Subjective Natural Language Problems: Motivations, Applications, Characterizations, and Implications

Cecilia Ovesdotter Alm
Department of English
College of Liberal Arts
Rochester Institute of Technology
coagla@rit.edu

Abstract

This opinion paper discusses subjective natural language problems in terms of their motivations, applications, characterizations, and implications. It argues that such problems deserve increased attention because of their potential to challenge the status of theoretical understanding, problem-solving methods, and evaluation techniques in computational linguistics. The author supports a more holistic approach to such problems; a view that extends beyond opinion mining or sentiment analysis.

1 Introduction

Interest in subjective meaning and individual, interpersonal or social, poetic/creative, and affective dimensions of language is not new to linguistics or computational approaches to language. Language analysts, including computational linguists, have long acknowledged the importance of such topics (Bühler, 1934; Lyons, 1977; Jakobson, 1996; Halliday, 1996; Wiebe et al, 2004; Wilson et al, 2005). In computational linguistics and natural language processing (NLP), current efforts on subjective natural language problems are concentrated on the vibrant field of opinion mining and sentiment analysis (Liu, 2010; Täckström, 2009), and ACL-HLT 2011 lists *Sentiment Analysis, Opinion Mining and Text Classification* as a subject area. The terms *subjectivity* or *subjectivity analysis* are also established in the NLP literature to cover these topics of growing inquiry.

The purpose of this opinion paper is not to provide a survey of subjective natural language problems. Rather, it intends to launch discussions about how subjective natural language problems have a vital role to play in computational linguistics and in shaping fundamental questions in the field for the future. An additional point of departure is that a continuing focus on primarily the fundamental distinction of *facts* vs. *opinions* (implicitly, denotative vs. connotative meaning) is, alas, somewhat limiting. An expanded scope of problem types will benefit our understanding of subjective language and approaches to tackling this family of problems.

It is definitely reasonable to assume that problems involving subjective perception, meaning, and language behaviors will diversify and earn increased attention from computational approaches to language. Banea et al already noted: "We have seen a surge in interest towards the application of automatic tools and techniques for the extraction of opinions, emotions, and sentiments in text (*subjectivity*)" (p. 127) (Banea et al, 2008). Therefore, it is timely and useful to examine subjective natural language problems from different angles. The following account is an attempt in this direction. The first angle that the paper comments upon is what motivates investigatory efforts into such problems. Next, the paper clarifies what subjective natural language processing problems are by providing a few illustrative examples of some relevant problem-solving and application areas. This is followed by discussing yet another angle of this family of problems, namely what some of their characteristics are. Finally, potential implications for the field of computational linguistics at large are addressed, with the hope that this short piece will spawn continued discussion.

Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics:shortpapers, pages 107–112,
Portland, Oregon, June 19-24, 2011.

2 Motivations

The types of problems under discussion here are fundamental language tasks, processes, and phenomena that mirror and play important roles in people's daily social, interactional, or affective lives. Subjective natural language processing problems represent exciting frontier areas that directly relate to advances in artificial natural language behavior, improved intelligent access to information, and more agreeable and comfortable language-based human-computer interaction. As just one example, interactional systems continue to suffer from a bias toward 'neutral', unexpressive (and thus communicatively cumbersome) language.

From a practical, application-oriented point of view, dedicating more resources and efforts to subjective natural language problems is a natural step, given the wealth of available written, spoken or multimodal texts and information associated with creativity, socializing, and subtle interpretation. From a conceptual and methodological perspective, automatic subjective text analysis approaches have potential to challenge the state of theoretical understanding, problem-solving methods, and evaluation techniques. The discussion will return to this point in section 5.

3 Applications

Subjective natural language problems extend well beyond sentiment and opinion analysis. They involve a myriad of topics–from linguistic creativity via inference-based forecasting to generation of social and affective language use. For the sake of illustration, four such cases are presented below (bearing in mind that the list is open-ended).

3.1 Case 1: Modeling affect in language

A range of affective computing applications apply to language (Picard, 1997). One such area is automatically inferring affect in text. Work on automatic affect inference from language data has generally involved recognition or generation models that contrast a range of affective states either along affect categories (e.g. angry, happy, surprised, neutral, etc.) or dimensions (e.g. arousal and pleasantness). As one example, Alm developed an affect dataset and explored automatic prediction of affect in text at the sentence level that accounted for different levels of affective granularity (Alm, 2008; Alm, 2009; Alm, 2010). There are other examples of the strong interest in affective NLP or affective interfacing (Liu et al, 2003; Holzman and Pottenger, 2003; Francisco and Gervás, 2006; Kalra and Karahalios, 2005; Généreux and Evans, 2006; Mihalcea and Liu, 2006). Affective semantics is difficult for many automatic techniques to capture because rather than simple text-derived 'surface' features, it requires sophisticated, 'deep' natural language understanding that draws on subjective human knowledge, interpretation, and experience. At the same time, approaches that accumulate knowledge bases face issues such as the artificiality and limitations of trying to enumerate rather than perceive and experience human understanding.

3.2 Case 2: Image sense discrimination

Image sense discrimination refers to the problem of determining which images belong together (or not) (Loeff et al, 2006; Forsyth et al, 2009). What counts as the sense of an image adds subjective complexity. For instance, images capture "both **word and iconographic sense distinctions** ... CRANE can refer to, e.g. a MACHINE or a BIRD; iconographic distinctions could additionally include birds standing, vs. in a marsh land, or flying, i.e. sense distinctions encoded by further descriptive modication in text." (p. 547) (Loeff et al, 2006). In other words, images can evoke a range of subtle, subjective meaning phenomena. Challenges for annotating images according to lexical meaning (and the use of verification as one way to assess annotation quality) have been discussed in depth, cf. (Alm et al, 2006).

3.3 Case 3: Multilingual communication

The world is multilingual and so are many human language technology users. Multilingual applications have strong potential to grow. Arguably, future generations of users will increasingly demand tools capable of effective multilingual tasking, communication and inference-making (besides expecting adjustments to non-native and cross-linguistic behaviors). The challenges of code-mixing include dynamically adapting sociolinguistic forms and functions, and they involve both flexible, subjective sense-making and perspective-taking.

3.4 Case 4: Individualized iCALL

A challenging problem area of general interest is language learning. State-of-the-art intelligent computer-assisted language learning (iCALL) approaches generally bundle language learners into a homogeneous group. However, learners are individuals exhibiting a vast range of various kinds of differences. The subjective aspects here are at another level than meaning. Language learners apply personalized strategies to acquisition, and they have a myriad of individual communicative needs, motivations, backgrounds, and learning goals. A framework that recognizes subjectivity in iCALL might exploit such differences to create tailored acquisition flows that address learning curves and proficiency enhancement in an individualized manner. Countering boredom can be an additional positive side-effect of such approaches.

4 Characterizations

It must be acknowledged that a problem such as inferring affective meaning from text is a substantially different kind of 'beast' compared to predicting, for example, part-of-speech tags.[1] Identifying such problems and tackling their solutions is also becoming increasingly desirable with the boom of personalized, user-generated contents. It is a useful intellectual exercise to consider what the general characteristics of this family of problems are. This initial discussion is likely not complete; that is also not the scope of this piece. The following list is rather intended as a set of departure points to spark discussion.

- **Non-traditional intersubjectivity** Subjective natural language processing problems are generally problems of meaning or communication where so-called *intersubjective agreement* does not apply in the same way as in traditional tasks.

- **Theory gaps** A particular challenge is that subjective language phenomena are often less understood by current theory. As an example, in the affective sciences there is a vibrant debate–indeed a controversy–on how to model or even define a concept such as emotion.

- **Variation in human behavior** Humans often vary in their assessments of these language behaviors. The variability could reflect, for example, individual preferences and perceptual differences, and that humans adapt, readjust, or change their mind according to situation details. Humans (e.g. dataset annotators) may be sensitive to sensory demands, cognitive fatigue, and external factors that affect judgements made at a particular place and point in time. Arguably, this behavioral variation is part of the given subjective language problem.

- **Absence of real 'ground truth'?** For such problems, *acceptability* may be a more useful concept than 'right' and 'wrong'. A particular solution may be *acceptable/unacceptable* rather than *accurate/erroneous*, and there may be more than one acceptable solution. (Recognizing this does not exclude that acceptability may in clear, prototypical cases converge on just one solution, but this scenario may not apply to a majority of instances.) This central characteristic is, conceptually, at odds with interannotator agreement 'targets' and standard performance measures, potentially creating an abstraction gap to be filled. If we recognize that *(ground) truth* is, under some circumstances, a less useful concept–a problem reduction and simplification that is undesirable because it does not reflect the behavior of language users–how should evaluation then be approached with rigor?

- **Social/interpersonal focus** Many problems in this family concern inference (or generation) of complex, subtle dimensions of meaning and information, informed by experience or socioculturally influenced language use in real-situation contexts (including human-computer interaction). They tend to tie into *sociolinguistic* and *interactional* insights on language (Mesthrie et al, 2009).

- **Multimodality and interdisciplinarity** Many of these problems have an interactive and humanistic basis. Multimodal inference is arguably also of importance. For example, written web texts are accompanied by visual mat-

[1] No offense intended to POS tagger developers.

ter ('texts'), such as images, videos, and text aesthetics (font choices, etc.). As another example, speech is accompanied by biophysical cues, visible gestures, and other perceivable indicators.

It must be recognized that, as one would expect, one cannot 'neatly' separate out problems of this type, but core characteristics such as *non-traditional intersubjectivity*, *variation in human behavior*, and recognition of *absence of real 'ground truth'* may be quite useful to understand and appropriately model problems, methods, and evaluation techniques.

5 Implications

The cases discussed above in section 3 are just selections from the broad range of topics involving aspects of subjectivity, but at least they provide glimpses at what can be done in this area. The list could be expanded to problems intersecting with the digital humanities, healthcare, economics or finance, and political science, but such discussions go beyond the scope of this paper. Instead the last item on this agenda concerns the broader, disciplinary implications that subjective natural language problems raise.

- **Evaluation** If the concept of "ground truth" needs to be reassessed for subjective natural language processing tasks, different and alternative evaluation techniques deserve careful thought. This requires openness to alternative assessment metrics (beyond precision, recall, etc.) that fit the problem type. For example, evaluating *user interaction* and *satisfaction*, as Liu et al (2003) did for an affective email client, may be relevant. Similarly, analysis of acceptability (e.g. via user or annotation verification) can be informative. MOS testing for speech and visual systems has such flavors. Measuring pejoration and amelioration effects on other NLP tasks for which standard benchmarks exist is another such route. In some contexts, other measures of quality of life improvements may help complement (or, if appropriate, substitute) standard evaluation metrics. These may include ergonomics, personal contentment, cognitive and physical load (e.g. counting task steps or load broken down into units), safety increase and non-invasiveness (e.g. attention upgrade when performing a complex task), or. Combining standard metrics of system performance with alternative assessment methods may provide especially valuable holistic evaluation information.

- **Dataset annotation** Studies of human annotations generally report on interannotator agreement, and many annotation schemes and efforts seek to reduce variability. That may not be appropriate (Zaenen, 2006), considering these kinds of problems (Alm, 2010). Rather, it makes sense to take advantage of corpus annotation as a resource, beyond computational work, for investigation into actual language behaviors associated with the set of problems dealt with in this paper (e.g. variability vs. trends and language–culture–domain dependence vs. independence). For example, label-internal divergence and intraannotator variation may provide useful understanding of the language phenomenon at stake; surveys, video recordings, think-alouds, or interviews may give additional insights on human (annotator) behavior. The genetic computation community has theorized concepts such as user fatigue and devised robust algorithms that integrate interactional, human input in effective ways (Llorà et al, 2005; Llorà et al, 2005). Such insights can be exploited. Reporting on sociolinguistic information in datasets can be useful properties for many problems, assuming that it is feasible and ethical for a given context.

- **Analysis of ethical risks and gains** Overall, how language and technology coalesce in society is rarely covered; but see Sproat (2010) for an important exception. More specifically, whereas ethics has been discussed within the field of affective computing (Picard, 1997), how ethics applies to language technologies remains an unexplored area. Ethical interrogations (and guidelines) are especially important as language technologies continue to be refined and migrate to new domains. Potential problematic implications of language technologies–

or how disciplinary contributions affect the linguistic world–have rarely been a point of discussion. However, there are exceptions. For example, there are convincing arguments for gains that will result from an increased engagement with topics related to endangered languages and language documentation in computational linguistics (Bird, 2009), see also Abney and Bird (2010). By implication, such efforts may contribute to linguistic and cultural sustainability.

- **Interdisciplinary mixing** Given that many subjective natural language problem have a humanistic and interpersonal basis, it seems particularly pivotal with investigatory 'mixing' efforts that reach outside the computational linguistics community in multidisciplinary networks. As an example, to improve assessment of subjective natural language processing tasks, lessons can be learned from the human-computer interaction and social computing communities, as well as from the digital humanities. In addition, attention to multimodality will benefit increased interaction as it demands vision or tactile specialists, etc.[2]

- **Intellectual flexibility** Engaging with problems that challenge black and white, right vs. wrong answers, or even tractable solutions, present opportunities for intellectual growth. These problems can constitute an opportunity for training new generations to face challenges.

6 Conclusion

To conclude: there is a strong potential–or, as this paper argues, a necessity–to expand the scope of computational linguistic research into subjectivity. It is important to recognize that there is a broad family of relevant subjective natural language problems with theoretical and practical, real-world anchoring. The paper has also pointed out that there are certain aspects that deserve special attention. For instance, there are evaluation concepts in computational linguistics that, at least to some degree, detract attention away from how subjective perception and production phenomena actually manifest themselves in natural language. In encouraging a focus on efforts to achieve 'high-performing' systems (as measured along traditional lines), there is risk involved–the sacrificing of opportunities for fundamental insights that may lead to a more thorough understanding of language uses and users. Such insights may in fact decisively advance language science and artificial natural language intelligence.

[2]When thinking along multimodal lines, we might stand a chance at getting better at creating core models that apply successfully also to signed languages.

Acknowledgments

I would like to thank anonymous reviewers and colleagues for their helpful comments.

References

Abney, Steven and Steven Bird. 2010. The Human Language Project: Building a Universal Corpus of the worlds languages. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden*, 8897.

Alm, Cecilia Ovesdotter. 2009. *Affect in Text and Speech*. VDM Verlag: Saarbrcken.

Alm, Cecilia Ovesdotter. 2010. Characteristics of high agreement affect annotation in text. *Proceedings of the LAW IV workshop at the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden*, 118-122.

Alm, Cecilia Ovesdotter. 2008. Affect Dataset. GNU Public License.

Alm, Cecilia Ovesdotter and Xavier Llorá. 2006. Evolving emotional prosody *Proceedings of INTERSPEECH 2006 - ICSLP, Ninth International Conference on Spoken Language Processing, Pittsburgh, PA, USA*, 1826-1829.

Alm, Cecilia Ovesdotter, Nicolas Loeff, and David Forsyth. 2006. Challenges for annotating images for sense disambiguation. *Proceedings of the Workshop on Frontiers in Linguistically Annotated Corpora, at the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, Sydney*, 1-4.

Banea, Carmen, Rada Mihalcea, Janyce Wiebe, and Samer Hassan. 2008. Multilingual subjectivity analysis using machine translation. *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, 127-135.

Bird, Steven. 2009. Last words: Natural language processing and linguistic fieldwork. *Journal of Computational Linguistics*, 35 (3), 469-474.

Bühler, Karl. 1934. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Stuttgart: Gustav Fischer Verlag.

Forsyth, David, Tamana Berg, Cecilia Ovesdotter Alm, Ali Farhadi, Julia Hockenmaier, Nicolas Loeff, and Gang Wang. Words and pictures: categories, modifiers, depiction, and iconography. In S. J. Dickinson, et al (Eds.). *Object Categorization: Computer and Human Vision Perspectives*, 167-181. Cambridge: Cambridge Univ. Press.

Francisco, Virginia and Pablo Gervás. 2006. Exploring the compositionality of emotions in text: Word emotions, sentence emotions and automated tagging. *AAAI-06 Workshop on Computational Aesthetics: Artificial Intelligence Approaches to Beauty and Happiness*.

Généreux, Michel and Roger Evans. 2006. Distinguishing affective states in weblog posts. AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs, 40-42.

Halliday, Michael A. K. 1996. Linguistic function and literary style: An inquiry into the language of William Golding's *The Inheritors*. Weber, Jean Jacques (ed). *The Stylistics Reader: From Roman Jakobson to the Present*. London: Arnold, 56-86.

Holzman, Lars E. and William Pottenger. 2003. Classification of emotions in Internet chat: An application of machine learning using speech phonemes. LU-CSE-03-002, Lehigh University.

Jakobson, Roman. 1996. Closing statement: Linguistics and poetics. Weber, Jean Jacques (ed). *The Stylistics Reader: From Roman Jakobson to the Present*. London: Arnold, 10-35.

Karla, Ankur and Karrie Karahalios. 2005. TextTone: Expressing emotion through text. *Interact 2005*, 966-969.

Liu, Bing. 2010. Sentiment analysis and subjectivity. *Handbook of Natural Language Processing*, second edition. Nitin Indurkhya and Fred J. Damerau (Eds.). Boca Raton: CRC Press, 627-666.

Liu, Hugo, Henry Lieberman, and Ted Selker. 2003. A model of textual affect sensing using real-world knowledge *International Conference on Intelligent User Interfaces*, 125-132.

Llorà, Xavier, Kumara Sastry, David E. Goldberg, Abhimanyu Gupta, and Lalitha Lakshmi. 2005. Combating user fatigue in iGAs: Partial ordering, Support Vector Machines, and synthetic fitness *Proceedings of the Genetic and Evolutionary Computation Conference*.

Llorà, Xavier, Francesc Alías, Lluís Formiga, Kumara Sastry and David E. Goldberg. Evaluation consistency in iGAs: User contradictions as cycles in partial-ordering graphs IlliGAL TR No 2005022, University of Illinois at Urbana-Champaign.

Loeff, Nicolas, Cecilia Ovesdotter Alm, and David Forsyth. 2006. Discriminating image senses by clustering with multimodal features. *Proceedings of the 21st International Conference on Computational Linguistics and the 44th ACL, Sydney, Australia*, 547-554.

Lyons, John. 1977. *Semantics* volumes 1, 2. Cambridge: Cambridge University Press.

Mesthrie, Rajend, Joan Swann, Ana Deumert, and William Leap. 2009. *Introducing Sociolinguistics*, 2nd ed. Amsterdam: John Benjamins.

Mihalcea, Rada and Hugo Liu. 2006. A corpus-based approach to finding happiness. *AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs*, 139-144.

Picard, Rosalind W. 1997. *Affective Computing*. Cambridge, Massachusetts: MIT Press.

Sproat, Richard. 2010. *Language, Technology, and Society*. Oxford: Oxford University Press.

Täckström, Oscar. 2009. A literature survey of methods for analysis of subjective language. SICS Technical Report T2009:08, ISSN 1100-3154.

Wiebe, Janyce, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. 2004. Learning subjective language. *Journal of Computational Linguistics* 30 (3), 277-308.

Wilson, Theresa, Janyce Wiebe, and Paul Hoffman. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, 347-354.

Zaenen, Annie. 2006. Mark-up barking up the wrong tree. *Journal of Computational Linguistics* 32 (4), 577-580.