Limitations of the Philosophy of Language Understanding Implicit in Computational Linguistics

Graeme Hirst
Department of Computer Science
University of Toronto

Contemporary computational linguistics (CL) strives to be a strongly empirical science. It has long ago abandoned the *a priori*, intuitionistic and introspectionist methods of earlier days, which it inherited from theoretical linguistics, in which CL researchers would simply invent, from their own linguistic intuition, the examples that motivated their approach to a problem and the data on which their systems would be tested. But CL has not really abandoned intuitionism. Rather, it has replaced the intuition of the researcher with the intuition of the **annotator**. In a typical experiment, a corpus of naturally occurring text is annotated by human 'judges' as to some linguistic property that one would like a computer to be able to recognize; a machine-learning algorithm then uses this **training data** to induce a model from features of the text that the researcher hypothesizes are indicative of the property. Typically, the property of interest is not overt in the text and at least some degree of judgement is inherent in its recognition. For example, each occurrence of an ambiguous word in a text may be tagged with the particular sense in which it is being used; sentences in a review of a consumer product may be tagged as either criticism or praise.

While this has the obvious benefit of avoiding researcher bias, what hasn't changed is the implicit philosophy of language understanding in which there is a single linguistic reality, a single understanding or interpretation of a text or of its elements, which is open to native-speaker introspection or intuition. The tacit assumption remains that all competent native speakers of a language (or dialect) will have the same intuition and hence will annotate any given text or any given linguistic element within a text the same way; or, if they do not, the differences can be attributed either *(a)* to carelessness, ignorance, or error on the part of the annotator, or *(b)* to an unclarity, vagueness, or other deficiency in the definition of the required annotations. To guard against this, it's highly preferable that a text be annotated independently by more than one person. Ideally, their annotations will be identical. Following work originally done in the context of content analysis (Krippendorff 1980), CL has developed sophisticated statistical methods (reviewed by Artstein & Poesio 2008) for measuring the **inter-annotator agreement** on a text, usually expressed as the κ (kappa) statistic, and for determining whether or not it is satisfactory (taking into account what level of agreement would occur merely by chance in the particular task).

The assumptions of this methodology are challenged in two ways.

1. Reader-based views of meaning and language understanding. The position that there is a unique understanding embodied in a text is famously challenged

by postmodernists such as Stanley Fish and Roland Barthes, among others. In this view, because a reader cannot be certain of a writer's intentions, they instead bring their own knowledge and experience to the interpretation of the text, which is not necessarily the same as that of the writer of any other reader. In particular, Fish (1980) claims that when readers *do* agree on an understanding — that is, have the same response, or, from the CL perspective, annotate a text or a linguistic element the same way — it is because they are members of the same **interpretive community**. In this view, annotation can be successful only if the annotators are drawn from the same interpretive community and only insofar as the purpose to which the annotated data is put ultimately serves (only) this same interpretative community. (As Wittgenstein might have remarked, "If a lion could annotate text, we could not use it as training data".)

    2. Individual differences in cognitive language comprehension processes. It is well established that, quite apart from issues of interpretive communities and of each individual language user's knowledge and experience, there are individual differences in cognitive language comprehension processes and sometimes in the interpretations that are their outcome. For example, Bever, Townsend, and colleagues (summarized by Hauser and Bever 2008) have shown that, because of differences in lateralization of language function in the brain, right-handers who have left-handed family members differ qualitatively in language comprehension from those who do not — for example, in the degree to which they rely on syntax rather than words and meaning (Townsend, Carrithers, & Bever 2001). Mulligan (2006) found individual differences in the cognitive processing strategies by which her subjects drew inferences from texts. While it is possible that these different processes could nonetheless lead to identical outcomes in interpretation and understanding (within an interpretive community), and they surely do in most instances, this is not always the case. As Ross put it long ago (1979; and see more generally Fillmore, Kempler, & Wang 1979), a language has a **core** on which there is general agreement among speakers, a **fringe** of less agreement, and a **bog** of disagreement.[1] For example, Mendelsohn and Pearlmutter (1999) demonstrated that there are individual differences in preferences in relative clause attachment that correlate with size of working memory; Gleitman and Gleitman (1970) found that the use (or violation) of syntactic constraints in the interpretation of noun phrases correlated with educational level.

    These two challenges make predictions that are consistent with what we see. It can be quite difficult to obtain satisfactory inter-annotator agreement, or, more generally, agreement on the "truth" of the data, for many kinds of annotations, and the κ values reported are often depressingly low. But the annotated data is used anyway — nothing better is available — with mediocre results. One particularly notable example was in the 2005 PASCAL Challenge on Recognizing Textual Entailment (Da-

---

[1] I avoid using the word *idiolect* here, as it is often construed, for example in studies of style and authorship, merely as a set of personal **preferences** or idiosyncrasies (or consistent clear error) in the choice of words, phrases, or syntactic constructions.

gan, Glickman, & Magnini 2006), in which the task for competing systems was to determine, given a pair of sentences, whether the second is textually entailed by the first. The dataset used in the competition as 'the right answers' was criticized because many researchers simply did not agree with many of the judgements it embodied. But present-day research in computational linguistics remains largely oblivious to the problems of a methodology and a philosophy of language understanding that inherently limit the quality and utility of the systems that it builds.

The alternative is a methodology in which aspects of language that are subject to notable individual differences are indeed modelled on an individual basis. That is, CL needs to start taking the idea of **user modelling** more seriously again — as it did in the pre-empirical era (Hirst 2007). But the user model must now include a perspective on the user's linguistic behaviour. One element of this will be the linguistic aspects of the user's individual purposes in using the system; but it will also include linguistic aspects of the user's knowledge and experience, as might be inferred, for example, from observations of their prior reading, and even, to the extent that they can be determined, aspects of the user's language processing strategies. (I discuss elsewhere (Hirst 2008) why such systems would have a high utility.) Such a model would start as a generic model; it would then adapt and accommodate itself to the individual user, becoming more precise and refined. This could occur through feedback from the user (or even through explicit training) and also, over time, through inferences based on the system's passive observation of the user's actions.

Of course, such systems are a distant goal. The point here is that for them to be a goal at all, in order for their utility to be recognized (Hirst 2008), CL will have to change its philosophy of language understanding and recognize the limitations of the annotation-based methodology.

References

Artstein, Ron & Poesio, Massimo (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4), 555–596.

Dagan, Ido; Glickman, Oren; & Magnini, Bernardo (2006). The PASCAL Recognising Textual Entailment Challenge. In: Quiñonero-Candela, J.; Dagan, I.; Magnini, B.; & d'Alché-Buc, F. (editors), *Machine Learning Challenges*. Springer-Verlag (Lecture Notes in Computer Science 3944), 177–190.

Fillmore, Charles J.; Kempler, Daniel; and Wang, William S-Y. (1979). *Individual Differences in Language Ability and Language*, Academic Press, NY.

Fish, Stanley (1980). *Is There a Text in this Class? The Authority of Interpretive Communities.* Harvard University Press.

Gleitman, Lila and Gleitman, Henry (1970). *Phrase and paraphrase.* Norton, NY.

Haenggi, Dieter; Gernsbacher, Morton Ann, & Bolliger, Caroline M. (1994). Individual differences in situation-based inferencing during narrative text comprehension. In: van Oostendorp, Herre & Zwaan, Rolf A. (editors), *Naturalistic Text Comprehension,* Ablex, Norwood, NJ, 79–96.

Hauser, Marc D. & Bever, Thomas (2008). A biolinguistic agenda. *Science*, 332 (#5904), 14 November 2008, 1057–1059.

Hirst, Graeme (2007). Views of text-meaning in computational linguistics: Past, present, and future. In: Dodig-Crnkovic, Gordana & Stuart, Susan (editors), *Computation, Information, Cognition — The Nexus and the Liminal*, Cambridge Scholars Publishing, Newcastle-upon-Tyne, 270–279.

Hirst, Graeme (2008). The future of text-meaning in computational linguistics. In: Sojka, Petr; Horák, Aleš; Kopeček, Ivan; & Pala, Karel (editors), *Proceedings, 11th International Conference on Text, Speech and Dialogue (TSD 2008),* Brno, Czech Republic, Springer-Verlag (Lecture Notes in Artificial Intelligence 5246), 1–9.

Krippendorff, Klaus (1980). *Content Analysis: An Introduction to its Methodology.* Sage, Beverly Hills, CA.

Mendelsohn, Aurora & Pearlmutter, Neal (1999). Individual differences in relative clause attachment preferences. *CUNY Conference on Human Sentence Processing*, New York.

Mulligan, Elizabeth J. (2006). *Assessment of Individual Differences in the Components of Comprehension Skill.* PhD thesis, University of Colorado, Boulder.

Ross, John Robert (1979). Where's English? In Fillmore, Kempler & Wang, 1979, 127–163.

Townsend, David J.; Carrithers, Caroline; & Bever, Thomas G. (2001). Familial handedness and access to words, meaning, and syntax during sentence comprehension. *Brain and Language*, 78: 308–331.