*, pages 484–491. Springer, 2013.

Ivan Habernal, Tomáš Ptáček, and Josef Steinberger. Sentiment analysis in czech social media using supervised machine learning. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 65–74, 2013.

Jan Hajič, Jarmila Panevová, Eva Hajičová, Petr Sgall, Petr Pajas, Jan Štěpánek, Jiří Havelka, Marie Mikulová, Zdeněk Žabokrtský, and Magda Ševčíková. Prague dependency treebank 2.0. *CD-ROM, Linguistic Data Consortium, LDC Catalog No.: LDC2006T01, Philadelphia*, 98, 2006.

Jan Hajič, Jan Votrubec, Pavel Krbec, Pavel Květoň, et al. The best of two worlds: Cooperation of statistical and rule-based taggers for czech. In *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing: Information Extraction and Enabling Technologies*, pages 67–74. Association for Computational Linguistics, 2007.

Jan Hajič Jr. *Popularita osob automaticky*. Bc. Thesis, Univerzita Karlova, 2011.

Eva Hajičová. *Negace a presupozice ve významové stavbě věty*. Academia, Praha, 1975.

Eva Hajičová. The functional generative description as a functionally motivated formal model of language. In *SLE - 45th Annual Meeting of the Societas Linguistica Europaea*, pages 130–131, Stockholm, Sweden, 2012. Stockholm University, Societas Linguistica Europaea.

Michael A.K. Halliday. *Functional grammar*. Edward Arnold, London, 1994.

Michael Alexander Kirkwood Halliday. *Explorations in the functions of language*. Edward Arnold, London, 1973.

Richard Mervyn Hare. *The language of morals*. Oxford Clarendon Press, 1961.

Vasileios Hatzivassiloglou and Kathleen R McKeown. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, pages 174–181. Association for Computational Linguistics, 1997.

Karel Hausenblas. On the characterization and classification of discourses. *Travaux linguistiques de Prague*, 1:67–84, 1964.

Karel Hausenblas. *Výstavba jazykových projevů a styl*. Universita Karlova, 1971.

Kenneth Heafield, Philipp Koehn, and Alon Lavie. Grouping language model boundary words to speed k-best extraction from hypergraphs. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2013)*, pages 958–968, 2013.

Milena Hnátková. Značkování frazémů a idiomů v českém národním korpusu s pomocí slovníku české frazeologie a idiomatiky. *Slovo a slovesnost,(63)*, pages 117–126, 2002.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8): 1735–1780, November 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735.

Jana Hoffmannová. *Stylistika a–: současná situace stylistiky*. Trizonia, Praha, 1997.

Jana Hoffmannová and Olga Müllerová. *Čeština v dialogu generací*. Academia, Praha, 2007.

Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM, 2004.

Susan Hunston. *Corpus approaches to evaluation: Phraseology and evaluative language*. Routledge, 2011.

Ken Hyland and Polly Tse. Metadiscourse in academic writing: A reappraisal. *Applied linguistics*, 25(2):156–177, 2004.

Nitin Indurkhya and Fred J Damerau. *Handbook of natural language processing*, volume 2. CRC Press, 2012.

Tomoko Izumi, Tomohide Shibata, Hisako Asano, Yoshihiro Matsuo, and Sadao Kurohashi. Constructing a corpus of japanese predicate phrases for synonym/antonym relations. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC)*, pages 1394–1400, 2014.

Roman Jakobson. Closing statement: Linguistics and poetics. *Style in language*, 350:377, 1960.

Martin Janečka. Korpusová analýza výrazů vyjadřujících postojovou a evaluativní modalitu k propozici výpovědi. In *Lingvistika Praha (online)*. Univerzita Karlova v Praze, Filozofická fakulta, 2014.

Hayeon Jang and Hyopil Shin. Language-specific sentiment analysis in morphologically rich languages. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 498–506. Association for Computational Linguistics, 2010.

Milan Jelínek. Stylistika. In *Příruční mluvnice češtiny*, pages 699–780. Nakladatelství Lidové Noviny, Praha, 1995.

Valentin Jijkoun and Katja Hofmann. Generating a non-english subjectivity lexicon: relations that matter. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 398–405. Association for Computational Linguistics, 2009.

Mahesh Joshi and Carolyn Penstein-Rosé. Generalizing dependency features for opinion mining. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 313–316. Association for Computational Linguistics, 2009.

Renata Kamenická and Jiří Rambousek. Společensky nezávadná mluva: Praktická slovníková příručka. *Literární noviny*, pages 33–36, 1995.

Petr Karlík, Marek Nekula, and Zdenka Rusínová. *Příruční mluvnice češtiny.* Nakladatelství Lidové Noviny, 1996.

Petr Karlík, Marek Nekula, and Jana Pleskalová. *Encyklopedický slovník češtiny.* Nakladatelství Lidové Noviny, 2002.

Soo-Min Kim and Eduard Hovy. Determining the sentiment of opinions. In *Proceedings of the 20th International Conference on Computational Linguistics*, pages 1367–1374. Association for Computational Linguistics, 2004.

Soo-Min Kim and Eduard Hovy. Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, pages 1–8. Association for Computational Linguistics, 2006.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

Manfred Klenner, Simon Clematide, Stefanos Petrakis, and Marc Luder. Compositional syntax-based phrase-level polarity annotation for german. *Linguistic Issues in Language Technology*, 7(1), 2012.

Nozomi Kobayashi, Kentaro Inui, and Yuji Matsumoto. Extracting aspect-evaluation and aspect-of relations in opinion mining. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 1065–1074, 2007.

Marie Krčmová. *Běžně mluvený jazyk v Brně*, volume 231. Univerzita J. E. Purkyně, 1981.

Ivona Kučerová. K subjekt-predikátové shodě v češtině. Master's thesis, Univerzita Karlova, 2000.

Akshi Kumar and Ritu Rani. Sentiment analysis using neural network. In *Next Generation Computing Technologies (NGCT), 2016 2nd International Conference on*, pages 262–267. IEEE, 2016.

Yeha Lee, Jungi Kim, and Jong-Hyeok Lee. Extracting domain-dependent semantic orientations of latent variables for sentiment classification. In *Computer Processing of Oriental Languages. Language Technology for the Knowledge-based Economy*, pages 201–212. Springer, 2009.

Ladislav Lenc and Tomás Hercig. Neural networks for sentiment analysis in czech. In *ITAT*, pages 48–55, 2016.

Vladimir I Levenshtein. Binary codes capable of correcting deletions, insertions and reversals. In *Soviet Physics Doklady*, volume 10, pages 707–710, 1966.

Beth Levin. *English verb classes and alternations: A preliminary investigation*. University of Chicago Press, 1993.

Bing Liu. *Web data mining: Exploring hyperlinks, contents, and usage data*. Springer Science & Business Media, 2007.

Bing Liu. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167, 2012.

Markéta Lopatková, Zdeněk Žabokrtský, Václava Kettnerová, Karolina Skwarska, Eduard Bejček, Klára Hrstková, Michaela Nová, and Miroslav Tichý. VALLEX 2.5 – Valency lexicon of Czech verbs, version 2.5. *Software prototype*, 2007.

Jean-Pierre Malrieu. *Evaluative semantics: Cognition, language and ideology*. Routledge, 2002.

J. Martin and P. White. *The Language of Evaluation: Appraisal in English*. Palgrave Macmillan, New York, 2005.

Vilém Mathesius. *A functional analysis of present day English on a general linguistic basis*. Walter de Gruyter, 1975.

Yvette Yannick Mathieu, Christiane Fellbaum, et al. Verbs of emotion in French and English. In *The 5th International Conference of the Global WordNet Association (GWC-2010)*, 2010.

Yoshiko Matsumoto. Variations in Japanese honorification—deviations or a change in the making. *Constructional reorganization*, pages 89–104, 2008.

Ryan McDonald, Fernando Pereira, Kiril Ribarov, and Jan Hajič. Non-projective dependency parsing using spanning tree algorithms. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 523–530. Association for Computational Linguistics, 2005.

Arun Meena and TV Prabhakar. *Sentence level sentiment analysis in the presence of conjuncts using linguistic analysis*. Springer, 2007.

Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai. Topic sentiment mixture: modeling facets and opinions in weblogs. In *Proceedings of the 16th international conference on World Wide Web*, pages 171–180. ACM, 2007.

Anna Mikulová. Emocionalita, expresivita a hodnocení-integrace v jazycích a jazykovědných koncepcích. In *Integrace v jazycích, jazyky v integraci*, pages 113–133. Nakladatelství Lidové Noviny, Praha, 2010.

Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolářová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, et al. Annotation on the tectogrammatical level in the Prague dependency treebank. *Annotation manual. Tech. Report*, 30, 2006.

George Miller and Christiane Fellbaum. *WordNet: An electronic lexical database*. MIT Press Cambridge, 1998.

Jozef Mistrík. *Štylistika*. Slovenské pedagogické nakladatel'stvo, 1985.

Lucie Mladová. Od hloubkové struktury věty k diskurzním vztahům (diskurzní vztahy v češtině a jejich zachycení v anotovaném korpusu). Master's thesis, Univerzita Karlova, 2008.

Karo Moilanen and Stephen Pulman. Sentiment composition. In *Proceedings of the Recent Advances in Natural Language Processing International Conference*, pages 378–382, 2007.

Ramanathan Narayanan, Bing Liu, and Alok Choudhary. Sentiment analysis of conditional sentences. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 180–189. Association for Computational Linguistics, 2009.

Jiří Nekvapil. Úvodem k monotematickému číslu „Analýza promluv a textů, analýza diskurzu". *Sociologický časopis/Czech Sociological Review*, 42(02):263–267, 2006.

Karel Jaroslav Obrátil. *Velký slovník sprostých slov*. Lege Artis, 2000.

Patrik Ouředník. *Šmírbuch jazyka českého: slovník nekonvenční češtiny, 1945-1989*. Paseka, 2005.

Karel Pala and P. Ševeček. Czech WordNet. Final Report Version 1, June 29, 1999, 1999.

Jarmila Panevová. *Formy a funkce ve stavbě české věty*, volume 13. Academia, Praha, 1980.

Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical Methods in Natural Language Processing*, pages 79–86. Association for Computational Linguistics, 2002.

Michał B. Paradowski and Łukasz Jonak. Understanding the social cascading of geekspeak and the upshots for social cognitive systems. In *Understanding and Modelling Collective Phenomena*, pages 27–32. University of Birmingham, 2011.

Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. Understanding the exploding gradient problem. *CoRR*, abs/1211.5063, 2012.

Soňa Pazderová. Subjektivita v novinovém zpravodajství: Evaluativní jazyk v případě únosu Burtonových. Master's thesis, Masarykova univerzita, 2007.

Veronica Perez-Rosas, Carmen Banea, and Rada Mihalcea. Learning sentiment lexicons in spanish. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 3077–3081, 2012.

Vladimír Petkevič. Reliable morphological disambiguation of Czech: A rule-based approach is necessary. *Insight into the Slovak and Czech Corpus Linguistics*, pages 26–44, 2006.

Lucie Poláková, Jiří Mírovský, Anna Nedoluzhko, Pavlína Jínová, Šárka Zikánová, and Eva Hajičová. Introducing the Prague Discourse Treebank 1.0. In *Proceedings of the 6th international joint conference on natural language processing*, pages 91–99, 2013.

Maria Pontiki, Haris Papageorgiou, Dimitrios Galanis, Ion Androutsopoulos, John Pavlopoulos, and Suresh Manandhar. Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, 2014.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammed AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Veronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Núria Bel, Salud Maria Jiménez-Zafra, and Gülşen Eryiğit. Semeval-2016 task 5 : aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30. Association for Computational Linguistics, 2016. ISBN 978-1-941643-95-2.

Martin Popel and Zdeněk Žabokrtský. TectoMT: modular NLP framework. In *Advances in natural language processing*, pages 293–304. Springer, 2010.

Ana-Maria Popescu and Orena Etzioni. Extracting product features and opinions from reviews. In *Natural language processing and text mining*, pages 9–28. Springer, 2007.

Katrin Přikrylová, Vladislav Kuboň, and Kateřina Veselovská. The role of conjunctions in adjective polarity analysis in czech. *Computación y Sistemas*, 20(3):377–386, 2016a.

Katrin Přikrylová, Vladislav Kuboň, and Kateřina Veselovská. Logical vs. natural language conjunctions in czech: A comparative study. In *ITAT*, pages 68–73, 2016b.

Robert R. Provine, Robert J. Spencer, and Darcy L. Mandell. Emotional expression online emoticons punctuate website text messages. *Journal of Language and Social Psychology*, 26(3):299–307, 2007.

Tomáš Ptáček, Ivan Habernal, and Jun Hong. Sarcasm detection on czech and english twitter. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers.*, pages 213–223, 2014.

Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. Opinion word expansion and target extraction through double propagation. *Computational linguistics*, 37(1):9–27, 2011.

Jonathon Read. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL Student Research Workshop*, pages 43–48. Association for Computational Linguistics, 2005.

Antonio Reyes, Paolo Rosso, and Davide Buscaldi. From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering*, 74:1–12, 2012.

Ellen Riloff and Janyce Wiebe. Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 105–112. Association for Computational Linguistics, 2003.

Eleanor Rosch. Cognitive representations of semantic categories. *Journal of experimental psychology: General*, 104(3):192, 1975.

Josef Ruppenhofer, Michael Ellsworth, Miriam RL Petruck, Christopher R Johnson, and Jan Scheffczyk. *FrameNet II: Extended theory and practice*. International Computer Science Institute, Berkeley, California, 2006.

Josef Ruppenhofer, Swapna Somasundaran, and Janyce Wiebe. Finding the Sources and Targets of Subjective Expressions. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pages 2781–2788, 2008.

Magdaléna Rysová. Alternative lexicalizations of discourse connectives in Czech. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 2800–2807, 2012.

Christopher Scaffidi, Kevin Bierhoff, Eric Chang, Mikhael Felker, Herman Ng, and Chun Jin. Red opal: product-feature scoring from reviews. In *Proceedings of the 8th ACM conference on Electronic commerce*, pages 182–191. ACM, 2007.

Klaus R. Scherer. Appraisal considered as a process of multilevel sequential checking. In *Appraisal processes in emotion: Theory, methods, research*, pages 92–120. Oxford University Press, 2001.

Karin Kipper Schuler. *VerbNet: A broad-coverage, comprehensive verb lexicon*. PhD thesis, University of Pennsylvania, 2005.

John R. Searle. *Speech acts: An essay in the philosophy of language*, volume 626. Cambridge University Press, 1969.

Petr Sgall. *Generativní popis jazyka a česká deklinace*. Academia, Praha, 1967.

Petr Sgall, Ladislav Nebeský, Alla Goralčíková, and Eva Hajičová. *A functional approach to syntax in generative description of language*. Elsevier Pub. Co., 1969.

Petr Sgall, Eva Hajičová, and Jarmila Panevová. *The meaning of the sentence in its semantic and pragmatic aspects*. Springer Science & Business Media, 1986.

Cecílie Šimáčková. *Sedmijazyčný slovník vulgarizmů*. Nakladatelství PLOT, 2009.

Jana Šindlerová and Kateřina Veselovská. Building a corpus of evaluative sentences in multiple domains. In *Corpus Linguistics 2013 – Abstract Book*, pages 273–275, Lancaster, UK, 2013. UCREL.

Jana Šindlerová, Kateřina Veselovská, and Jan Hajič Jr. Tracing sentiments: Syntactic and semantic features in a subjectivity lexicon. In *Proceedings of the XVI EURALEX International Congress: The User in Focus*, pages 405–414, Bolzano/Bozen, Italy, 2014. EURAC research.

Pavel Smrž. Automatic acquisition of semantics-extraction patterns. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06)*, pages 271–274, 2006a.

Pavel Smrž. Using WordNet for opinion mining. In *Proceedings of the Third International WordNet Conference*, pages 333–335. Masaryk University, 2006b.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642, 2013.

Swapna Somasundaran, Janyce Wiebe, and Josef Ruppenhofer. Discourse level opinion interpretation. In *Proceedings of the 22nd International Conference on Computational Linguistics*, pages 801–808. Association for Computational Linguistics, 2008.

Swapna Somasundaran, Galileo Namata, Janyce Wiebe, and Lise Getoor. Supervised and unsupervised methods in employing discourse relations for improving opinion polarity classification. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 170–179. Association for Computational Linguistics, 2009.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15 (1):1929–1958, January 2014. ISSN 1532-4435.

Josef Steinberger, Polina Lenkova, Mijail Alexandrov Kabadjov, Ralf Steinberger, and Erik Van der Goot. Multilingual entity-centered sentiment analysis evaluated by parallel corpora. In *Proceedings of the 8th International Conference Recent Advances in Natural Language Processing (RANLP*, pages 770–775, 2011.

Josef Steinberger, Mohamed Ebrahim, Maud Ehrmann, Ali Hurriyetoglu, Mijail Kabadjov, Polina Lenkova, Ralf Steinberger, Hristo Tanev, Silvia Vázquez, and Vanni Zavarella. Creating sentiment dictionaries via triangulation. *Decision Support Systems*, 53(4):689–694, 2012.

Josef Steinberger, Tomáš Brychcín, and Michal Konkol. Aspect-level sentiment analysis in Czech. In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 24–31, 2014.

Andreas Stolcke et al. Srilm-an extensible language modeling toolkit. In *INTERSPEECH*, pages 901–904, 2002.

Philip J Stone, Dexter C Dunphy, and Marshall S Smith. *The General Inquirer: A Computer Approach to Content Analysis.* MIT press, 1966.

Veselin Stoyanov and Claire Cardie. Topic identification for fine-grained opinion analysis. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 817–824. Association for Computational Linguistics, 2008.

Carlo Strapparava and Rada Mihalcea. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 70–74. Association for Computational Linguistics, 2007.

Gregory T. Stump. How peculiar is evaluative morphology? *Journal of linguistics*, 29(01):1–36, 1993.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307, 2011.

Aleš Tamchyna and Kateřina Veselovská. UFAL at semeval-2016 task 5: Recurrent neural networks for sentence classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: International Workshop on Semantic Evaluation (SemEval)*, pages 367–371, Stroudsburg, PA, USA, 2016. ACL, Association for Computational Linguistics (ACL). ISBN 978-1-941643-95-2.

Marina Terkourafi. Don't go V-ing in Cypriot Greek: Semantic, pragmatic, and prosodic aspects of a prohibitive construction. *Constructions and Frames*, 2(2):208–241, 2010.

Tun Thura Thet, Jin-Cheon Na, Christopher SG Khoo, and Subbaraj Shakthikumar. Sentiment analysis of movie reviews on discussion boards using a linguistic approach. In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pages 81–84. ACM, 2009.

Zhiqiang Toh and Jian Su. Nlangp: Supervised machine learning system for aspect category classification and opinion target extraction. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 496–501, Denver, Colorado, June 2015. Association for Computational Linguistics.

Peter D. Turney. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 417–424, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics. doi: 10.3115/1073083.1073153.

Zdeňka Urešová. *Valence sloves v Pražském závislostním korpusu*. Ústav formální a aplikované lingvistiky, Univerzita Karlova, Praha, 2011.

Kateřina Veselovská. Členská negace a způsob jejího vyjadřování v současné češtině. Master's thesis, Univerzita Karlova, 2010.

Kateřina Veselovská. *Členská negace a způsoby jejího vyjadřování v datech ČNK a PDT 2.0*, pages 49–61. Nakladatelství Lidové Noviny, Praha, 2011a. ISBN 978-80-7422-116-3.

Kateřina Veselovská. Sentence-level polarity detection in a computer corpus. In *WDS'11 Proceedings of Contributed Papers, Part I*, pages 167–170, Praha, 2011b. Matfyzpress. ISBN 978-80-7378-184-2.

Kateřina Veselovská. Sentence-level sentiment analysis in Czech. In *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics*, pages 65–69. ACM, 2012.

Kateřina Veselovská. Czech Subjectivity Lexicon: A Lexical Resource for Czech Polarity Classification. In *Proceedings of the Seventh International Conference Slovko 2013; Natural Language Processing, Corpus Linguistics, E-learning*, pages 279–284, Lüdenscheid, Germany, 2013. RAM-Verlag. ISBN 978-3-942303-18-7.

Kateřina Veselovská. Fear and trembling: Annotating emotions in czech holocaust testimonies. In *Proceedings of the 5th International Workshop on Emotion, Social Signals, Sentiment & Linked Open Data*, pages 41–45, Reykjavík, Iceland, 2014a. Harpa Centre, ELRA.

Kateřina Veselovská. On the linguistic structure of evaluative meaning. In *Language Use and Linguistic Structure*, pages 269–277, Olomouc, 2014b. Univerzita Palackého. ISBN 978-80-244-4060-6.

Kateřina Veselovská and Jan Hajič Jr. Why words alone are not enough: Error analysis of lexicon-based polarity classifier for Czech. In *Sixth International Joint Conference on Natural Language Processing – Proceedings of the 3rd Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2013)*, pages 1–5, 2013.

Kateřina Veselovská and Aleš Tamchyna. ÚFAL: Using hand-crafted rules in aspect based sentiment analysis on parsed data. In *Proceedings of the Eighth International Workshop on Semantic Evaluation (SemEval 2014)*, pages 694–698. Dublin City University, 2014. ISBN 978-1-937284-96-1.

Kateřina Veselovská, Jan Hajič Jr., and Jana Šindlerová. Creating annotated resources for polarity classification in Czech. In *Empirical Methods in Natural Language Processing: Proceedings of the Conference on Natural Language Processing (KONVENS 2012)*, pages 296–304, Wien, 2012. ÖGAI.

Kateřina Veselovská, Jr. Jan Hajič, and Jana Šindlerová. Subjectivity lexicon for Czech: Implementation and improvements. *Journal for Language Technology and Computational Linguistics*, 29(1):47–61, 2014. ISSN 2190-6858.

Khuong Vo, Dang Pham, Mao Nguyen, Trung Mai, and Tho Quan. Combination of domain knowledge and deep learning for sentiment analysis. In *International Workshop on Multi-disciplinary Trends in Artificial Intelligence*, pages 162–173. Springer, 2017.

Adam Westerski. Sentiment analysis: Introduction and the state of the art overview. *Universidad Politecnica de Madrid, Spain*, pages 211–218, 2007.

Janyce Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. Learning subjective language. *Computational linguistics*, 30(3):277–308, 2004.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2-3):165–210, 2005.

Michael Wiegand and Dietrich Klakow. The role of knowledge-based features in polarity classification at sentence level. In *Proceedings of the 22nd International FLAIRS Conference.*, 2009.

Michael Wiegand, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, and Andrés Montoyo. A survey on the role of negation in sentiment analysis. In *Proceedings of the workshop on negation and speculation in natural language processing*, pages 60–68. Association for Computational Linguistics, 2010.

Theresa Wilson, Paul Hoffmann, Swapna Somasundaran, Jason Kessler, Janyce Wiebe, Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. Opinionfinder: A system for subjectivity analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 34–35. Association for Computational Linguistics, 2005a.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 347–354. Association for Computational Linguistics, 2005b.

Theresa Ann Wilson. *Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states*. PhD thesis, University of Pittsburgh, 2008.

Suwon Yoon. *'Not'in the mood: the syntax, semantics, and pragmatics of evaluative negation*. PhD thesis, University of Chicago, 2011.

Jaroslav Zima. *Expresivita slova v současné češtině: studie lexikologická a stylistická*. Nakl. Československé akademie věd, 1961.

Jan Žižka and František Dařena. Automatic sentiment analysis using the textual pattern content similarity in natural language. In *Text, Speech and Dialogue*, pages 224–231. Springer, 2010.