Last Words

The Shrinking Horizons of Computational Linguistics

Ehud Reiter*
University of Aberdeen

1. The Problem

Understanding language is one of the great challenges of science, and language-related technology is one of the great opportunities of Information Technology. Consequently, many different kinds of researchers work on language issues. Within the computer science community, language is studied by the "ACL community," by which I mean researchers who regularly publish in Association for Computational Linguistics (ACL) venues, such as the journal *Computational Linguistics* and ACL conferences. But language-related research is also carried out by researchers in other areas of computer science, including knowledge representation, cognitive modeling, vision and robotics, and human–computer interaction communities. Additionally, there are even more people outside computer science who study language, including linguists, psycholinguists, philosophers, and sociolinguists.

This is fine; understanding language and developing language technology are huge problems, and it is very useful to have many research communities from diverse backgrounds working on language. This will be especially true if the different research communities are aware of each other, so they can share insights, observations, problems, and so forth.

Unfortunately, my impression is that the ACL community is much less interested in research with other language-related research communities than it used to be. This impression is mostly based on discussions I have had with researchers who are on the border between ACL and another language-research community. Several such people have told me that whereas ten years ago they occasionally submitted papers to ACL venues and attended ACL conferences, now they do not bother, because they believe that the ACL community has no interest in their research.

In attempt to quantify this insight, I have analyzed citations from papers published in *Computational Linguistics* in 1995 and in 2005. Specifically, I extracted all citations from *Computational Linguistics* (CL) articles (excluding book reviews) in these years to journal papers. I then classified the cited journal papers into one of the categories shown in Table 1; whenever possible this classification was based on the subject category assigned by ISI Journal Citation Reports (JCR) to the cited journal. For example, a citation of a paper in *Cognitive Science* would count as a psychology citation, since ISI JCR classifies *Cognitive Science* as "Psychology, Experimental." I counted citations myself, rather than relying on ISI JCR's count, as there were some mistakes in JCR's counting. I also created my own "other NLP and speech" classification (that is, references to speech and NLP

* Department of Computing Science, Aberdeen AB24 3UE, UK. Email: ereiter@csd.abdn.ac.uk.

Table 1
Percentage of citations to journal papers in different areas, from papers in *Computational Linguistics* (raw number of cites in parentheses). Recent cites means papers published within the past ten years.

	All cites		Recent cites	
Area	(2005) % (n)	(1995) % (n)	(2005) % (n)	(1995) % (n)
Computational Linguistics	37 (56)	26 (50)	44 (36)	37 (47)
other NLP and speech	16 (25)	9 (17)	23 (19)	13 (16)
other artificial intelligence	12 (18)	13 (25)	19 (15)	13 (17)
human-computer interaction	0 (0)	3 (5)	0 (0)	3 (4)
other computer science	11 (17)	19 (36)	4 (3)	8 (10)
linguistics	8 (13)	14 (26)	5 (4)	13 (16)
psychology and psycholinguistics	1 (2)	12 (23)	0 (0)	12 (15)
mathematics and statistics	10 (15)	3 (5)	2 (2)	0 (0)
other	5 (7)	2 (4)	2 (2)	1 (1)

journals other than *Computational Linguistics*), as there is no such category in ISI JCR. This includes journals such as *Computer Speech and Language*, *Natural Language Engineering*, and *IEEE Transactions on Speech and Audio Processing*. I did not look at citations to workshops or conference papers, partly in order to simplify the analysis task, and partly because I am not aware of an independent subject classification of conferences and workshops similar to ISI JCR's journal classifications. I suspect that the main effect of including conference and workshop citations would be to decrease the percentage of citations to non-CS areas (psychology, linguistics, mathematics, other), because CS researchers are much more likely to publish important findings in conferences than researchers in other disciplines.

Table 1 shows the numbers of citations from CL papers to journals in different subject areas. It also shows the number of citations to papers that are less than ten years old; this gives us an idea as to whether CL authors are influenced by recent work in other areas. I draw the following conclusions from this table:

- CL authors were generally more inward looking in 2005 than in 1995; 53% of journal citations in 2005 were to NLP and speech journals, compared to only 35% in 1995.
- CL authors were much less likely to cite psychology and HCI journals in 2005 than in 1995: 1% of citations in 2005, compared to 15% in 1995.
- CL authors were less likely to cite recent work in linguistics and journals in 2005 than in 1995: 5% of recent citations in 2005, compared to 13% in 1995.
- CL authors were more likely to cite mathematical and statistical journals in 2005 than in 1995: 10% of citations in 2005, compared to 3% in 1995.

In short, 2005 CL authors were much less influenced by other research communities investigating language than 1995 CL authors. I find the steep drop in citations of psychology research to be especially surprising, as the past ten years has seen great progress made in psychology and psycholinguistics, partially driven by more widespread availability of new technology such as eye trackers and brain scanners.

I also looked at citations from 2005 papers published in *Natural Language Engineering*, which is another journal in which I believe many members of the ACL community publish. The citation pattern from 2005 NLE papers is overall quite similar to the citation pattern from 2005 CL papers. In particular, looking at references to other language research communities, 2005 NLE papers have a few more references than 2005 CL papers to papers in psychology journals (3% vs. 1%) and HCI journals (2% vs. 0%), but on the other hand have fewer references to papers in linguistic journals (5% vs. 8%). I could not examine citations from NLE papers in 1995, as the journal did not exist then.

I have not analyzed papers in ACL conferences in this way, but certainly my impression is that the situation here is similar, and that most ACL conference papers do not make many references to papers from other language research communities.

2. Impact

Does this matter? The ACL community seems to be focusing more on specific niches of the "language research" space, such as low-level syntactic analyses based on statistical corpus-based techniques. This is certainly an important niche, and progress in it will benefit the larger language community; so perhaps the ACL community is right to ignore the larger community and focus on making progress in its specific niches?

This is a logically defensible position, but I for one feel that it is a mistake, and in particular such a focus makes ACL conferences and *Computational Linguistics* less intellectually exciting and vigorous. This inward focus also goes against the belief in the larger scientific community that we need more inter-disciplinary work, and more interaction between researchers coming from different backgrounds.

The lack of references from papers published in ACL venues to work in the broader language community also decreases the awareness of the broader community in ACL research. This is because many researchers track references to their papers using citation-tracking services such as Google Scholar. Hence a side effect of researcher A citing researcher B is that researcher B is likely to become aware of researcher A's work.

I am also disappointed that *Computational Linguistics* and ACL conferences do not publish a broader range of papers on "problems involving natural language and computation" (especially as the ACL Web site states that ACL is "the international scientific and professional society for people working on problems involving natural language and computation"). In Table 2, I list recent language-related articles published in *Artificial Intelligence* (which is the most prestigious general AI journal). These articles address topics that I rarely see in ACL venues these days (but which I did see 10–20 years ago), such as user modeling, knowledge representation, integration of linguistic and visual information, and computational cognitive modeling. These articles, incidentally, have a very different citation pattern from *Computational Linguistics* papers (Table 1); only 11% of journal citations are to NLP and speech journals, whereas 31% are to AI journals, and 25% (!!) are to psychology journals. In other words, language-related papers published in *Artificial Intelligence* (unlike papers in *Computational Linguistics*) show considerable awareness of the broader language research community.

The last eight articles in Table 2, in fact, appeared in a special issue on Connecting Language to the World, which I co-edited (Roy and Reiter 2005). When I organized this

Table 2
Language-related articles in *Artificial Intelligence* published in August 2005 and August 2006.

Title	Selected keywords
Generating and evaluating evaluative arguments	user tailoring, preferences
Linguistic quantifiers modeled by Sugano integrals	knowledge representation, fuzzy logic
Discovering the linear writing order of a two-dimensional ancient hieroglyphic script	deciphering, unsupervised learning
Explorations in engagement for humans and robots	dialogue, human-robot interaction
Connecting language to the world (survey)	grounding language
Word sense disambiguation with pictures	image auto-annotation
Learning to talk about events from narrated video in a construction grammar framework	language acquisition
Dynamically structuring, updating, and interrelating representations of visual and linguistic discourse context	reference generation and resolution, cross-modal representations, salience
Protocols from perceptual observations	symbol grounding
Choosing words in computer-generated weather forecasts	information presentation, idiolect
Semiotic schemas: A framework for grounding language in action and perception	situated language, cross-modal
The emergence of compositional structures in perceptually grounded language games	language evolution, grammar induction

special issue, I originally suggested having it in *Computational Linguistics*, but I switched to *Artificial Intelligence* because many potential authors told me that they had no interest in publishing in *Computational Linguistics*, because it was not a journal that they and their research colleagues read. In other words, they did not see *Computational Linguistics* as a venue that published papers about their kind of research.

3. Solutions?

If this is indeed a problem, what can be done about it? I believe that one useful initial step would be to broaden the range of papers in ACL conferences and *Computational Linguistics* (I believe that many ACL members would like this to happen in any case). This could be done by making the paper-reviewing process less conservative (Church 2005); by putting more emphasis on the poster track and less on the full-paper track at ACL conferences (because the poster track tends to be broader than the full-paper track); and by including a broader range of people in journal editorial boards and conference program committees.

Another useful step would be to include a broader range of workshops, tutorials, and invited talks at ACL conferences and survey articles in *Computational Linguistics* on other areas of language-related research (I was very pleased to see a recent announcement that CL intended to publish such articles).

I believe that such measures are worth trying, because both the ACL community and the language-research community in general would benefit if the ACL community took more interest in the broader community, as indeed it did ten years ago.

References

Church, Ken. 2005. Reviewing the reviewers. *Computational Linguistics*, 31:575–578.

Roy, Deb and Ehud Reiter. 2005. Connecting language to the world. *Artificial Intelligence*, 167:1–12.