

# The Role of Conjunctions in Adjective Polarity Analysis in Czech

Katrin Prikrylová, Vladislav Kuboň, Kateřina Veselovská

Charles University in Prague, Faculty of Mathematics and Physics,  
Czech Republic

{prikrylova,vk,veselovska}@ufal.mff.cuni.cz

**Abstract.** Adjectives very often determine the polarity of an utterance. This paper studies the role of conjunctions in the analysis of adjective polarity. The study is being performed on Czech language on the basis of an existing algorithm for English. The algorithm has been modified in order to reflect the differences between the two typologically different languages. The results of the original and modified algorithm are being compared and discussed. The paper also contains a thorough discussion of exceptions and special cases supported by a number of examples from a large corpus of Czech.

**Keywords.** Adjective polarity, conjunctions, Czech language.

## 1 Introduction

Adjectives constitute a very important factor in sentiment analysis, i.e. automatic extraction of emotions from text (see [3]). They play an important role mainly in polarity detection. The aim of polarity detection is to define whether the opinion expressed in a text is positive or negative. The influence of adjectives on overall sentence polarity is very often decisive, as we can see it in the following example:

*He was a good father.*

*He was a bad father.*

As also emphasized e.g. by [7], if we know the polarity of an adjective, we can more easily determine the overall polarity of a sentence. However, things are not that easy as it might seem from the previous example. One problematic issue is the fact that very often the polarity of an adjective depends on a context. The same adjective may have different polarity depending on

different circumstances. Another important issue concerns the complexity of a sentence – very often the sentence of which polarity we want to analyze contains more than a single adjective. In these cases, the adjectives as such, the phrases the adjectives are part of or even the whole clauses are connected by means of punctuation marks or conjunctions.

These issues call for the development of a set of rules or an algorithm which would determine the polarity of an individual adjective in a given context. In the following step, it would analyze the polarity of a sentence or of an expression based on the polarity of individual adjectives and their combination. Some experiments in this direction have been already performed for English (see [8, 9]). The article most corresponding to the research direction we want to take is the paper of Hatzivassiloglou and McKeown [1], in which the authors determine the polarity of English adjectives by means of conjunctions. The main goal of this paper is to adapt their method to a typologically different language, namely Czech.

### 1.1 The Method of Hatzivassiloglou and McKeown

In the paper [1], the authors aimed at determining the adjective polarity without taking context into account. The output of their experiments consisted of sets of labeled adjectives. They have observed that for English, in almost 78% of the cases when two adjectives are connected by conjunctions **and** and **or**, the expression connects two adjectives of the same polarity. However, this rule does not hold for a conjunction **but**.

Based on the above mentioned observations, Hatzivassiloglou and McKeown created a set of the following rules:

1. If the adjectives are connected by a conjunction **but**, let's suppose that they have the opposite polarity.
2. If the adjectives are morphologically related, then let's suppose that they have the opposite polarity.
3. In all other cases, let's suppose that the connected adjectives have the same polarity.

The application of these simple rules led to an overall precision of almost 82%, while the second and third rule had only minor influence on the result, improving it only marginally by a couple of percent.

## 2 Adaptation of the Method for Czech

In this section, we would like to investigate the differences between English and a typologically different language, Czech. The main differences (apart from the lexical differences where a translated adjective may have a slightly different polarity) between the two languages which might influence the polarity are the fleective nature and a relatively free word order of Czech language, as it has been indicated e.g. in [5].

### 2.1 Data

Two main sources of the data have been exploited in our experiment, namely the corpus SYN2010 [2] and a Czech Subjectivity Lexicon [6]. SYN2010 is a large scale general corpus of Czech annotated at a morphological level. We have collected all relevant sentences from the corpus, namely those which contained positive and/or negative adjectives connected by a relevant conjunction.

### 2.2 Conjunctions under Investigation

The paper [1] primarily concentrates on English conjunctions **and** and **but**. These conjunctions correspond to Czech words **a** and **ale**. Therefore, it is natural that we have also started our analysis with these two "core" conjunctions. Moreover, we have added the conjunction **nebo** (corresponding to English **or**)<sup>1</sup>, in order to get a complete picture of the two most frequently used logical conjunctions. The rules for the latter had to be modified for Czech due to grammatical rules which handle the use of this conjunction.

As a starting hypothesis we have adopted the same basic rules which were proposed for English:

1. The conjunction **a** (and) connects two expressions of the same polarity.
2. The conjunction **ale** (but) connects two expressions of opposite polarity.
3. The conjunction **nebo** (or) connects two expressions of the same polarity, if it is not preceded by a comma. When preceded by a comma, it connects two expressions of the opposite polarity.<sup>2</sup>

Let us now further investigate particular conjunctions and discuss the observations discovered in the course of the analysis of corpus data.

### 2.3 Conjunction **a** (and)

If the conjunction **a** (and) connected two positive or two negative adjectives, it was in a vast majority of cases (96.7 %) a connection which could be analyzed as a normal conjunction, i.e. the expected behavior of the conjunction. For example:

*Tvůj otec mě obdařil nádherným darem – tebou – krásnou a dokonalou dcerou.* (Your father gave

<sup>1</sup>The conjunction **or** is also briefly mentioned in the paper [1] with the remark that it connects two adjectives of the same polarity.

<sup>2</sup>The reason for this distinction is the fact that the conjunction **nebo** (or) can be used as both the disjunction (in these cases, it connects two expressions of the same polarity) and exclusive disjunction (when it should connect two expressions of the opposite polarity). It is the presence of comma which according to the grammatical rules of standard Czech should make a difference between a simple and exclusive disjunction.

**Table 1.** The conjunction **a** (and) and the polarity of adjectives

Polarity	Sentences	% ratio
positive and positive	6 021	61 %
negative and negative	2 454	25 %
positive and negative	1 351	14 %
Total	9 826	100 %

me a wonderful gift – you – **a beautiful and perfect** daughter.)

The remaining cases (in which the conjunction **a** (and) behaved differently) included the following features:

1. Particular parts of the expression belonged to the different syntactic levels: although the two adjectives are located next to each other in the sentence and the conjunction connects them, they are syntactically bound to different expressions and the conjunction thus does not actually connect them. Let us consider the following example:  
*Přesto byly chvíle, kdy se mezi ostatními dívkami cítila osamělá, **odlišná a zvláštním** způsobem cizí.*  
(Even though there were moments when she felt lonely, **different and in a certain** way foreign among other girls.)
2. The conjunction connected two expressions, which in fact divided some reality into two parts – for example:  
*Je Polsko rozdělené na **sociální a liberální** část, jak o tom hovořili v kampani oba kandidáti?*  
(Is Poland divided into a **social and liberal** part, as it was claimed by both candidates in their campaigns?)  
In these cases, even though individual adjectives may have some polarity, the whole sentence is not evaluative. Such sentences are, unfortunately, very difficult to identify automatically.
3. The conjunction **a** (and) actually played a role of a disjunction. However, as we are going to show in the subsequent sections, this configuration does not require any special

treatment because in case of a disjunction using the conjunction **nebo** (or) we will also require identical polarity of adjectives. It was quite interesting that in our corpus this case involved solely adjectives in plural, as for example in the sentence:

*Někdy jsem měla pocit, že chci, aby už se něco stalo, a dokonce i **strašlivé a tragické** věci mi připadaly lepší než nic.*

(Sometimes I had a feeling that I want something to happen, even **horrible and tragic** things seemed to me better than nothing.)

An interesting, although also very infrequent phenomenon (0.6%) has been the gradation of the identical adjective, for example:

*Pokaždé, když se dostal do toho začarovaného kruhu, byly vyjeté koleje **hlubší a hlubší**.* (Every time, when he got into that vicious circle, the beaten tracks were **deeper and deeper**.)

Even more interesting situation is observed in sentences in which a positive adjective is connected with a negative one – these are in fact the counterexamples to our presumption. This happened in 14% of cases.

If we search for a connection between a really negative and a really positive adjective (in a given context), we realize that in 9% of cases the use of a pair with opposite polarity is indicated lexically, by means of words like **zároveň** (at the same time), **současně** (simultaneously) or **mezi** (between) (for example: *Na první pohled byla měra **krásná a hrůzná** zároveň.* (On the first sight, the moth was **beautiful and horrifying** at the same time.))

Let us also mention that in the sentences containing the word **mezi** (between), the adjectives<sup>3</sup> are only used in metacontext, i.e. are not intentionally evaluative<sup>4</sup>. From the point of view of sentiment analysis it is important that the words which are only mentioned do not take part in the evaluation of sentence polarity. They are present only as auxiliary expressions describing certain reality which is further discussed subsequently. The adjectives appearing in phrases with **mezi** (between) were also in majority of cases opposites

<sup>3</sup>Or the complete noun phrases containing them.

<sup>4</sup>More about this distinction can be found in [4]

– speakers wanted to point out a contrast between some conditions in reality.<sup>5</sup>

## 2.4 Conjunction **ale** (but)

Our initial hypothesis concerning the conjunction **ale** (but) supposed that this conjunction connects two expressions (adjectives) of the opposite polarity. Unfortunately, there were 37% of sentences in which this conjunction connected adjectives of the same polarity, thus contradicting our hypothesis.

**Table 2.** Combination of adjectives by means of the conjunction **ale** (but)

Polarity	Sentences	% ratio
positive and positive	109	28 %
negative and negative	37	9 %
positive and negative	244	63 %
Total	390	100 %

### 2.4.1 The conjunction **ale** (but) and the adjectives of opposite polarity

Firstly, let us survey the connections which conform to our hypothesis. Roughly one tenth of the cases contained a connection at different syntactic levels, as, e.g., *Levou pěst si tiskl k rameni, jako by mu takovíhle šupáci byli **odporní, ale pravou** napřáhoval k seznamovacímu pozdravu.*<sup>6</sup> (He pushed his left fist to his shoulder, as if he these bastards were **loathsome**, but he was stretching the **right** one in the greeting.)

In less than 10% cases one of the adjectives has been negated, as, e.g., in the sentence: *Není to sport **agresivní, ale čestný.*** (It is not **an aggressive** sport, but **a honest** one.) These sentences usually contain some kind of correction – the speaker rejects the first adjective and tries to persuade the listeners that the reality is exactly opposite. These sentences will be discussed in

<sup>5</sup>Similar behavior can also be attributed to other words, as, e.g., *rozdíl* (the difference), *rozdělit* (to divide), *rozlišit* (to distinguish), *střídavě* (alternately), *existovat* (to exist), and also the word *jako* (like, as).

<sup>6</sup>Let us mention that in this example we are able to recognize a different syntactic level on the basis of a difference in the gender, number or case of both adjectives.

more detail below. For the moment, it is important that they actually represent a minor issue.

Let us suppose that we are able to recognize the connection at a different syntactic level. In that case, we would find out that the remaining 88% of sentences actually conforms to our hypothesis, i.e. they combine positive and negative adjective without additional problems or conditions.

### 2.4.2 The conjunction **ale** (but) and the adjectives of the same polarity

The combination of the two adjectives of the same polarity seems to be much more interesting. These are actually the combinations which go against our hypothesis. Similarly as in the previous case, a substantial part of sentences (39%) contains combinations at different syntactic levels, as, e.g.: *Kariéra, zpívání je **nádherné, ale skutečný** život je prioritá.* (The career, singing is **beautiful**, but a **real** life has a priority.)

The constructions, in which the combination has been lexically negated, therefore it actually contained two expressions of opposite polarity, constituted one fifth of all cases, for example: *Nevyhrál **lepší, ale šťastnější** tým.* (**Not the better, but the happier** team won.) Let us mention that these cases are actually, due to the application of negation, conforming to our hypothesis.

The rest, i.e. slightly more than a half of cases, consisted of connections which actually contradicted our hypothesis. Let us look at the example: *Je to **vídné, přesto naléhavé, je to zábavné, ale poučné, je to cíleně dojemné, ale bez klišé.*** (It is kind, but still urgent, it is **entertaining, but educational** as well, it is intentionally touching, but without clichè.) It is important that although both adjectives have the same polarity, they are used in a certain contrast. In our example, there is the contrast between entertainment and education.

If the combination of two adjectives of the same polarity is preceded by the word **nejen** (not only), it is actually not an adversative relation, but a synthetic one. Both adjectives have the same polarity in these cases.

### 2.4.3 The conjunction **ale** (but) and a negation

In our corpus, 9% of detected combinations containing the conjunction **ale** (but) also contained a negated adjective.

Let us note that in the combination of two positive or two negative adjectives combined by a conjunction **ale** (but) it is not possible to use all possible pairs (some of them do not make sense from the semantic point of view). For example, it is impossible to say: *Byla to hezká, ale krásná žena.* (She was a **a handsome but beautiful** woman.)<sup>7</sup> or: *Prodavač byl nepoctivý, ale nepříjemný.* (The shop assistant was **dishonest, but unpleasant**)<sup>8</sup>. In the case of connections containing one negated expression, we are actually in a situation in which we are connecting two phrases of the opposite polarity. We should therefore not be surprised that it is possible to say: *Nebyla to hezká, ale krásná žena.* (She was not **a handsome, but beautiful** woman.). In this way we can create sentences independently of how semantically close or distant are the adjectives used. In our example they both describe the same property, both are considered to be positive, but they express a different intensity (of beauty, in this case). Nevertheless, it makes sense to use them in the combination with the conjunction **ale** (but) and a negation. The main reason for such construction is the fact that it actually works as some kind of refinement, correction of the speaker, rectification etc.

Even more interesting is the opposite configuration, i.e. the connection of two adjectives of opposite polarity. If we negate one of them, we actually get two phrases of the same polarity. Surprisingly, in this case we can ignore the semantic constraints applied to the combination of two adjectives of the same polarity (as they were discussed above). We can say: *Nebyla to ošklivá, ale krásná žena.* (She was not an **ugly, but beautiful** woman.) – if we negate the word *ošklivá* (ugly), it acquires positive polarity and, at

<sup>7</sup>In this construction it is not possible to create a contrast between two adjectives of the same polarity describing the same property or quality.

<sup>8</sup>Although – as we have already shown – there is a number of adjectives of the same polarity, which can be connected by the conjunction **ale** (but), as, e.g.: *Koupila si pohodlné, ale luxusní boty.* (She bought **comfortable but luxuries** shoes.)

the same time, it describes a similar property as the word *krásná* (beautiful). The situation is similar to the one discussed above, but now it does not contradict anything. The reason seems to be an implicit message, carried by the negation, namely the correction or rectification of some fact. The question, whether it is this kind of connection, can be answered positively either in the case that the conjunction **ale** (but) can be substituted by a conjunction **nýbrž** (not ... but), or in the case that it is a combination of opposites. The whole phrase then carries the polarity of the second adjective.

All these observations actually have a negative consequence for sentiment analysis. If the combination containing the conjunction **ale** (but) also contains negation, it is impossible to determine whether the adjectives should have identical or opposite polarity.

If we limit our investigation only to combinations not containing the negation, the results would be only slightly better.<sup>9</sup> Even worse, although we might consider two adjectives of the same polarity with one negation as two phrases of the opposite polarity, the improvement would also be only marginal.

### 2.5 Conjunction **nebo** (or) as a Copulative Conjunction

According to its function, conjunction **nebo** (or) can work both as a copulative or disjunctive conjunction.<sup>10</sup> When used as a copulative conjunction, it connects expressions of the same polarity. In structures with disjunctive meaning, **nebo** (or) connects expressions of the opposite polarity and is preceded by no comma. The disjunctive meaning is marked by a comma preceding **nebo** (or). Based on this convention, the data were divided into two separately analyzed parts.

Within the first part of the data representing adjectives connected exclusively by **nebo** (or) without a comma, only half of the examples complied with the similar polarity hypothesis.

<sup>9</sup>For example 65 % of combinations with opposite polarity instead of the original 63%.

<sup>10</sup>Linguistic term "copulative conjunction" here corresponds to disjunction in logic, whereas "disjunctive conjunction" corresponds to exclusive disjunction.

**Table 3.** Adjectives connected by **nebo** (or) without a comma

Polarity	Sentences	% Ratio
positive and positive	99	17 %
negative and negative	185	32 %
positive and negative	292	51 %
Total	576	100 %

### 2.5.1 Conjunction **nebo** (or) and Adjectives of the Same Polarity

If we further investigate structures from Table 3, we realize that in two thirds of the cases, the authors used conjunction **nebo** (or) to express alternative or even synonymy. Sometimes, it seemed that the author intended to correct him/herself or tried to find more suitable wording, e.g. (...) *je stejně chyba, abyste se označil za špatného nebo hrozného.* (... it is actually a mistake to mark yourself as **bad or awful**.)

### 2.5.2 Conjunction **nebo** (or) and Negation

However, it is interesting to analyze in detail a use of **nebo** (or) in connection with negation, see the following example:

- *Není to příliš zábavné nebo zajímavé ani poučné nebo inspirující divadlo.* (It is neither a very **funny or interesting nor informative or inspirational** performance.)

These structures follow the scheme  $\neg(A \vee B)$ , which can be, according to de Morgan's Laws, converted to  $\neg A \wedge \neg B$ . In such sentences, Czech language can employ conjunction **ani** (nor). When applying this conjunction, we generate a correct sentence both in terms of logic and grammar<sup>11</sup>.

Whereas conjunction **ani** (nor) can connect also phrases of the opposite polarity, for conjunction **nebo** (or) in the copulative function this possibility seems rather inappropriate.

<sup>11</sup>And by the way in terms of a **language sense** as well.

### 2.5.3 Conjunction **nebo** (or) and Adjectives of the Opposite Polarity

As we have shown at the beginning of this section, a good half of the cases does not convey with our hypothesis. In these cases, conjunction **nebo** (or) connects adjectives of the opposite polarity. However, in only one tenth of the cases it expresses either synonymy or alternative choice, as in the sentence *Totěž se stává, když jste rozrušená nebo zamilovaná.* (It also happens when you are **excited or lovesick**.) 43% of the cases represent structures influenced by the lexical indicators, namely expressions (or phrases) like **ať (už)** (no matter how), **jestli** (if) and **zda** (whether) – for particular percentage see Table 4. It is also interesting that these phenomena mostly concern opposite couples. The writer usually announces his or her own ignorance, e.g. *Zda chytřejší nebo hloupější, nevím.* (Whether **cleverer or dumber**, I don't know.). This also concerns utterances expressing an indirect question: *Sám posuďte, jestli je lepší nebo horší.* (It is up to you whether it is **better or worse**.) Often, the structures implied indifference, e.g. *Příroda je slepá a zná jen jednu cestu – kupředu – a to bez ohledu na to, jestli je dobrá nebo špatná.* (Nature is blind and it admits only one way – forward – regardless of whether it is **good or bad**.) In sentences with the conjunction **ať (už)** (no matter how) the indifferent meaning is quite clear: *Ať už je ale technologie dobrá nebo špatná, jsme uzamčeni v jejím středu.* (No matter how **good or bad** the technology is, we are locked in its center.) The important feature of these structures is that they do not contain evaluative adjectives. The adjectives are used not to describe a particular entity, but in the metacontext describing a potential situation/possible world. Similarly, yet not that unequivocally, behave also the words **že** (that) and **jako** (as).

In the rest of the cases, we found a contrast of the two adjectives, while the ratio of opposites and non-opposites was 52 to 11. That means that if we were able to distinguish opposite couples, there would remain one fifth of the cases in which it makes sense to connect adjectives of the opposite polarity by the conjunction **nebo** (or).

**Table 4.** The types of contrast in combination of opposite polarity adjectives by means of the conjunction **nebo** (without a comma)

Polarity	Sentences	% Ratio
<b>ať (už)</b>	22	24 %
<b>jestli</b>	19	21 %
<b>zda</b>	9	10 %
contrast of opposites	38	41 %
simple contrast	4	4 %
Total	92	100 %

### 2.5.4 Conjunction **nebo** (or) as a Disjunctive Conjunction

Conjunction **nebo** (or) works as a negative conjunction if it connects two sentences / two constituents and is at the same time preceded by a comma. In logic, this phenomena would be called *exclusive disjunction*. In the article [1] the authors do not take this opportunity into account.<sup>12</sup> However, we could use this feature to determine a hypothesis for **nebo** (or) as a disjunctive conjunction: we suppose that it connect adjectives of the opposite polarity.

This time, the results are clear – in 89 % of the cases conjunction **nebo** (or) connects parts of the opposite polarity. The cases with the same polarity concerned adjectives used to express contrast - even though they have the same polarity.

**Table 5.** Combination of adjectives by means of the conjunction **nebo** (or) (and a comma)

Polarity	Sentences	% ratio
positive and positive	8	7 %
negative and negative	5	4 %
positive and negative	107	89 %
Total	120	100 %

However, we noticed more interesting situation when it came to the adjectives of the opposite polarity. First of all, in only 2% of the cases **nebo** (or) was not connecting opposites, as for example in the following sentence: *Rodiče si často pletou závislost s úctou a můžou být laskaví, nebo*

<sup>12</sup>They include it for the rest of the conjunctions for which they suppose that they connect expressions of the same polarity.

*přísní, úzkostliví, nebo sebevědomí, na tom nesejde.* (Parents very often confuse addiction with respect and they can be **kind or strict, anxious or confident**, it does not matter.)

All the other cases concerned a connection including opposites, or a connection on different syntactic levels. The opposites were used in contrastive meaning mostly in questions, both direct (questions finished by a question mark) and indirect, in which we can find expressions like **jestli** (if) or **zda** (whether), see the following examples:

*Je návrh zavrženého blobu-chobotnice na Letné krásný, nebo hnusný?* (Is the drawing of the rejected blob-octopus at Letná **nice or ugly?**)  
*Nebyla jsem si úplně jistá, jestli je to dobré, nebo špatné znamení.* (I was not sure whether it is a **good or a bad sign.**)

These sentences contain no evaluation either.

## 3 Rules Suggestions

This section covers all of the rules we deduced from our observations:

If the structure contains two adjectives connected by a conjunction...

- ...**a** (and) and if there is an expression *zároveň* (as well) or *současně* (at the same time) present in a sentence, the polarity of these adjectives is opposite.
- ...**a** (and) and if there is an expression *mezi* (between), *rozdíl* (contrast), *rozdělit* (divide), *rozlišit* (distinguish), *střídavě* (alternately), *existovat* (exist) and *jako* (like) , the polarity of these adjective is opposite, but it is more likely that the structure contains no evaluation.
- ...**a** (and) and we know that the adjectives have the opposite polarity, it is more likely that the structure contains no evaluation.
- ...**a** (and) and both of the adjectives have the same lemma, they also have the same polarity and the overall structure expresses gradation.
- ...**a** (and) following comma, it expresses a connection on different syntactic levels.

- In all the other cases, conjunction **a** (and) connects two adjectives of the same polarity.
- ...**ale** (but), one of them is negated by a lexical negation and the structure concerns opposites, the overall structure expresses the polarity of the second adjective.
- ...**ale** (but) and one of them is negated by a lexical negation, it is more likely that given structure expresses the opposite polarity.
- ...**ale** (but) and the whole structure is preceded by the expression **nejen** (neither), both of the adjectives have the same polarity.
- In all the other cases, it is more likely that conjunction **ale** (but) connects two adjectives of the opposite polarity.
- ...**nebo** (or) as a copulative conjunction and if the sentence contains negation, it is more likely that it connects adjectives of the same polarity.
- ...**nebo** (or) as a copulative conjunction and the sentence contains following expression: **jestli** (if), **zda** (whether), **-li** (in case), **ať (už)** (no matter how), the adjectives have the opposite polarity. However, they are not part of the evaluation.
- ...**nebo** (or) in a copulative conjunction and if we know they are opposites, it is more likely that the structure contains no evaluation.
- In all the other cases, conjunction **nebo** (or) used as a copulative conjunction connects adjectives of the same polarity.
- ...**nebo** used as a disjunctive conjunction and if the sentence contains expressions **jestli** (if), **zda** (whether), **ať (už)** (no matter how), the adjectives have the opposite polarity (and they are more likely to be opposites). However, the sentence contains no evaluation.
- ...**nebo** used as a disjunctive conjunction in question, the adjectives have the opposite polarity (and they are more likely to be opposites). However, the sentence contains no evaluation – the speaker only asks for the opinion.

- In all the other cases, conjunction **nebo** (or) used as a disjunctive conjunction connects adjectives of the opposite polarity.

## 4 Evaluation

To get a result which would represent all the rules, we chose 150 random corpus sentences containing one of our conjunctions connecting two adjectives (at least one of which was found in the subjectivity lexicon). By this selection, we got a mixture of sentences in which we take into account the frequency of conjunction in a corpus (apart from the approach in which we would use the average results for particular conjunctions).

In total, we correctly identified 79 % of cases by the rules.<sup>13</sup>

Although it may seem that the results reported in [1] are better (Hatzivassiloglou and McKeown reached almost 82% accuracy by applying all their rules), it is necessary to take into account the different nature of Czech. As a fleective language with high degree of word order freedom it contains phenomena which are more difficult to handle by natural language processing and thus the tools and methods developed for Czech typically perform by a couple of percent worse than those developed for English, regardless whether they deal with morphology (taggers) or syntax (parsers etc.). This claim may also be supported by the data introduced in this paper - while Hatzivassiloglou and McKeown started from the baseline of 78%, our baselines for individual conjunctions were much lower: 73% for **a** (and), 60% for **ale** (but), 51% for **nebo** (or) without a comma. Only **nebo** (or) preceded by a comma gives in Czech slightly higher baseline, probably due to the strictness of grammatical rules which govern its use. The baseline in this case reached 82%, thus increasing the average baseline precision for all four conjunctions to 66.5%. This means that our algorithm has increased the overall baseline precision by 11.5% compared to 4% mentioned in [1].

<sup>13</sup>Which by the way reflect the results for particular conjunctions – the average mean of the percents makes 78%.



## 5 Conclusions

The experiment described in this paper actually brought several important findings. First of all, it has showed that determining the polarity of an expression containing adjectives connected by conjunctions is more difficult for a flective language with high degree of word order freedom than for English. This fact has been reflected by the complexity of rules which we had to develop in order to obtain a substantial increase of precision against the baseline. The set of rules, which led to a precision of 79% on a random sample of Czech sentences, is a second important result of this experiment.

The third important result of our experiment was the discovery that the hand-crafted rules do not identify only the polarity of adjectives, but they can also predict the evaluative character of the whole structure. Moreover, we emphasize that these rules were created only based on observations of the conjunctions connecting two adjectives. However, we suppose that they will not behave differently when connecting other constituents (or even phrases or the whole sentences). This hypothesis will definitely be investigated in the future.

## Acknowledgments

This research was supported by the grant GA15-06894S of the Grant Agency of the Czech Republic and by the SVV project number 260 224. This work has been using language resources stored and/or distributed by the LINDAT-Clarín project of MŠMT (project LM2010013).

## References

1. **Hatzivassiloglou, V. & McKeown, K. R. (1997).** Predicting the Semantic Orientation of Adjectives. *Proceedings of the 8th conference on European chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, Stroudsburg, PA, USA.
2. **Křen, M., Bartoň, T., Cvrček, V., Hnátková, M., Jelínek, T., Koček, J., Novotná, R., Petkevič, V., Procházka, P., Schmiedtová, V., & Skoumalová, H. (2010).** *SYN2010: balanced corpus of written Czech*. LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague, Praha. Available online: <http://hdl.handle.net/11858/00-097C-0000-0023-119F-6>.
3. **Liu, B. (2012).** *Sentiment analysis and opinion mining*, volume 5. Morgan & Claypool Publishers.
4. **Marvan, T. (2010).** *Otázka významu*. Togga, Praha. In Czech.
5. **Veselovská, K. (2011).** Sentence-level polarity detection in a computer corpus. *Proceedings of the 20th Annual Conference of Doctoral Students. WDS*, pp. 167–170.
6. **Veselovská, K. & Bojar, O. (2013).** *Czech SubLex 1.0*. LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague, Praha. Available online: <http://hdl.handle.net/11858/00-097C-0000-0022-FF60-B>.
7. **Wiegand, M. & Klakow, D. (2009).** The role of knowledge-based features in polarity classification at sentence level. *FLAIRS Conference*.
8. **Wilson, T., Wiebe, J., & Hoffmann, P. (2005).** Recognizing contextual polarity in phrase-level sentiment analysis. *Proceedings of the conference on human language technology and empirical methods in natural language processing*, Association for Computational Linguistics, pp. 347–354.
9. **Wilson, T. A. (2008).** *Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states*. ProQuest.

**Katrin Příkrylová** is a PhD student in the Institute of Formal and Applied Linguistics, Charles University in Prague. Her objective are geographical named entities, the problem of theirs recognition in the text and theirs disambiguation. During her master studies she was connecting the area of logic and sentiment analysis and she carried out a research on the topic how the connectives influence the sentiment polarity of adjectives and what kind of relation is between logical connectives and connectives from natural language.

**Vladislav Kuboň** is an associate professor of the Institute of Formal and Applied Linguistics, Charles University in Prague. His research interests include syntactic analysis of natural languages, machine translation of closely related languages, sentiment analysis and formal properties of natural languages. Currently, Vladislav leads a project "Linguistic structure of evaluative meaning in Czech", funded by the Grant Agency of the Czech Republic.

**Kateřina Veselovská** is a Senior Research Associate at the Institute of Formal and Applied Linguistics, Charles University in Prague. She is mostly concerned with sentiment analysis and

polarity detection, especially in connection with the dependency treebanks and functional generative description. Previously, Kateřina Veselovská participated in the projects of Tectogrammatical description of language for automated speech retrieval and machine translation or A Multilingual Archive of Verbal Valency Characteristics. Also, she run her own project called Sentence-Level Polarity Detection in a Computer Corpus. At the moment, Dr. Veselovská is a team member of the project concerning linguistic structure of evaluative meaning in Czech.

Article received on 13/02/2016; accepted on 07/03/2016.  
Corresponding author is Katrin Příkrylová.