Dimensions of Subjectivity in Natural Language

Wei Chen
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213, USA
weichen@cs.cmu.edu

Abstract

Current research in automatic subjectivity analysis deals with various kinds of subjective statements involving human attitudes and emotions. While all of them are related to *subjectivity*, these statements usually touch on multiple dimensions such as non-objectivity[1], uncertainty, vagueness, non-objective measurability, imprecision, and ambiguity, which are inherently different. This paper discusses the differences and relations of six dimensions of *subjectivity*. Conceptual and linguistic characteristics of each dimension will be demonstrated under different contexts.

1 Introduction

Natural language involves statements that do not contain complete, exact, and unbiased information. Many of these are subjective, which share the common property described in narrative theory (Banfield, 1982) as "(subjective statements) must all be referred to the speaking subject for interpretation". Wiebe (1990) further adapted this definition of *subjectivity* to be "the linguistic expression of private states (Quirk et al., 1985)". So far, linguistic cues have played an important role in research of subjectivity recognition (e.g. (Wilson et al., 2006)), sentiment analysis (e.g. (Wilson et al., 2005; Pang and Lee, 2004)), and emotion studies (e.g. (Pennebaker et al., 2001)). While most linguistic cues

---

[1]We use the term "non-objectivity" to refer to the property of creating a bias from a speaker's point of view that is not supported by sufficient objective evidence. It is not identical to the *subjectivity* that involves all the dimensions we discuss in this paper.

are grouped under the general rubric of *subjectivity*, they are usually originated from different dimensions, including:

- non-objectivity
- uncertainty
- vagueness
- non-objective measurability
- imprecision
- ambiguity

These dimensions all mingle in various applications that deal with subjective statements. For example, opinion extraction processes statements involving non-objectivity and uncertainty. Evaluation and sentiment analysis deal with vague words, which often covers the issue of non-objective measurability and imprecision. Ambiguity sometimes involves implicit subjectivity that is hard to recognize from linguistic patterns, which leads to great challenge of identifying and understanding subjective statements.

Since multiple dimensions are involved in *subjectivity*, discriminating them may be helpful in understanding *subjectivity* and related concepts. The following sections discuss characteristics and relations of the six dimensions of *subjectivity*.

2 Dimensions of Subjective Statements

2.1 Non-objectivity

In this paper, we define non-objectivity as the property of creating a bias according to personal beliefs, judgments and emotions. This does not include the kind of subjectivity originated from particular properties of linguistic units that lead to personal interpretations. Non-objectivity exists in subjective

statements such as opinions, evaluations, and persuasive statements. Non-objectivity can be recognized from linguistic patterns including words explicitly expressing thoughts, beliefs, speculations, and postulations such as "think", "believe", "hope" and "guess". Although linguistic cues are found to be reliable, there are cases of non-objectivity that cannot be identified merely from lexical, syntactical or morphological cues. For example, sentence (1) and sentence (2) are very similar in linguistic structures, but only sentence (2) is non-objective.

(1) Living things cannot survive without water.

(2) He cannot survive without music.

Apart from linguistic patterns and conceptual characteristics of non-objectivity, there are two main issues in non-objectivity recognition. First, non-objectivity cannot be clearly identified without knowledge about its source (Wiebe et al., 2005). For example, "Bob says the red team is about to win" is objective with respect to the position of the speaker of the sentence, who objectively stated a speech event. But the fragment "the red team is about to win" is an opinion of Bob. Hence, whether a statement is an opinion depends on both the scope of the statement and the source of that statement. Second, non-objectivity always lies in a context, which cannot be ignored (Wiebe, 1990). For example, "Pinocchio's nose" is likely to be objective when used within the context of the famous fairy tale. But the same phrase can be used subjectively as a metaphor in other contexts, where it may indicate non-objectivity.

2.2 Uncertainty

Uncertainty can indicate either subjectivity or objectivity. Flagged by words such as "probably" and "maybe", statements expressing uncertainty are usually considered subjective because "being uncertain" itself can be a subjective mental activity. However, uncertainty is not a subtype of *subjectivity*. Consider the following sentences:

(3) Bob has probably already finished his homework.

(4) A poll of recent public opinions shows that Bob is likely to win the nomination.

Sentence (3) is a subjective statement, where the speaker expresses his/her postulation of "Bob finished his homework" through the uncertainty indicated by "probably". On the contrary, sentence (4) is an objective statement, although uncertainty about a future event exists. This sentence reports a conclusion drawn from sufficient evident that Bob takes the majority vote based on the survey, which does not rely on a particular speaking subject for interpretation. In this case, uncertainty does not necessarily imply *subjectivity*.

On the other hand, people sometimes explicitly indicate uncertainty to avoid being subjective.

(5) It is *possible* that the red team will win.

(6) It is *likely* that the red team will win.

(7) The red team will win.

We could easily imagine a scenario where sentence (5) is more objective than sentence (6) and (7). For example, the speaker may believe that the red team will lose, but in order to avoid personal bias, he/she may instead say: "It is possible that the red team will win (but the blue team has a better chance)." In general, explicitly showing uncertainty can imply postulation, but it can also convey the intention of being objective by not excluding other possibilities.

Uncertainty sometimes exists in statements where no linguistic cues are present. For example, the linguistic pattern of sentence (7) is similar to that of "I will have an exam tomorrow", but the later one is usually used to describe an objective future event while sentence (7) can be semantically identical to sentence (6)[2], although the indicator of uncertainty in sentence (7) is not shown explicitly.

2.3 Vagueness, Non-objective Measurability, and Imprecision

Vagueness refers to a property of the concepts that have no precise definitions. For example, gradable words such as "small" and "popular" are sometimes treated as linguistic cues of vagueness, and they are found to be good indicators of *subjectivity* (Hatzivassiloglou and Wiebe, 2000).

Especially, gradable words are vague if there is no well-defined frame of reference. This in some cases

---

[2]These two are identical as long as the game is not fixed.

leads to two issues: comparison class and boundary. In the sentence "Elephants are big", the comparison class of "elephants" is unclear: we could compare the size of elephants with either land animals or all the animals including both land and aquatic creatures[3]. Also, there is no clear boundary between "being small" and "not being small". Different individuals usually have their own fuzzy boundaries for vague concepts. As such, vague words are usually treated as important cues for *subjectivity*. However, learning which words are vague is non-trivial, because vagueness cannot be hard-coded into lexicons. For example, the gradable word "cold" is vague in sentence (8) but not in sentence (9). The difference between these two is the one in sentence (9) has a known boundary which is the temperature for liquid water to exist, and the one in sentence (8) simply reflects personal perception.

(8) It is *cold* outside.

(9) It is too *cold* during the night on the moon for liquid water to exist.

Vagueness is often a strong indicator of *subjectivity* because it involves personal explanation of a concept. But there are exceptions. For example, the definition of "traditional education" can be vague, but talking about "traditional education" may not necessarily imply *subjectivity*.

When speaking of qualities, there are two major dimensions related to vagueness: non-objective measurability and imprecision. Attributes like height, length, weight, temperature, and time are objectively measurable, whereas things like beauty and wisdom are usually not objectively measurable. Vagueness exists at different levels for non-objectively and objectively measurable qualities. For non-objectively measurable qualities, vagueness exists at the conceptual level, where it intersects with non-objectivity. In the sentence "He is not as charming as his brother", the word "charming" refers to a quality whose interpretation may vary among different cultures and different individuals. For objectively measurable qualities, vagueness exists at the boundary-setting level, where either subjectivity or common sense comes into play. Sentence

(10) shows an example of the objectively measurable quality "long time" indicating an opinion that the speaker is unsatisfied with someone's work. On the contrary, an objective meaning of "long time" in sentence (11) can be resolved by common sense.

(10) You finally finished the work, but it took you a *long time*.

(11) Intelligent life took a *long time* to develop on Earth.[4]

Statements involving objectively measurable quantities often have an imprecision problem, where vagueness is usually resolved from common agreements on small variations of values. For example, "Bob is six feet tall" usually implies that the height is "around" six feet[5], with a commonly acceptable precision of about an inch. Generally, specific precisions are determined by variations tied to measurement technologies for specific quantities: the precision for the size of a cell may be around a micron, and the error tolerance for the distance between stars can be on the order of light years. Imprecision can also indicate subjectivity when used for subjective estimation. For instance, "Bob needs two days to finish his homework" is usually not telling an exact period of time, but a personal estimation.

2.4 Ambiguity

While vagueness exists at the conceptual level, ambiguity lies at the level of linguistic expressions. In other words, an ambiguous statement contains linguistic expressions that can refer to multiple explanations, whereas a vague statement carries a concept with unclear or soft definition.

Previous studies have explored the relationship between ambiguity and *subjectivity*. They have shown that subjectivity annotations can be helpful for word sense disambiguation when a word has distinct subjective senses and objective senses (Wiebe and Mihalcea, 2006).

Lexical and syntactical ambiguity usually can be resolved from contextual information and/or common consensus. But when ambiguity is used intentionality, identifying and understanding the ambiguity become creative and interactive procedures,

---

[3] Other comparison classes are also possible.

[4] Sentence fragment adapted from *Astrobiology Magazine (Dec 02, 2002)*.

[5] It could also mean "at least six feet tall" in some cases.

which usually indicate *subjectivity*. The sentence "I'd like to see more of you" is an example of this kind, which could be used to indicate multiple meanings under the same context [6].

3   Mixtures of Multiple Dimensions

In many cases, subjective statements involve multiple of the dimensions discussed in previous sections. For example, the subjectivity of the sentence "It's a nice car" comes from three dimensions: non-objectivity, vagueness and ambiguity. First, "a car being nice" is usually a personal opinion which may not be commonly acceptable. Second, the gradable word "nice" indicates vagueness, since there is no clear boundary for "being nice". Third, the sentence is also ambiguous because "nice" could refer to appearance, acceleration, angle rate, and many other metrics that might affect personal evaluations.

For information retrieval systems, processing natural queries such as "find me the popular movies of 2007" requires proper understanding of the vague word "popular". Besides, non-objectivity and ambiguity also take part in the query: on the non-objectivity side, the definition of "popular" may differ according to different individuals; on the ambiguity side, the word "popular" may refer to different metrics related to the popularity of a movie such as movie ratings and box office performance.

In applications requiring certain level of language-understanding, things can get even more complicated while different dimensions weave together. As in sentence (5), the speaker may bias towards the blue team while he/she shows uncertainty towards the red team. Correctly understanding this kind of subjective statements would probably need some investigation in different dimensions of *subjectivity*.

4   Conclusion

In this paper, we demonstrated that subjectivity in natural language is a complex phenomenon that contains multiple dimensions including non-objectivity, uncertainty, vagueness, non-objective measurability, imprecision and ambiguity. These dimensions pattern together in various kinds of subjective state-

ments such as opinions, evaluations and natural queries. Since these dimensions have different behaviors in subjective statements, discriminating them in both linguistic and psychological aspects would be necessary in subjectivity analysis.

Acknowledgments

References

Ann Banfield. 1982. *Unspeakable Sentences: Narration and Representation in the Language of Fiction*. Routledge and Kegan Paul, Boston.

Vasileios Hatzivassiloglou and Janyce Wiebe. 2000. Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of the 18th conference on Computational linguistics*, pages 299–305.

Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the ACL*, pages 271–278.

James Pennebaker, Martha Francis, and Roger Booth. 2001. *Linguistic Inquiry and Word Count: LIWC*. Lawrence Erlbaum Associates, Mahwah.

Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman, New York.

Janyce Wiebe and Rada Mihalcea. 2006. Word sense and subjectivity. In *Proceedings of the ACL*, pages 1065–1072.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. In *Language Resources and Evaluation*, volume 39, pages 165–210.

Janyce Wiebe. 1990. *Recognizing Subjective Sentences: A Computational Investigation of Narrative Text*. Ph.D. thesis, SUNY Buffalo Dept. of Computer Science.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 347–354.

Theresa Wilson, Janyce Wiebe, and Rebecca Hwa. 2006. Recognizing strong and weak opinion clauses. *Computational Intelligence*, 22(2):73–99.

---

[6] Kent Bach, Ambiguity. *Routledge Encyclopedia of Philosophy*, http://online.sfsu.edu/ kbach/ambguity.html