Think before You Talk: The Role of Cognitive Science in Natural Language Processing

Amy Neustein

Linguistic Technology Systems, 800 Palisade Avenue, Suite 1809,
Fort Lee, NJ 07024 U.S.A.
amy.neustein@verizon.net

Abstract. This paper examines the advantages of invoking a cognitive science framework to resolve some of the persistent Natural Language Processing (NLP) problems that compromise the performance of natural language systems. I use the term "cognitive science" in its most *generic* sense – an interdisciplinary field of study concerned with how information is represented, processed and transformed – even though certain disciplines that may be useful to the design and fine-tuning of NLP programs, such as conversation analysis whose origins are in sociological studies of talk-in-interaction, are theoretically outside the scope of cognitive studies. The task of fully simulating human language use may prove awesome and impracticable even though we can reasonably point to nearly six decades of considerable progress with the development of computers that can interact with humans by using natural speech. Nevertheless, an interdisciplinary approach to NLP may bring us closer to this goal. Using a cognitive science approach, three different areas of NLP are analyzed in this paper: 1) Simulation of Human Language Use in Spoken Dialogue Systems; 2) Reference Generation and Referential Practices; and 3) Word Sense Disambiguation (WSD).

1 Introduction

While our global world requires sophisticated natural language processing (NLP) programs, there still exists noticeably poor integration of the various disciplines which prove essential in helping machines to better understand human (natural) language. These disciplines consist of the following: discourse analysis and computational linguistics; sociolinguistics, sociology and conversation analysis; speech act theory and linguistic philosophy; artificial intelligence and soft computing; psycholinguistics and cognitive psychology; and other branches and subfields of what can be broadly called human language studies. Given the broad applications of natural language technology (e.g., machine translation of web pages, extraction of medical information from electronic health records, IVR-driven semi or fully automated call centers, and speech enabled smartphones and other mobile devices), it is imperative that we take a step back and consider how to better integrate such diverse disciplines that have taken up the task of researching the interactions between computers and human language.

In this paper, I will demonstrate the advantage of invoking a cognitive science framework to resolve some of the persistent NLP problems that compromise the performance of natural language systems. I use the term "cognitive science" in its most *generic* sense – an interdisciplinary field of study concerned with how information is represented, processed and transformed – even though certain disciplines that may be useful to the design and fine-tuning of NLP programs, such as conversation analysis (whose origins are in sociological studies of talk-in-interaction), are theoretically outside the scope of cognitive studies.

2 An Interdisciplinary Approach to the Simulation of Human Language Use in Spoken Dialogue Systems

The task of fully simulating human language use may prove awesome and impracticable even though we can reasonably point to nearly six decades of considerable progress with the development of computers that can interact with humans by using natural speech [18]. Bel-Enguix and Jimenez-Lopez [3], writing for a special issue of the *International Journal of Speech Technology* which had been aptly titled "Rethinking Natural Language Processing for Speech Technology," showed that while it is difficult to "fully simulate human language use [one may] take advantage of research on human language in order to improve conversational interfaces" (p. 219). Using Conversational Grammar Systems, or CGS, the authors modeled dialogue as inter-action by looking at dialogue as "a sequence of acts performed by two or more agents in a common environment" (p. 209). The authors were at pains to show that given the "investigation and modeling of human language is clearly an *interdisciplinary task*...methods for language technology have to come from different disciplines" (emphasis supplied) (p. 219). To define their modeling of dialogue as inter-action, the authors drew from conversation analytic studies, which they combined with their knowledge of computational linguistics, formal language theory, and speech act theory.

For example, in addressing one of the key tasks of dialogue management, which is to "control when to speak and what to say," the authors made use of what they referred to as "the important selection techniques in dialogue such as [the] turn-taking system and adjacency pairs [an area of study that has been one of the main research foci of conversation analysts for nearly four decades]…formalized in CGS by introducing derivation modes that control how long an agent can act in the environmental state; by endowing agents with an *internal control* that contains start/stop conditions that allow agents to recognize places where they can start their activity, as well as places where they should stop their actions and give others the chance to act; and by defining mapping that fulfills in CGS a function analogous to the one carried out by adjacency pairs, reactive pressures or discourse expectations in their respective conversational models " (p. 219).

In another paper appearing in the same issue of the *Journal*, John Barnden [1] explored the problem presented by the use of metaphors, in both text and speech, when performing natural language processing tasks. He pointed out that while the problem of the metaphor may be viewed "as a peripheral problem (perhaps mostly to do with poetry and other literary language) it is in fact a pervasive feature of mundane language…" (p. 121). To make this point, Barnden reached across the methodological divide and buttressed his discourse analytic findings with the helpful discoveries of conversation analysts Paul Drew and Elizabeth Holt [4], who showed how speakers employ the art of metaphor to achieve topic transition in conversation.

Neustein [13-17] showed how the turn-taking features of talk may prove critical to the linguist's pragmatics-centered approach to human-computer dialogue modeling in spoken language systems. Let us consider for a moment multi-party dialogue systems [19] and the importance of the dialogue manager's correct reading of a "concessive connector" [11], also known as a "clue word" – "so," "anyway", "now," "but" – that connects the various utterances that comprise a multi-utterance speaking turn. Computational linguists [20] have introduced the term "context space" to describe the current set of things being talked about in any given dialogue and have shown how such "clue words" can be instrumental in effecting a *transition* between these spaces. Sidner [23] and Grosz and Sidner [5] use various focus registers and spaces to represent the focus of attention in a dialogue where shifts in focus are made by speakers making utterances that either add or subtract items from the current focus registers. McTear [10] has adapted some of these techniques for constructing natural language interfaces.

It may seem quite plausible for clue words or concessive connectors to link one utterance to the next in a multi-utterance turn, until one looks more closely at some of the more subtle features of the turn-taking process, which constitutes the topic of study explored in painstaking detail by conversation analysts for nearly four decades. Here is an example of how such certitude about connectors stringing together the parts of a multi-utterance speaking turn may begin to erode. Consider the following two scenarios where a speaker uses concessive connectors, each time for entirely different purposes.

In the first scenario, the current speaker, immediately after producing the concessive connector ("so"), displays a "holding" silence [7,8] – an abrupt silence accompanied by marked inhalation that indicates the speaker's intent to "hold" the turn and continue speaking. In the second scenario, however, the exact opposite occurs: the speaker's production of the concessive connector ("so") is followed by a "trail off" silence – a gradual silence accompanied by exhalation – the kind of silence that provides a clear transition relevance place for the other speaker to begin to speak [7,8]. What this shows is that a concessive connector or clue word does *not* always serve the purpose of connecting the utterances of a multi-utterance turn, because as demonstrated in the second scenario, the speaker's intent may be to yield his turn to the next speaker rather than to continue speaking. This is amply displayed by his use of a clue word followed by a "trail off" silence, which indicates the speaker's intent to relinquish his turn to the next speaker. Dialogue managers, to be able to competently assess the dialogue progress (which requires an accurate determination of when a speaking turn is about to end or, alternatively, when it is being continued) must be able to recognize some of the formal properties of talk-in-interaction as defined by conversation analysts, such as the differentiation between "holding" silences and "trail off" silences and what they signify for the ensuing turn-taking activities.

All in all, an interdisciplinary approach to simulating human language use may not only help to resolve some of the more challenging NLP problems facing designers of spoken dialogue systems, but may also serve as "checks and balances" for interpreting dialogue entries correctly in order to assist the dialogue manager in gauging what is happening at each turn.

3 Interactional Approaches to Reference Generation and Referential Practices

Both Zock et al. [24] and Moore [12] have studied the communicative competencies (pragmatic knowledge) associated with referential expressions, though each has approached this fascinating domain of study using a different methodological approach. Zock and his colleagues' work focus on foreign language learning: in particular the analysis of WebREG, a web-based application for learning referring expressions. This application presents a language learning experiment whose purpose is to remedy the hit or miss process in which individuals learn in the natural course (through their own experiences) what the "referring expression" [22] stands for. Moore's research reveals how referential practice is organized in the context of search engine interactions, in which users employ naming (or alternatively use generic descriptions if they don't know the name at first, and then formulate their subsequent search queries using the newly learned name) in their search queries to refer to the objects/subjects of their search. But whether the study of referential activity concerns foreign language learning or online search, both Zock et al. and Moore approach the employment of reference expressions as an *interactive* event.

What this means, if we take a look at the work of Zock et al. [24] for a moment, is that in the process of performing reference generation (a sub-task of language production), the speaker must move from his own "egocentric point of view…to the listener's position" so that referring expressions are aimed at the *listener's* frame of reference and cognitive state. Likewise, in Moore's [12] discussion of his research findings on users' GUI-driven online search, the author shows that just as speakers demonstrate that "choosing a form of reference in [conversational] interaction is a function of both the speaker's knowledge and the presumed knowledge of the recipient," web searchers display *interactional* competencies (found in conversational dialogue) in the course of their query formulations produced during an Internet search. They do this, first, by formulating their queries using names for the entity occupying their online search; second, by resorting to generic descriptions when they don't know the name of the entity in question; and third, use of the name, as opposed to generic descriptions, in *all* subsequent searches.

Conversation analysts Sacks and Schegloff [21] showed the "preferences" displayed by speakers in their organization of reference(s) in conversational dialogue. The first is "minimization [which] involves use of a single reference form"; the second is "recipient design [which] involves the preference for 'recognitionals'" (p. 15). In fact, this orderly practice of organizing reference is additionally displayed in conversational interaction when speakers use names "not only because the person is known but also in *preparation for subsequent use* in the conversation even when the person is not already known by the recipient/hearer" [emphasis supplied]. In this way, all future references will be made with the name that has been introduced by the other speaker which points to the organizational preference for minimization and recognitionals.

As shown in Moore's [12] study, when entering referring expressions into the search box, online search users display similar organizational practices for using referring expressions as they do when talking to other speakers. Like Zock et al. [24], who points to a web-based application for teaching foreign speakers how to perform reference generation through this special kind of experimental design (rather than relying on *random* learning about referential expressions), Moore points to several proposals that may be adopted by search engines to help users perform reference generation (presumably at the very beginning of their search) rather than risk the haphazard navigation of the search engine until the correct referential expression may be found. For example, Moore suggests that search engines may better accommodate to users' referential practices by trying to "include a link labeled 'I don't know what it's called?' which would activate an alternate mode of displaying results optimized for finding entity names." Whatever the final solutions may be for performing effective reference generation – whether in verbal interactions with other speakers or through query-related online search – both Zock et al. and Moore are on the right course with their application of an *interactional* approach to reference generation and referential practices which clearly takes into account the pragmatic knowledge and communicative competencies that are invoked when using referring expressions.

4 Combining Cognitive Science with Computational Strategies for Word Sense Disambiguation

Ion and Tufis [6] correctly state that "[the] most difficult problems in natural language processing stem from the inherently ambiguous nature of human languages" (p. 113). Kwong [9] carefully analyzed NLP's persistent problem of Word Sense Disambiguation (WSD), making the astute observation that the lack of synergy between computational and cognitive paradigms has resulted in "plateaued performance in the state-of-the-art [natural language] systems" in which "computational linguists have focused heavily, if not entirely, on the technical aspects of the systems, and few had attempted to give any serious or extensive account of the cognitive aspects of WSD" (p. vii). Kwong explains why this lapse cannot be dismissed too easily: "While individual systems could be fine-tuned by engineering the learning algorithms and the feature sets, the obstacle to overcoming the currently plateaued performance obtained from supervised systems may rest with some intrinsic properties of words and senses closely related to our cognition" (p. vii).

In assessing the deficiencies in today's approach to WSD, Kwong [9] points out that this wasn't always the case: "[I]f we look at the history of WSD, which is more than half a century now, computational linguists and psycholinguists had backed up each other in the early days, but they had then diverted from each other, more and more" (p. vii). She wisely suggests that one "take a step back to re-examine the computational strategies (by machines) and the cognitive strategies (by humans) for WSD in parallel, and to explore alternative classification senses which might shed light on their differential information susceptibility" (pp. vii-viii). To accomplish this difficult task Kwong proposes as a first step that some important questions be asked about lexical sensitivity when performing word sense disambiguation, given the fact that "WSD has…long been realized as a lexically sensitive task":

- If some knowledge sources contribute better to target words of a particular POS, then among target words of the same POS, do the same knowledge sources contribute in a similar way?
- If POS cannot adequately differentiate the contributions of the knowledge sources, are there other ways for classifying the words and senses which may better account for the lexical sensitivity?
- If such alternative classification(s) can be identified, will their explicit modeling in WSD systems improve disambiguation performance, allowing them to overcome the current plateau?

Using these unanswered questions to inspire investigation of new approaches to automatic WSD research, Kwong propounds *a lexically sensitive model of WSD* that would, for example, engage in "separating words, or more precisely word senses, into fairly distinct groups (or *sense types*) according to their responses to disambiguation with different knowledge sources" (p. 85). Such sense types "would be beyond simple linguistic categories like POS, and are likely to be more semantic and perceptual" (p. 85). Thus, by separating words into sense types, "the knowledge on the information susceptibility…(which refers to the relation between the intrinsic properties of a word and the effectiveness of various types of lexico-semantic knowledge to characterize and disambiguate it…[and which] is absent from existing lexical resources) of individual target words is important for fine-tuning WSD systems and informing the optimal combination of knowledge sources for disambiguation" (pp. 85-86).

What Kwong's innovative approach to the study of lexical sensitivity demonstrates is that a lexically sensitive model for WSD, one that combines both a cognitive and computational perspective, will better inform automatic systems with psycholinguistic evidence instead of "resting entirely and helplessly with specific machine learning algorithms and their feature selection mechanisms" (p. 92). In practical terms, the benefits of a cognitive approach to WSD may prove to be helpful to the resolution of a number of the major problems encountered with machine translation, such as word sense ambiguities that compromise the accuracy of the translation of text from one language to another, or sense ambiguities found in the performance of cross-lingual text mining, in which material that is subject to machine translation is extracted from a specific portion of the text.

There is yet another practical benefit of integrating cognitive strategies with computational approaches. In today's world we have many under-resourced languages and regional dialects, which naturally produce small quantities of parallel corpus data needed for machine translation. There are more than 6,000 languages and dialects in the world, whereas only a very small number of these have sufficient resources for the development of well-performing NLP systems. Basque, Gallego, Latvian, Khmer, Lao, and Swahili are among the thousands of such under-resourced languages and regional dialects. Since computational (statistical) models have come to depend on large amounts of corpus data to perform with a high level of recognition accuracy, languages that have woefully insufficient corpus data can hamper the building of NLP programs, which require large amounts of data to accomplish among other things the disambiguation of sense meaning of words. In spite of such impediments to machine learning, the under-resourced regions of the world are just as much in need of natural language technology as the more developed countries; they therefore require natural language processing to perform automatic machine translation and other related NLP tasks which can be easily compromised by limited parallel corpus data.

Cognitive paradigms for WSD may offer a forward-looking solution to building NLP programs for under-resourced languages. Because in providing cognitive solutions to the sense ambiguities in language, such programs can make up to some degree for the lack of large parallel data samples. Using semantic speech recognition in the Basque context, Barroso et al. [2] showed how "cross-lingual approaches…[and] data optimization and soft computing methods [that are] oriented to complex environment are used in order to overcome the lack of resources" in the Basque language. The author and her co-authors point out the useful benefits of their employing semantic speech recognition. "Nowadays, our work is oriented to Information Retrieval and mainly to small Internet mass media. In these cases the available resources for Basque in general, and for this task in particular, are very few and complex to process because of the noisy environment. Thus, the methods employed in this development (ontology-based approach or cross-lingual methodologies oriented to profit from more powerful languages) could suit the requirements of many under-resourced languages" (p. 33).

What we have seen from the examples given above is that in complementing computational (statistical) models of automatic machine translation and Information Retrieval, cognitive science may help NLP programs perform better in translating news broadcasts, newspaper articles, cartographical and travel information and other textual data, and help fill the gap between the desired amount of parallel corpus data and the limited parallel corpus data available. Certainly, this combined approach of pairing cognitive science with computational – statistical – methods for building machine translation programs can serve as a beacon of light in those regions of the world punctuated by under-resourced languages.

5 Conclusions

The nearly 60-year history of NLP has been marked by cyclical developments, as evidenced by the fluctuations in the advancement and recession of cognitive approaches to modeling of human language. Over the years, those in the field of NLP will have witnessed huge pendulum swings; at times cognitive modeling strategies have been placed on the front burner for practically every researcher of human language technology, while at other times such strategies have receded into the background so much so in fact that cognitive methods have been viewed in some sectors as the "step child" of computational modeling. Nonetheless, with all the advancements of statistical approaches to natural language modeling, some of the biggest hurdles to making NLP programs work effectively cannot be overcome without employing cognitive science, or, even more broadly, other disciplines and branches of human language sciences such as those that closely study the organizational properties of conversation interaction. The discussion above gives some examples of how cognitive and interactional studies of human language may complement computational methods for designing and improving natural language programs. Both the standard and the cutting edge technologies that use natural language deserve no less than a detailed and well thought out interdisciplinary approach, one that will enable NLP to work at its very best.

6 References

1. Barnden, J. A., 2008. Challenges in natural language understanding: the case of metaphor (commentary). International Journal of Speech Technology, 11(3-4): 121-123.
2. Barroso, N., Lopez de Ipina, K, Barroso, O, Ezeiza, A, Hernandez, C., and Grana, M., 2012. Semantic speech recognition in the Basque context, Part I: cross-lingual approaches. International Journal of Speech Technology, 15(1): 33-40.
3. Bel-Enguix, G., Jimenez-Lopez, M.D., 2008. Modelling dialog as inter-action. International Journal of Speech Technology, 11(3-4): 209-221.
4. Drew, P., Holt, E., 1998. Figures of speech: Figurative expressions and the management of topic transition in conversation. Language in Society, 27(4): 495-522.
5. Grosz, B. J., Sidner, C. L., 1986. Attention, intention and the structure of discourse. Computational Linguistics, 12(3): 175-204.
6. Ion, R., Tufis, D., 2009. Multilingual versus monolingual word sense disambiguation. International Journal of Speech Technology, 12(2-3): 113-124.
7. Jefferson, G., 1983. On a failed hypothesis: "Conjuctionals" as overlap vulnerable. Tilburg Papers in Language and Literature, 28: 29-33.
8. Jefferson, G., 1986. Notes on 'latency' in overlap onset. Human Studies, 9(2-3): 153-183.
9. Kwong, O., 2012. New Perspectives on Computational and Cognitive Strategies for Word Sense Disambiguation. SpringerBriefs, Series in Speech Technology (Neustein, A., Ed.) Springer-Verlag, Berlin Heidelberg New York.
10. McTear, M. F., 1987. The Articulate Computer. Blackwell, Oxford.
11. Moeschler, J., Reboul, A., 1994. Dictionnaire Encyclopedique de pragmatique. Seuil, Paris.
12. Moore, R. J., forthcoming. A Name is Worth a Thousand Pictures: Referential Practice in Search Engine Interactions. In A. Neustein and J. M. Markowitz (Eds.), Machine Talk: The Next Generation of Natural Language Processing and Speech Technology. Springer-Verlag, Berlin Heidelberg New York.
13. Neustein, A., 2001. Using sequence package analysis to improve natural language understanding. International Journal of Speech Technology, 4(1): 31-44.
14. Neustein, A., 2004. Sequence Package Analysis: a new natural language understanding method for performing data mining of help-line calls and doctor-patient interviews. In B. Sharp (Ed.), Proceedings of first international workshop on natural language understanding and cognitive science, ICEIS 2004, University of Portugal, (April 13) pp. 64-74.
15. Neustein, A., 2006. Sequence Package Analysis: A new natural language understanding method for improving human response in critical systems. International Journal of Speech Technology, 9(3-4): 109-120.
16. Neustein, A., 2007. Sequence Package Analysis: A new method for intelligent mining of patient dialog, blogs and help-line calls. Journal of Computers, 2(10): 45-51.
17. Neustein, A., 2011. Sequence Package Analysis and Soft Computing: Introducing a new hybrid method to adjust to the fluid and dynamic nature of human speech. In E. Corchado, V. Snasel, J. Sedano, A.E. Hassanien, J.L. Calvo, and D. Slezak (Eds.) Soft Computing Models in Industrial and Environmental Applications, 6th International Conference SOCO 2011: Advances in Intelligent and Soft Computing, Volume 87, Springer-Verlag, Berlin Heidelberg New York, pp. 1-10.
18. Pieraccini, R., 2012. The Voice in the Machine. MIT Press, Cambridge, Mass.

19. Popescu, V., Caelen, J., and Burileanu, D., 2009. A constraint satisfaction approach to context-sensitive utterance generation in multi-party dialogue systems. International Journal of Speech Technology, 12(2-3): 95-112.
20. Reichman, R., 1985. Getting Computers to Talk Like You and Me. MIT Press, Cambridge, Mass.
21. Sacks, H., Schegloff, E. A., 1979. Two preferences in the organization of reference to persons in conversation and their interaction. In G. Psathas (Ed.), Everyday Language: Studies in Ethnomethodology. Irvington Publishers, Inc., New York, pp. 15-21.
22. Searle, J., 1969. Speech acts: An essay in the philosophy of language. Cambridge University Press, Cambridge, UK.
23. Sidner, C. L., 1983. Focusing in the comprehension of definite anaphora. In M. Brady and R.C. Berwick (Eds.), Computational Models of Discourse. MIT Press, Cambridge, Mass., pp. 267-330.
24. Zock, M., Lapalme, G., and Yousfi-Monod, M., 2012. Learn to speak like normal people do: The case of object descriptions. In B. Sharp and M. Zock (Eds.), Proceedings of 9th International Workshop on Natural Language Processing and Cognitive Science. Wroclaw University of Economics, Wroclaw, Poland (June 28, 2012).