Sociocognitive Language Processing – Emphasising the Soft Factors

Björn W. Schuller and Michael F. McTear

Abstract Sociocognitive Language Processing (SCLP) is the idea of coping with everyday language, including slang and multi-lingual phrases and cultural aspects, and in particular with irony / sarcasm / humour, and paralinguistic information such as the physical and mental state and traits of the dialogue partner (e. g., affect, age groups, personality dimensions), and social aspects. By that, multimodal aspects such as facial expression, gestures or bodily behaviour should ideally be included in the analysis where possible. At the same time, SCLP can render future dialogue systems more 'chatty' by not only feeling natural but being truly emotionally and socially competent, ideally leading to a more symmetrical dialogue. For that, the computer should itself have a 'need for humour', an 'increase of familiarity', etc., i. e., enabling computers to experience or at least better understand emotions and personality such that they have 'a feel' for these concepts. Beyond these ideas, the broader idea of SCLP includes verbal behaviour analysis, a closer coupling between language understanding and generation incorporating social and affective information, and new language resources to meet these ends. By that, SCLP unites expertise from psychology, social sciences, and (natural) language processing. Here, we give a short introduction.

Björn W. Schuller
University of Passau, Chair of Complex and Intelligent Systems, Passau, Germany &
Imperial College London, Department of Computing, London, U. K. e-mail: schuller@IEEE.org

Michael F. McTear
University of Ulster, Computer Science Research Institute, Ulster, U. K.
e-mail: mf.mctear@ulster.ac.uk

1 Introduction

Sociocognitive Language Processing (SCLP)[1] can be considered as a term to respect 'soft factors' in communication. Likewise, it can be seen as a specific kind of and arguably also as an extension to the more traditional and broadly defined field of Natural Language Processing (NLP). the idea of coping with everyday language, including slang and multi-lingual phrases and cultural aspects, and in particular with irony / sarcasm / humour, and paralinguistic information such as the physical and mental state and traits of the dialogue partner (e. g., affect, age groups, personality dimensions), and social aspects. By that, multimodal aspects such as facial expression, gestures or bodily behaviour should ideally be included in the analysis where possible. At the same time, SCLP can render future dialogue systems more 'chatty' by not only feeling natural but being truly emotionally and socially competent, ideally leading to a more symmetrical dialogue. For that, the computer should itself have a 'need for humour', an 'increase of familiarity', etc., i. e., enabling computers to experience or at least better understand emotions and personality such that they understand these concepts. Beyond these ideas, the broader idea of SCLP includes verbal behaviour analysis, a closer coupling between language understanding and generation incorporating social and affective information, and new language resources to meet these ends. By that, SCLP links NLP expertise more closely with such of psychology, the social sciences, and related disciplines.

In this short paper, we will exemplify the principle by focusing on the analysis side of NLP, known as Natural Language Understanding (NLU). Further, we will limit the example to *spoken* language understanding (SLU). The principles do, however, similarly apply to Natural Language Generation (NLG) in any form.

2 Example: Spoken Language Understanding

In SLU, the text output by the speech-to-text (speech recognition) component is analysed in order to determine its meaning. Meaning can take various forms. In spoken dialogue systems frame-based representations consisting of sets of attribute-value pairs are used to capture the information in an utterance that is relevant to the application [13]. Thus, the representation of an utterance such as *book a flight to London on Friday* would be something like: *flight, destination = London, day = Friday*. In other cases, a deeper semantic representation is required that reflects subtle distinctions in meaning that are conveyed by the syntactic form of the utterance [1]. For example, there is only one word that is different in the following sentences (S) but the meaning representations should be different:

- S1: *List all employees of the companies who are based in the city centre*
- S2: *List all employees of the companies that are based in the city centre*

[1] As the field of *Spoken* Language Processing is usually abbreviated as SLP, we suggest S*C*LP as a short notation for Sociocognitive Language Processing.

The interpretation of S1 asks for a listing of employees who are based in the city centre while the interpretation of S2 asks for a listing of employees who are not necessarily based in the city centre but who work for companies based there. This difference can only be picked up by an analysis that reflects the difference between the use of *who* and *that* in such sentences. Up until the 1980s NLP was dominated by the knowledge-based (or symbolic) paradigm in which grammars were handcrafted and meaning was represented using logic-based formalisms. At this time the focus was on parsing written texts. In the 1990s a paradigm shift occurred in which probabilistic and data-driven models that had already been deployed successfully in speech recognition were now applied to language processing [31]. The availability of large corpora of spoken and written text led to an increasing use of machine learning techniques so that the previously handcrafted rules of the knowledge-based paradigm could be learnt automatically from labelled training data. At the same time, attention turned to SLU, for which statistical methods have proved to be more robust as they degrade gracefully with input that is previously unseen and potentially ill-formed. Currently, SLU focuses primarily on the textual form of a message and on its propositional content. However, additional information is conveyed in the prosodic features of a spoken utterance – its phrasing, pitch, loudness, tempo, and rhythm – that can indicate differences in the function of an utterance as well as expressing emotional aspects such as anger or surprise [29]. Other information to support the interpretation of an utterance may come from sensors that provide data about the environmental context, biosensors that report on the user's physical and emotional state, and machine vision systems that can detect non-verbal accompaniments of speech, such as gestures and facial expressions.

3 The Sociocognitive View

Above (cf. Section 2), we introduced SLU and showed that, in principle, works in this field *do* consider prosody, and often also multimodal information such as considering the facial expression of speakers. One may thus ask what makes the term SCLP different or justified. In [10], the authors find in two experiments that it seems plausible to consider language understanding as "a special case of social cognition". This is based on a model to predict an "interaction between the speaker's knowledge state and the listener's interpretation" [10]. Similarly, the authors in [8] attest the high relevance of "social, cognitive, situational, and contextual aspects" when dealing with language. Further, the fields of Affective Computing (AC) [21], Social Signal Processing (SSP) [20, 33] or Behavioural Signal Processing (BSP) [16], suggest consideration of affective or emotional and social cues for computing systems used, e .g., in human-computer interaction [17], or human dialogue analysis. In particular in speech analysis [24, 2, 6, 26] and synthesis [14, 15], such information was considered rather early. Further, sub-fields of NLP deal with such information, most noteworthy the discipline of Sentiment Analysis (SA) [18, 32, 11, 5, 28]. Including also acoustic speech feature information, Computational Paralinguistics (CP)

[27] provides a broader view on speaker states and traits beyond sentiment, emotion, or social signals including also biological traits such as age, gender, height, or race, personality traits, or health-related state and trait information (cf. also [9, 19]) alongside physiological states such as eating or exercising, and cognitive load.

Likewise, AC and SSP/BSP provide a general beyond-language paradigm for computer analysis and synthesis of affective and behavioural cues – each of which focusing on one of emotional or social intelligence; SA (and related sub-disciplines of NLP such as Opinion Mining) focus on the analysis of a single aspect – here sentiment – and neglect the synthesis side, and CP deals mainly with speech and language analysis and synthesis without a linking model or component such as a dialogue model. This makes NLP the definition that comes arguably closest to SCLP; however, while it aims to deal with 'natural' language, its emphasis on the soft factors of communication is rather weak. By SCLP, we advocate a strong link between language and the authoring or speaking person's state and trait as related to the spoken content and conveyed 'message'. Likewise, a decision such as in the example in Section 2 regarding the interpretation of S1 potentially in the sense of S2 could be supported by estimation of the social and cultural background of the speaker such as "native speaker" (or not) or the personality such as "conscientious" (then, likely taking it for the actual sense of S1).

4 Conclusion

In this short contribution, we introduced the term of Sociocognitive Language Processing (SCLP). We further motivated its introduction reviewing key relevant literature in a constructive and synthetic manner with the aim to highlight borders between related existing disciplines and terms such as NLP and SA, and at the same time confine SCLP as compared to broader fields such as AC and SSP.

We exemplified the considerations by language analysis looking at NLU and leading to a sociocognitive view. In the same vein, in NLG, the sociocognitive view lends more weight on the 'soft factors' in communication, such as synthesising irony, or nonverbal fillers, and behaviour that fit these and accompany the linguistic content. We believe that, by integrating SCLP principles, one can render future dialogue systems more 'chatty' making them not only feel 'natural' but truly emotionally and socially competent (cf., e. g., [25, 7]). Ideally, this will lead to a more 'symmetrical' dialogue where both ends – humans and computer systems or intelligent machines – will integrate and comprehend soft factors in the communication.

Arguably, for that, communicative technical systems should themselves have a 'need for humour' [22, 4, 30], an 'increase of familiarity' during repeated or prolonged interactions, etc.. In other words, it appears required for genuine SCLP to enable computers to experience and have or at least better understand emotions and personality, such that they have 'a feel' for these concepts (cf., e. g., [3]). For example, the degree of conscientiousness of a system might be the decisive factor

between taking S1 in our example in the sense of S1 or rather S2 in addition to its interpretation of 'what the user is like' (based also on increased familiarity) in relation to the best interpretation. This will, however, need to be further expanded upon in follow-up considerations and studies including 'Sociocognitive Dialogue Processing'. Beyond these ideas, the broader idea of SCLP includes verbal behaviour analysis, a closer coupling between language understanding and generation incorporating social and affective information, and new language resources to meet these ends. By that, SCLP unites expertise from psychology and social sciences with NLP on the way to enable genuine conversational dialogue systems [12] or emotionally and socially aware computer-mediated communication [23].

Acknowledgements The first author acknowledges funding from the ERC under grant agreement no. 338164 (iHEARu), and the European Union's Horizon 2020 Framework Programme under grant agreements nos. 645378 (ARIA-VALUSPA), 644632 (MixedEmotions), and 645094 (SEWA).

References

1. Allen, J., Manshadi, M., Dzikovska, M., Swift, M.: Deep linguistic processing for spoken dialogue systems. In: Proceedings of the Workshop on Deep Linguistic Processing, pp. 49–56. ACL, Prague (2007)
2. Bachorowski, J.A., Owren, M.J.: Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. Psychological science **6**(4), 219–224 (1995)
3. Ball, J.E., Breese, J.S.: Modeling and projecting emotion and personality from a computer user interface (2001). US Patent 6,212,502
4. Binsted, K., Bergen, B., O'Mara, D., Coulson, S., Nijholt, A., Stock, O., Strapparava, C., Ritchie, G., Manurung, R., Pain, H., Waller, A., O'Mara, D.: Computational humor. IEEE Intelligent Systems **21**(2), 59–69 (2006)
5. Cambria, E., Schuller, B., Xia, Y., Havasi, C.: New Avenues in Opinion Mining and Sentiment Analysis. IEEE Intelligent Systems Magazine **28**(2), 15–21 (2013)
6. Dellaert, F., Polzin, T., Waibel, A.: Recognizing emotion in speech. In: Proceedings of the Fourth International Conference on Spoken Language Processing, vol. 3, pp. 1970–1973. IEEE (1996)
7. DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Scherer, S., Stratou, G., Suri, A., Traum, D., Wood, R., Xu, Y., Rizzo, A., Morency, L.P.: Simsensei kiosk: A virtual human interviewer for healthcare decision support. In: Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems, pp. 1061–1068. International Foundation for Autonomous Agents and Multiagent Systems (2014)
8. Dietrich, R., Graumann, C.F.: Language processing in social context. Elsevier (2014)
9. Furnham, A.: Language and personality. In: H. Giles, P. Robinson (eds.) Handbook of language and social psychology. John Wiley & Sons (1990)
10. Goodman, N.D., Stuhlmüller, A.: Knowledge and implicature: Modeling language understanding as social cognition. Topics in cognitive science **5**(1), 173–184 (2013)
11. Liu, B.: Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies **5**(1), 1–167 (2012)
12. McTear, M.F.: Spoken dialogue technology: enabling the conversational user interface. ACM Computing Surveys (CSUR) **34**(1), 90–169 (2002)

13. Mesnil, G., Dauphin, Y., Yao, K., Bengio, Y., Deng, L., Hakkani-Tur, D., He, X., Heck, L., Tur, G., Yu, D., Zweig, G.: Using recurrent neural networks for slot filling in spoken language understanding. IEEE/ACM Transactions on Audio, Speech, and Language Processing **23**(3), 530–539 (2015)
14. Murray, I.R., Arnott, J.L.: Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. The Journal of the Acoustical Society of America **93**(2), 1097–1108 (1993)
15. Murray, I.R., Arnott, J.L.: Implementation and testing of a system for producing emotion-by-rule in synthetic speech. Speech Communication **16**(4), 369–390 (1995)
16. Narayanan, S., Georgiou, P.G.: Behavioral signal processing: Deriving human behavioral informatics from speech and language. Proceedings of the IEEE **101**(5), 1203–1233 (2013)
17. Paiva, A.: Affective interactions: toward a new generation of computer interfaces? Springer (2000)
18. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, pp. 79–86. ACL (2002)
19. Pennebaker, J.W., Graybeal, A.: Patterns of natural language use: Disclosure, personality, and social integration. Current Directions in Psychological Science **10**(3), 90–93 (2001)
20. Pentland, A.S.: Social signal processing [exploratory dsp]. Signal Processing Magazine, IEEE **24**(4), 108–111 (2007)
21. Picard, R.W., Picard, R.: Affective computing, vol. 252. MIT press Cambridge (1997)
22. Ritchie, G.: Current directions in computational humour. Artificial Intelligence Review **16**(2), 119–135 (2001)
23. Riva, G.: The sociocognitive psychology of computer-mediated communication: The present and future of technology-based interactions. Cyberpsychology & behavior **5**(6), 581–598 (2002)
24. Scherer, K.R., Banse, R., Wallbott, H.G., Goldbeck, T.: Vocal cues in emotion encoding and decoding. Motivation and emotion **15**(2), 123–148 (1991)
25. Schröder, M., Bevacqua, E., Cowie, R., Eyben, F., Gunes, H., Heylen, D., Maat, M.T., McKeown, G., Pammi, S., Pantic, M., Pelachaud, C., Schuller, B., de Sevin, E., Valstar, M., Wöllmer, M.: Building autonomous sensitive artificial listeners. IEEE Transactions on Affective Computing **3**(2), 165–183 (2012)
26. Schuller, B.: Towards intuitive speech interaction by the integration of emotional aspects. In: Proceedings IEEE International Conference on Systems, Man and Cybernetics, vol. 6. IEEE, Yasmine Hammamet, Tunisia (2002). 6 pages
27. Schuller, B., Batliner, A.: Computational paralinguistics: emotion, affect and personality in speech and language processing. John Wiley & Sons (2013)
28. Schuller, B., Mousa, A.E.D., Vasileios, V.: Sentiment Analysis and Opinion Mining: On Optimal Parameters and Performances. WIREs Data Mining and Knowledge Discovery **5**, 255–263 (2015)
29. Shriberg, E., Stolcke, A.: Prosody modeling for automatic speech recognition and understanding. In: Mathematical Foundations of Speech and Language Processing – The IMA Volumes in Mathematics and its Applications, vol. 138, pp. 105–114. Springer, New York (2004)
30. Strapparava, C., Stock, O., Mihalcea, R.: Computational humour. In: Emotion-oriented systems, pp. 609–634. Springer (2011)
31. Tur, G., De Mori, R.: Spoken language understanding: Systems for extracting semantic information from speech. John Wiley & Sons (2011)
32. Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th annual meeting on association for computational linguistics, pp. 417–424. ACL (2002)
33. Vinciarelli, A., Pantic, M., Bourlard, H., Pentland, A.: Social signal processing: state-of-the-art and future perspectives of an emerging domain. In: Proceedings of the 16th ACM international conference on Multimedia, pp. 1061–1070. ACM (2008)