

AGNIESZKA PLUWAK

Institute of Slavic Studies, Polish Academy of Sciences, Warsaw, Poland  
a.pluwak@ispan.waw.pl

## TOWARDS THE APPLICATION OF SPEECH ACT THEORY TO OPINION MINING

### Abstract

The paper refers to the pragmatics' perspective on opinion mining in Polish and English, inspired by the discrepancy between the coverage of sentiment analysis and the market demand. An analysis of speech acts expressed in opinion texts reveals that almost half of all opinions include ways of indirect evaluation that might not get extracted while applying traditional methods of sentiment analysis based on direct evaluative vocabulary and polarity lexicons. Coding of sentiment with respect to speech acts could vastly broaden data mining results within an NLP-system.

**Keywords:** sentiment analysis; implicit sentiment; speech acts

## 1 Introduction

Given the lifetime achievement award at ACL 2013<sup>1</sup> Jerry Hobbes was asked about the future of computational linguistics. The famous linguist and logician replied that he sees it in projects, such as FrameNet, which he perceives as an application of the findings of modern linguistics to computer science, rather than attempts to cover language processing with purely statistical methods. Nevertheless, if so much can be achieved via mathematics on supervised and unsupervised corpora, why bother understanding the theory of linguistics and devote years to linguistic analysis? To provide an example: if a statistical parser can be built in two hours, why build a grammar rule-based one for two years with a high risk of failure and a hardly comprehensible and modifiable set of millions of rules (consulted with various computer scientists at the ACL 2013 conference; for appreciation of statistical over rule-based extraction systems see also e.g. Chiticariu, Li, & Reiss, 2013)? Will they really make a significant difference in the results of the analysis? Finally, the key question is: which linguistics findings could be applied and where could they be implemented?

It would seem that one of them is *the speech act theory* by Austin (1962), further elaborated on by Searle (1969) and other authors (e.g. Wierzbicka, etc.), and its application to sentiment analysis, or — to be more precise — data mining of opinion reviews. One of the major issues in current NLP-based sentiment analysis is the coverage of **implicit evaluation**, so far approached only on selected exemplary case studies and not yet unified within one methodology (Liu, 2012; Cambria, Schuller, Xia, & Havasi, 2013). According to Liu (2012, p. 37), the general assumption of sentiment analysis is that *no opinion is usually regarded as neutral* and even *objective sentences can imply opinions or sentiments due to desirable and undesirable facts* (Liu, 2012, p. 20). However, in

---

<sup>1</sup>The Annual Meeting of Association of Computational Linguistics held in Sofia, Bulgaria, in August 2013.

order to prove this, one has to look at opinion reviews not from the perspective of the traditional lexical approach (sentiment expressed within predefined polarized words or concepts), but rather from the point of view of **communication and stylistics (secondary orality of written texts**, see Ong, 1982) and pragmatics (**speech acts expressed in opinions**, e.g. *advice, hint, discouragement*, as in *I'm never going back there!* etc.).

This entails the perception of opinion reviews as **written-down spoken language**, which implies the reception and analysis of the speakers' discursive intentions, their description with indications of syntactic and semantic elements (constructions) and, finally, their rule-based coding (deep syntactic parsing with semantic elements, rather than key word or statistical analysis).

Speech acts cover the inferred evaluative content of opinion reviews, thus expanding the range and coverage of traditional sentiment analysis by e.g. catering for valuable hints for product improvement expressed without explicitly polarized vocabulary. As we shall see below, **implicit content** is included in at least **half of the analyzed texts**. As a result, an opposition between *direct or explicit sentiment analysis* (e.g. *delicious cheese*) and both *direct and indirect data mining from opinion texts* (*delicious cheese + I wanna have their blueberry pie again!*) might arise, the latter being a wider term and incorporating the former.

Such an approach also seems to meet the market demand for the analysis of opinion reviews (what is the customer's opinion of our company and products, expressed in a variety of ways?) by taking into consideration content which has until now been beyond the scope of statistical NLP-systems. To provide just one example, customer recommendations, such as *Bring back the old version (of this application)!*, are not an expression of direct polarization, such as *great place, nice atmosphere*, but through an inference of the illocutionary (or even perlocutionary) act of advice or demand convey both: 1) the customer's discontent with the new version and 2) preference for the previous version — both of which are valuable items of information and an expression of a negative evaluation of the new product.

Such acts of speech might seem to involve indirect (inferred from world knowledge) polarization as opposed to positive or negative adjectival phrases, but within pragmatics they are known as *direct speech acts* (direct acts of expression of e.g. a *warning* or *demand*) in contrast to e.g. *It's hot in here* posing an *indirect* hint for the addressee to open the window (Searle, 1969; Grzegorzczkova, 2010; Tokarski, 2014). Therefore, there seems to be a need to understand *opinions and evaluation* as expressions of *attitude* or *positive and negative feelings* in a broader sense than *sentiment as polarization of parts of speech*, especially since identification of adequate constructions (grammatical patterns used to express given speech acts) enable coding of such information in opinion reviews.

This paper includes: 1) The available approaches to sentiment analysis and opinion studies 2) The pragmatic' perspective on opinion reviews in English and Polish and an analysis of speech acts expressed in opinion texts 3) Examples of its application to NLP-systems in a rule-based fashion.

## 2 Available Approaches to Sentiment Analysis and Opinion Studies

According to S. Somasundaran (2010, p. iii):

*(...) opinion analysis deals with subjective phenomena such as judgments, evaluations, feelings, emotions, beliefs and stances. The availability of public opinion over the Internet and in face to face conversations, coupled with the need to understand and mine these for end applications, has motivated a great amount of research in this field in recent times. Researchers have explored a wide array of knowledge resources for opinion analysis, from words and phrases to syntactic dependencies and semantic relations.*

Cambria et al. (2013) name four major categories of *knowledge resources* related to sentiment analysis: *keyword spotting, lexical affinity, statistical methods and concept-based techniques*.

The authors discuss the methodological development from the first basic techniques based on pre-defined polarized vocabulary, together with their core problems, such as no extraction of any important meaning related to grammatical constructions, negation, compound sentences or semantic relations (e.g. co-reference or anaphora), through polarized lexicons and up to the application of syntactic text processing. Another analogical and, logically almost simultaneous, trend in sentiment analysis research has been the shift from whole-text or paragraph parameterization (general sentiment score based on the number of polarized words in a given text) towards sentence and segment — level analysis, up to the evaluation of single objects (see Liu, 2012).

After exploring a range of studies in search of an appropriate methodology, it would appear that there is a discrepancy between the humanist and engineering / scientific fields with regards to dealing with the analysis of opinions and evaluation. The most important distinction occurs between the narrow but precise perspective of *sentiment analysis* (especially in computational linguistics) and the broad, thoroughly elaborated but often IT-inapplicable, understanding of *opinion research* and studies of *evaluative — or even biased — language* (esp. in linguistic axiology, historical ethno-linguistics and discourse analysis, or even sociology). Somewhere in between there are also psychological (e.g. Smogfarm) approaches, focusing on the investigation of basic emotions, such as e.g. fear, stress, etc. expressed in evaluative texts and conceptual lexicons (SenticNet), which collect ideas or concepts (rather than vocabulary) with respect to their polarity.

As far as *NLP-based sentiment analysis* is concerned, it is most broadly understood as content (vocabulary but recently also concepts) classification into positively or negatively evaluated categories (polarity). A typical approach to sentiment analysis in the field of computational linguistics consists of:

- an analysis of vocabulary (parts of speech) of specific positive or negative character, be it adjectives (*bad/tasty*), nouns (*a flop*), verbs (*to disappoint*), adverbs (*generously*), etc. (see e.g. polarity lexicons, such as: Bing Liu’s Opinion Lexicon, MPQA Subjectivity Lexicon, SentiWordNet, Harvard General Inquirer, LIWC);
- in many (if not most) approaches published in the field of computational linguistics (e.g. Meng & Wang, 2009; Hu & Liu, 2004; Brody & Elhadad, 2010), we come across an even narrower scope of the problem, reducing its spectrum of analysis to the pairs of adjectives + nouns described by them only.

The statistical approach is realized mostly via probabilistic research of corpora, that is, either supervised or unsupervised machine learning. In the first case, corpora are annotated with sentiment tags by linguists and then the system is trained to automatically provide tags for further texts. In the unsupervised approach, the Web is treated as the corpus, where concepts (e.g. TV set) are extracted together with their typical features (price, model, etc.; Meng & Wang, 2009; Hu & Liu, 2004). In both approaches, typical adjectives describing the items (sometimes called aspects) are also extracted and problems such as context-dependent meaning (e.g. *warm soup* — positive, *warm ice cream* — negative) are also solved at the level of probability of occurrence in texts from a given domain (with a particular vocabulary surrounding). An indisputable advantage of these methods is that some of them are domain-independent, supporting their data-bases with unlimited Web data and that sentiment is mostly expressed via positive and negative vocabulary. Nevertheless, only direct polarization is covered.

A slightly broader understanding of sentiment analysis as *polarization of concepts* (knowledge or experience) rather than lexemes can be found both in linguistics and computational linguistics. Here, sentiment analysis is conceived of as requiring common knowledge of the world, necessary for the interpretation of statements, which do not include any vocabulary of a positive or negative sense, but which humans judge as positive or negative objects, facts, ideas or features, based on their common or individual experience. We can distinguish at least two approaches:

- data-bases of concepts expressing negative or positive experience (to and of the human kind), e.g. *failure* is always negative to its agent, e.g. SenticNet, or supported by the valence studies

(*failure* is negative to the agent expressed in the subject, as in the approach by Nasykawa & Yi, 2003);

- concepts and / or statements / situations of a negative or positive character from the speaker's perspective (subjectivity of meaning) — e.g. *This restaurant has a parking place*, e.g. Liu, 2012 or news analysis by Balahur and Steinberger (2009);

These approaches are also valuable, as they introduce the importance of knowledge and experience to the language study and are thus a clear step towards linguistic opinion research involving the notions of *perspective* and *perspectivizing*. The latter approach, especially, takes into consideration objective polarized statements (see below, Liu, 2012). However, there are more possibilities of implicit content expressions which should be studied and coded.

Finally, a very broad understanding of sentiment analysis on a conceptual level is to be found in the studies of *different stances* (e.g. political) and their related values expressed by the language. Such an approach is only offered by a purely linguistic (not computational application) approach to opinion research:

- general approaches to axiology of meaning and the system of values standing behind a given concept (Lakoff, 2002, etc.), that is concepts and / or statements / situations of a negative or positive character for a given person (subjectivity of meaning), where the decision about the favorability depends on the values (e.g. political stance) reflected by the language, e.g. *a war on drugs* is positive from the conservative point of view and negative from the liberal one,
- socio-linguistic and discourse analysis approaches to perspectivizing in the public discourse, e.g. (biased or framed) coverage of events (e.g. Ensink & Sauer, 2003; Meyer et al., 2000; Sadaba, 2006), similar to the previous example, but focusing on the media as an author.

The evident advantage of the above is a thorough discourse analysis of public events, with their influence on public opinion. The problem with these approaches, from the point of view of opinion mining, is that they refer to a different type of discourse than opinion reviews (historical or political and public discourse). Furthermore, they have not yet been coded or formalized and rely more on the discourse and pragmatics level than syntax and, therefore, probably harder to capture computationally (an exception to this is an approach to sentiment analysis in media news coverage by Balahur and Steinberger, 2009, which is computational and linguistically inspired).

Nevertheless, in none of these publications has there been described a thorough approach to implicit content analysis of opinion reviews on the pragmatic level. All of the studies described above are extremely useful, especially the parsing approach, supported by a sentiment dictionary, which tends to cover more than a half of useful cases of sentiment analysis. However, there are several things still missing with regards to sentiment analysis and opinion mining of user reviews — namely pragmatics and the description of the situation of communication: **the act of communication with the sender, the addressee, the message and the goal**. Employing such a perspective as a starting point of analysis guarantees the right interpretation of seemingly objective statements such as: *There is no wifi at the restaurant*, which in reality implies that some previous customer was unhappy with the lack of Internet connection and tries to warn future customers about this inconvenience. Even though the linguistic construction is a present tense in an indicative mood and no negative polarization is detected, the pragmatic analysis of the situation enables one to ascribe it to a *warning* or a *complaint*.

An opinion — including those expressed within customers' reviews — contains more valuable data than target expressions or concepts and their evaluation, but also discursive goals such as e.g. explanations (Zhang, Qian, Chen, Kang, & Huang, 2013), feelings (Somasundaran, 2010), experience with the product and product recommendations for the producer and prospective users. According to the latest research (Zhang et al, 2013) and to the best of our knowledge, an approach covering all of these aspects of reviews analysis is in demand.

### 3 Implicit Content Defined in a Rule-based Fashion

The closest to achieving the goal of implicit evaluation coverage have been several separate studies discussed by Liu (2012, p. 19). Liu distinguishes between **explicit** (subjective statements giving regular or comparative opinions like *Coke tastes great*) and **implicit opinions** (objective statement implying a regular or comparative opinion *I bought a mattress a week ago, and a valley has formed*) and comments that the former are easier to detect and classify, whereas (most likely as a result of this) *relatively less work has been done on implicit opinions* (Zhang & Liu, 2011; Liu, 2012). On the other hand, a subjective sentence does not necessarily have to express a sentiment (e.g. *I think that he went home*), whereas an objective one can imply opinions and sentiments *due to desirable and undesirable facts* (Liu, 2012, p. 20). He therefore advocates characterizing utterances as **opinionated or not opinionated** as a more appropriate term (Liu, 2012, p. 38).

Liu then introduces the concept of **rules of opinion and compositional semantics**, which are based on the linguistic assumptions that certain compound expressions, be they syntactic, semantic or mixed, express implications of positive and negative sentiments. Their analysis, description and identification, sometimes also requiring a description of world knowledge, enables the extraction of related data. Examples by Liu (2012, p. 53) are given mostly on a conceptual level (with possibilities of extension or elaboration) and include a definition of sentiment shifters e.g. *negation words* (*never, none, not, etc.*), *modal auxiliary verbs* (e.g. *would, could, should, might, etc.*), *presupposition markers of discontent* (e.g. *barely, hardly; fails* — e.g. *to impress me, etc.*). Also *high, low, increased or decreased quantity of an opinionated item* can mean positive or negative opinion, e.g. *reducing pain* is positive of a drug, contrary to *reducing concentration skills*. *A reduced price of a camera* is positive, *a high price* means the contrary (Liu, 2012, p. 55). Similar studies have been conducted for *desirable and undesirable facts* like a hole in a mattress after two weeks of its use and *production of a large and small resource quantity* (e.g. *a battery using up a lot of energy*). Nevertheless, the aforementioned rules are deemed to be highly domain-dependent, which is a drawback of this approach to sentiment analysis.

A compositional semantics approach translated to rule-based sentiment analysis modules or systems is still a very good method to examine how sentiment can be expressed via objective discourse. However, a syntactic or a domain-dependent approach might not provide for many possible types of implicit evaluations found in opinion reviews. Therefore, an experiment has been conducted which aims towards a more complex constructional approach — starting from pragmatics, through semantic-pragmatic groups of opinions and up to syntactic-semantic constructions involving speech acts, in order to discover how to extend the findings of case-and domain-limited rule-based compositional semantic studies to different ranges and sources of opinion reviews.

### 4 The Pragmatic Perspective on Opinion Reviews in Polish and English

The speech act theory by Austin (1962) assumes a taxonomy of five basic types of illocutionary acts: *verdictive, expositive, exercitive, behabitive* and *commissive*. The theory was then further elaborated upon by Searle (1969), who introduced the distinction between illocutionary verbs and acts, as well as direct and indirect ways of expressing them: “ (...) *there are five general ways of using language, five general categories of illocutionary acts. We tell people how things are (Assertives), we try to get them to do things (Directives), we commit ourselves to doing things (Commissives), we express our feelings and attitudes (Expressives), and we bring about changes in the world through our utterances (Declarations).*” Searle (1969, pp. vii–viii). Searle’s (1969, p. X) intention was to open *the discussions of the relation between literal sentence meaning and intended speaker’s utterance meaning*. Grzegorzyczkowa (2010, s. 85) also indicates that such an approach bears a similarity to de Sesseure’s distinction between *la langue and la parole*, between language as a theoretical system and practically used speech. Nevertheless, both are necessary,

since according to Searle (1969, p. 1) “ (...) different basic illocutionary types are realized in the syntax of a natural language (...)”. We believe that defining different types of constructions (pairings of form and meaning, see Goldberg, 2003) in the speech act fashion can help foster the process of opinion mining in search of content not extracted with traditional polarization methods.

The significant added value of this theory in application to opinion reviews is the distinction between *direct* and *indirect speech acts* (Searle, 1969, p. 88). A direct hint is a direct reference to the interlocutor with a request or plea, and can be expressed via an imperative (e.g. *Open the window!*). An indirect hint can be expressed with an assertive or constative, as in *It's hot in here*. As we can see, some of the expressions might actually be double speech acts: a *direct assertive*, for example, can become an *indirect directive* (e.g. *Sir, you are standing on my foot!* Searle, 1969, p. 88). Similar constructions can be investigated within the texts of opinion reviews, assuming that they bear the characteristics of a spoken discourse (written — down speech, Ong, 1982).

In order to ascertain the occurrence (frequency, types, etc.) of speech acts in opinion reviews, a sample set of 55 stylistically different texts written by the users of review websites, such as Foursquare (about 15 short opinions of the quality of restaurants in Cracow), Trip Advisor (10 long opinions of the services of the Intercontinental hotel in Warsaw), Opineo (about 10 medium-length opinions of an HP laptop) and Appstore (about 20 short reviews concerning a mobile application used to learn how to play the piano), was analyzed. The texts ranged in length between one phrase-texts (e.g. two to four words, *Najlepsza czekolada w mieście / The best chocolate in town*) to elaborated 10-sentence long opinions, in order to reflect the variety of writing styles and discourse created within this Internet genre.

Their major feature, typically omitted in NLP publications, was *secondary orality*, meaning that — as an Internet genre — opinions are characterized by “the use of electronics for oral communication” (Ong, 1982, p. X). Contemporary changes in the means of communication have resulted in a hybrid character of Internet discourse, such as an opinion, which now displays the characteristics of both written and spoken language, with the latter prevailing (Bauman, 2013).

On the sample of texts described above, a pragmatic and semantic analysis of customer reviews has been conducted to specify the repeated types of content (expressions of specific speech acts with their corresponding grammatical patterns) to identify about 20 different types of constructions. Below are presented a few selected examples:

Example	Type and directness of speech act	Addressee	Goal of communication	Inferred or presupposed evaluation / sentiment	Grammatical construction indicative of a speech act
<i>Więcej tam nie idę / I'm not going back there.</i>	— <i>direct commissive (promise), indirect directive (warning)</i>	<i>Teleologue — the owner of the place or another customer</i>	To the other customer: <i>don't go there — I'm not going.</i>  To the owner: <i>Your service / product is so bad, I don't ever want to use it again.</i>	bad, poor service / product	Expression of intention or the lack of it: EN: <i>1<sup>st</sup> PS + not going to</i> PL: <i>1<sup>st</sup> PS + nie zamierzam</i> EN: <i>1<sup>st</sup> PS + not + present continuous;</i> PL: <i>1<sup>st</sup> PS + nie + czas terażniejszy prosty</i> EN: <i>1<sup>st</sup> PS + not + future simple</i> PL: <i>1<sup>st</sup> PS + nie + czas przyszły prosty</i> <i>Most constructions in both languages include:</i> — <i>time / frequency adv</i> — <i>action verbs</i>

Chciałem kupić do niego płytę główną tylko nie wiem gdzie i jak, gdyż na aukcjach nie mogę znaleźć tego podzespołu... / I wanted to buy the main board, but I don't know where and how, because I cannot find this piece at auctions...	— <i>direct assertive</i> , — <i>indirect directive</i>	Prospective customer	Watch out: if it breaks down, spare parts are hard to buy	Bad equipment	I cannot + find / buy, etc. / it is impossible to find / buy + synonymous constructions expressing lack of possibility ( <i>it's impossible to + action V</i> )
Przywróćcie starą wersję! / Bring back the old version!	— <i>direct directive (demand)</i>	Owner or producer	I liked the previous version more	Bad evaluation	Action verb in Imperative + 1 <sup>st</sup> PPl + object
Oby tak dalej! / Keep it like that!	— <i>direct directive (demand)</i> — <i>indirect directive / expressive (hint)</i>	Owner or producer	I like it. Don't change it — keep it this way.	Positive evaluation	PL: Particle oby + Particle tak + Adv / V EN: Imperative + 1 <sup>st</sup> S / PPl + object
Mieliśmy czekać 30min. 40minut trwało podanie wody. Okazało się, że w międzyczasie skończyły się zupy. Po godzinie (!) 3 na 7 osób dostało dania. / We were supposed to wait 30 minutes. It took them 40 minutes to bring us water. It turned out that in the meantime they run out of soups. After an hour (!) 3 of 7 people got their meals.	— <i>direct assertive</i> , — <i>indirect directive (warning)</i>	Prospective customer	Unless you want to wait forever, do not enter this place	Bad service / evaluation	— typically more than one short sentence in a row, — PS in all of them + 1 <sup>st</sup> PS or 1 <sup>st</sup> PL in the first sentences, then change of subject (to the object described, e.g. dishes or the name of the product or company) — Action verbs

## 5 Quantitative and Qualitative Results

Within the sample of opinion reviews described above, 24 out of 55 texts include 27 examples of illocutionary speech acts (sometimes more than one example within the same text). This means that almost a half of all opinions do not get extracted while applying traditional methods of sentiment analysis (including direct evaluative vocabulary). Of course, one can argue that these expressions of indirect evaluation sometimes (in some other cases) overlap with direct evaluation expressed in adjectives, adverbs, verbs and nouns, but — as we have shown in the aforementioned examples — this does not have to be the case. One only needs to look at the last example — an event description at the Intercontinental hotel — to conclude that some reviews do not have to include a single evaluative lexeme to imply a negative statement.

The basic types of speech acts included in opinion reviews are hints and warnings directed at prospective customers or hints, demand and laudation directed at owners or producers. We therefore have encountered four major types of speech acts and two types of addressees (prospective customer/user and company or place owner/product producer).

The majority (11) of utterances could be classified as *hints*, which is compatible with the results of the types of addressees (prevailing prospective customers). This fact confirms that opinions are mostly hints for prospective customers or producers. The second result — warnings are mostly directed at future customers. Finally, direct demand or laudation are only directed at producers or owners (3 opinions only).

Quantitatively speaking, prospective users and customers prevail as addressees of communication. We typically encounter three types of text with respect to categories of addressees —

directed at:

- Prospective customers/ users,
- Owners/ producers,
- Directed at both of them or wider audience (*teleologue* — more than one possible addressee, no specific addressee).

Judging by the number of texts clearly directed at owners or users, there can be said to be more than three times as many texts directed at customers/users than at owners/ producers. In total, most texts (10) were directed at customers/users, about 30% more than in the case of teleologue. However, if we included the results of a collective category (10 in total) into the results for both groups, we would see approximately twice as many reviews directed at prospective customers than at owners and producers.

This sample, although probably not 100% statistically exact and reflecting an average type of review or addressee, provides an interesting outlook for bigger corpora of opinion reviews. What is also significant is that direct references represent a minority of speech acts, whereas indirect hints and warnings prevail. This in turn might prove to be useful with regards to questions about which constructions to give priority to within the programming process.

## 6 Syntactic-Semantic Constructions

The translation of opinion reviews from Polish into English was conducted according to the methodology of searching for semantic equivalents (see Kubaszczyk, 2006, 1999), rather than using direct translations. The application of this approach was possible due to the fact that the translator is also a linguist with experience in analysis of English opinion reviews for the purposes of data extraction.

The constructions formalizing the semantic-pragmatic expression of indirect evaluation via speech acts, such as hints or warnings, turn out to be fairly non-language specific. Differences at the level of grammar tend to occur, but they are the ones already described in comparative descriptive grammars of both languages. Bearing in mind the methodology of translation (searching for pragmatic-semantic and formal equivalents) we can say that these constructions are made up of elements representing their semantic, lexical or formal syntactic equivalents.

With respect to specific types of opinion reviews, one has to note that:

- *Hints* in opinion reviews (in both languages) frequently consist of:
  - ◇ constructions of necessity or willingness, often with modal verbs and adverbs of frequency (e.g. *Muszę spróbować raz jeszcze!* / *I have to try once again*)
  - ◇ the use of recommendation verbs such as *recommend* in the present simple, first person singular (*Polecam!* / *I recommend!*), especially, in Polish, often concluding the whole review;
  - ◇ fixed expressions of completeness implying a customer does not have to look for anything else (almost taken from TV commercials): *all + in + one + noun* (e.g. *all in one application*)
  - ◇ expressions of doubt (verbs of thinking or hesitation) in simple past plus, linking words such as *ale/but* and expressions of no doubt (verbs of evaluation) in the present simple (Polish) or simple past (English), e.g. *Po opiniach na Gastronautach zastanawiałam się czy tu wejść, ale jednak spróbowałam i nie żałuję* / *Having read opinion reviews at Gastronauts I hesitated, if I should enter this place or not, but in the end I did and I didn't regret it;*
  - ◇ sets of synonyms of positive features concerning specific aspects (world knowledge), e.g. *convenient location* being *central, centric, close to the railway station*, etc.



- ◇ *could, ought to, should* and other modal constructions with reference to features of a given object, expressing lack of satisfaction with it, e.g. *wanna mogłaby być większa / the bathtub could be bigger* (as in Liu, 2012);
  - ◇ direct recommendations of action, such as *worth + action verb + (that)*, e.g. *Warto zadbać o to, aby pokój był na najwyższym piętrze (ze względu na widok) / It's worth booking a room in the last floor (because of the view)*. Such hints are implications of satisfaction with nice view in the last floor and lack of satisfaction with the view in lower floors.
- *Warnings* (in both languages) in opinion reviews are expressed via:
- ◇ Expressions of *intentions* such as *future tense* or specific phrases such as *as going to* or their negations (*I'm not going back there, you'll never use their services again*, etc.);
  - ◇ Conditional phrases (*Nie są podawane z automatu, jak się o nie nie zapyta. / They are not served if you don't order them yourself*);
  - ◇ Action verbs denoting a break-down, often metaphorical uses, such as *padł* (direct translation: *fell down*, or *fell dead* equivalent — broke down) or *nawalił* (failed) used either in the simple past or present simple tense with a time or frequency adverb (*I zasilacz mi padł i używam zastępczego / and the charger went down and I'm using a substitute*);
  - ◇ Constructions of lack of possibility, such as *possible + action verb in infinitive* or *modal verbs* (e.g. *can*) in the simple past or present simple (e.g. *Chciałem kupić części zamienne, ale to niemożliwe. / I wanted to buy spare parts, but it's impossible*);
  - ◇ Aspects and features as in Liu (2012), e.g. *W środku często traci się zasięg. / Internet access goes off frequently in this restaurant*. Typically sentences in the present simple;
  - ◇ A type of a mirroring expression to aspects and features in the case of descriptions of product experience are *stories* frequently occurring with reference to services. They consist of a few sentences with action verbs in the simple past (they can be recognized by action-verbs simple past sentences in a row, typically starting with *I* or *we* as a subject and moving to description of a reaction, such as *it turned out that...* or no reaction e.g. *No reply*), describing a negative solution provided to a customer's problem.
- *Demands* are uttered with the use of the imperative mode — which is the same in both languages,
- *Laudations* are sometimes expressed via idiomatic expressions (*punkt dla was / chalk one up for you*) or, especially in English (not encountered in Polish so far) — a direct reference in present simple (1<sup>st</sup> person Plural, e.g. *you rock!*). Here we can also encounter the imperative mode in English expressions, as in *Keep it like that! Don't change it!* and specific particle constructions in Polish as in *Oby tak dalej!*

Another interesting aspect is that laudations and demands directed towards owners and producers could also be deemed to be implicit positive and negative opinions as perceived by prospective customers. Any demand to change a specific product is also an implicit statement of dissatisfaction with it. Any laudation — on the contrary — of doubtless satisfaction and recommendation.

A thorough description of all the encountered cases and examples will soon be provided in a separate paper by the same author. The speech acts have therefore become an *interlanguage* — a *tertium comparationis* (see Sinclair, 1996; Koseska-Toszewa, Korytkowska, & Roszko, 2007) and a semantic starting point for analysis, from which syntactic or syntactic and lexical-semantic (if we use e.g. WordNet to extend search results) patterns can be used to build coding rules. However, what is more important is that such an approach seems fairly domain independent (hints and warnings occur regardless of whether the reviewers are expressing an opinion on hotels, restaurants or smart phone applications), which requires validation via further studies on a larger sample of texts.

## 7 Comments for Coding

The data provided above can be of use and aid while coding. Namely, there are certain syntactic and semantic features, characteristic of, and defined within, these constructions. One of the differences with respect to traditional sentiment analysis is that semantic and syntactic criteria are used apart from POS-tagging.

With respect to their makeup, the following types of constructions for coding rules can be defined:

- purely syntactic — generally the easiest ones for coding (e.g. imperative), although we can also come across some more complex ones in compound sentences (e.g. conditional sentences);
- syntactic and semantic (syntactic constructions + semantic categories to be defined, such as a semantic subclass of verbs, e.g. action verbs). While building such rules, WordNet or other similar lexical databases could be of significant help, while providing semantic subgroups of verbs and other word classes;
- lexical semantic, where first a set of key expressions has to be defined and then broadened with synonyms or other useful semantic relations (e.g. *center*, *centric*, *centrally located*). Typically, synonyms cannot be limited to the scope of one grammatical category, such as nouns, since the whole range of options will not be covered.

As can be deduced from the above constructions, thanks to semantic tools such as Wordnet and FrameNet, quite unusual coding rules can be built which consist of both semantic and syntactic components, such as morpho-syntactic (e.g. 2<sup>nd</sup> person plural) and syntactic structures (e.g. future tense) and semantic sets (e.g. action verbs of a restaurant/ordering frame).

## 8 Conclusions

Since there are a variety of options in which one can express a direct or indirect warning or piece of advice, this method requires further studies in constructions and descriptive grammar. However, it seems to the authors that it is worth shedding new light on review analysis and sentiment research. It also requires work on both pragmatics and semantics (often not separated by cognitive scientists, see Goldberg, 2003 or Grzegorzczkova, 2010), since it requires both contextual knowledge and world knowledge interpretation. Nevertheless, it seems to the authors, it is worth the effort, since it can provide additional significant data to what can be achieved via traditional polarization studies.

The analysis presented above shows that:

- coverage of what is interesting and significant from the point of view of prospective customers or producers requires a broader perspective than a morpho-syntactic one with sentiment dictionaries involved — it engages pragmatics, semantics, syntax and morphology for a full scope;
- recently there has been some intuition that plain sentiment methods are not enough — e.g. psychological approaches by ... They are also useful, but they seem to focus on one aspect and there seems to be a need for interdisciplinary research;
- in many cases, elements indicative of sentiment (regardless of their direct or indirect character) span across more than one sentence, which still poses a challenge to many sentiment analysis models (e.g. *I had been wondering whether to enter or not, but in the end I did and I don't regret it*). This example shows an indisputable advantage of application of rule-based parsing methods to such problems of sentiment analysis;
- As with any grammatical (e.g. modality, time, etc.) or lexical category (e.g. force, see Talmy), or grammatical construction (e.g. di-transitive construction, see Goldberg, 1995), evaluation can be expressed in a variety of ways, be it in a direct way with specific polarized

- vocabulary (e.g. *badly located*), or indirect ways via use of specific grammatical forms (tenses or types of verbs, e.g. modal verbs, e.g. *could be better located*) or via hints requiring deduction by the interlocutor in discourse (e.g. centrally located);
- in the majority of examples we are given indirect meaningful hints, which could not be captured and covered by traditional systems of sentiment analysis.

As can be deduced from the above constructions, thanks to semantic tools, such as Wordnet and FrameNet, quite unusual coding rules can be built which consist of both semantic and syntactic components, such as morpho-syntactic (e.g. 2<sup>nd</sup> person plural) and syntactic structures (e.g. future tense) and semantic sets (e.g. action verbs of a restaurant/ordering frame)

Speech acts have already found several applications in computational linguistics, retrievable from the ACL Anthology. Applications, such as in dialogue systems (e.g. Reschke, Vogel, & Jurafsky, 2013), unsupervised detection of speech acts in English and Italian (Novielli, Strapparava, 2009), analysis of indirect acts of requesting and informing (Perrault & Allen, 1980), study of differences between effects and consequences of speech acts (Field & Ramsay, 2006) and formal logic for description of speech acts (Vanderveken, 1990) have been presented at a number of conferences. Nevertheless — to the best of our knowledge — none of the authors has so far tested the written text of opinion reviews from the perspective of secondary orality via the application of speech acts analysis to opinion reviews.

## References

- Austin J. (1962). *How to do things with words*. New York: Oxford University Press.
- Balahur A., Steinberger R. (2009). Rethinking Sentiment Analysis in the News: from Theory to Practice and back. *Proceedings of WOMSA*.
- Bauman Z. (2013). Lecture "Człowiek. Granice", Malta Festival, 28th June 2013 Poznań.
- Brody S., Elhadad N. (2010). An unsupervised aspect-sentiment model for online reviews. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 804–812.
- Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2), 15–21. <http://dx.doi.org/10.1109/mis.2013.30>.
- Chiticariu, L., Li, Y., & Reiss, F. R. (2013). Rule-based information extraction is dead! Long live rule-based information extraction systems! In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 827–832). Seattle, Washington, USA, 18–21 October 2013.
- Eiser, J. R. (2007). Attitudes and the use of evaluative language: A two-way process. *Journal for the Theory of Social Behavior*, 5(2), 235–248. <http://dx.doi.org/10.1111/j.1468-5914.1975.tb00354.x>
- Ensink T. & Sauer Ch. (2003). *Framing and Perspectivising in Discourse*. John Benjamins Publishing Company.
- Field, D. & Ramsay, A. (2006). How to change a person's mind: Understanding the difference between the effects and consequences of speech acts. In *Proceedings of the 5th Workshop on Inference in Computational Semantics (ICoS-5), 20–21 April 2006* (pp. 27–36). Retrieved from <http://anthology.aclweb.org/W/W06/W06-3903.pdf>
- Grzegorzczkova R. (2010). *Wprowadzenie do semantyki językoznawczej*. PWN.
- Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences*, 7(5):219–224.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.
- Hu, M. & Liu, B. (2004). Mining opinion features in customer reviews. In *Proceedings of Nineteenth National Conference on Artificial Intelligence* (pp. 755–760).
- Koseska-Toszewa, V., Korytkowska, M., & Roszko, R. (2007). *Polsko-bułgarska gramatyka konfrontatywna*. Warszawa: Dialog.
- Kubaszczyk, J. (1999). Kognition und Übersetzen: Über das Übersetzen von Konzepten. *Glottodidactica*, 27, 77–89.
- Kubaszczyk, J. (2006). Übersetzungsäquivalente der substantivierten Adjektivwörter vom Typ das Neue, das Schöne (neutrale Abstrakta) im Polnischen. *Glottodidactica*, 32, 73–85.

- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167. <http://dx.doi.org/10.2200/S00416ED1V01Y201204HLT016>.
- Lakoff G. (2002). *Moral Politics: How Liberals and Conservatives Think*. University of Chicago Press.
- Meng, X. & Wang, H. (2009). Mining user reviews: From specification to summarization. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers* (pp. 177–180). Suntec, Singapore, 4 August 2009. Stroudsburg, PA, USA: Association for Computational Linguistics. <http://dx.doi.org/10.3115/1667583.1667637>.
- Meyer M., Titscher S., Vetter E., Wodak R. (2000). *Methods of Text and Discourse Analysis*. Sage Verlag, London, Thousand Oaks, New Delhi.
- Nasykawa, T. & Yi, J. (2003). *Sentiment analysis: Capturing favorability using natural language processing*. K-CAP'03, October 23–25, 2003, Sanibel Island, Florida, USA.
- Novielli, N. & Strapparava, C. (2009). Towards unsupervised recognition of dialogue acts. In *Proceedings of the NAACL HLT Student Research Workshop and Doctoral Consortium* (pp. 84–89). Boulder, Colorado, June 2009. 2009 Association for Computational Linguistics. <http://dx.doi.org/10.3115/1620932.1620947>.
- Ong, W. (1982). *Orality and literacy: The technologizing of the world*. London: Routledge.
- Perrault, C. R. & Allen, J. (1980). A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics*, 6(3–4).
- Pluwak, A., Korczyński, W., & Kisiel-Dorohinicki, M. (2016). Adapting a constituency parser to user-generated content in Polish opinion mining. *Computer Science*, 17(1), 23. <http://doi.org/10.7494/csci.2016.17.1.23>.
- Reschke, K., Vogel, A., & Jurafsky, D. (2013). Generating recommendation dialogs by extracting information from user reviews. In *Proceedings of the 51<sup>st</sup> Annual Meeting of the Association for Computational Linguistics* (pp. 499–504). Sofia, Bulgaria, August 4–9 2013.
- Sadaba T. (2006). *Framing: Una Teoria Para Los Medios de Comunicacion*. Ulzama, Pamplona.
- Searle, J. R. (1969). *Speech acts — an essay in the philosophy of language*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CB09781139173438>.
- Sinclair, J. (1996). The search for units of meaning. *Textus*, 9(1), 75–106.
- Somasundaran, S. (2010). *Discourse-level relations for opinion analysis* (Unpublished Doctoral Dissertation at John Hopkins University).
- Tokarski R. (2014). *Światy za słowami. Wykłady z semantyki leksykalnej*. Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.
- Vanderveken, D. (1990). *Meaning and speech acts* (Vol. 1, *Principles of language use*). Cambridge, England: Cambridge University Press.
- Zhang L., Bing L. (2011). Extracting Resource Terms for Sentiment Analysis. *Proceedings of the International Joint Conference on Natural Language Processing* (pp. 1171–1179).
- Zhang, Q., Qian, J., Chen, H., Kang, J., & Huang, X. (2013). Discourse level explanatory relation extraction from product reviews using first-order logic. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 946–957). Seattle, Washington, USA, 18–21 October 2013.

---

## Acknowledgment

This work was supported by a core funding for statutory activities from the Polish Ministry of Science and Higher Education.

The author declares that she has no competing interests.

This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 PL License (<http://creativecommons.org/licenses/by/3.0/pl/>), which permits redistribution, commercial and non-commercial, provided that the article is properly cited.