Chapter 5
Lexical Challenges in the Intersection of Applied Linguistics and ANLP

Scott Jarvis
Ohio University, USA

ABSTRACT

The investigation of natural language processing in the field of Applied Linguistics is pursued with both theoretical and practical aims, such as arriving at a clearer understanding of the nature of language knowledge, the rules that govern its use, how it is acquired, how unique it is to individual speakers, and how it can best be taught to learners. The purpose of this chapter is to draw attention to some of the prominent areas of overlap between Applied Linguistics and ANLP, highlighting the problems they face in relation to the characterization of lexical deployment, and focusing particularly on challenges related to the measurement of lexical diversity and the representation of the unique lexical signatures of individual samples of natural language use. The bulk of the chapter is devoted to describing preliminary solutions to these challenges.

INTRODUCTION

It is useful to begin this chapter with a definition of what Applied Linguistics means and which areas of research it entails. This turns out not to be such a straightforward task given that Applied Linguistics seems to mean different things to different people. At one extreme, one prominent scholar in the field has told me personally that he sees Applied Linguistics as being limited to the pursuit of improved methods for language learning and language teaching. This definition emphasizes the *applied* term in the field's name, but overlooks applications not directly linked to

language learning and teaching. On the opposite extreme, Applied Linguistics is understood by some people as encompassing all questions related to language acquisition, language pedagogy, language assessment, language use, and language practices and policies that are not purely theoretical (e.g., which at least involve some sort of data collection and analysis and/or attempts to derive practical applications). The scope of the annual meeting of the American Association for Applied Linguistics is closer to this latter extreme, as it hosts topic strands in numerous areas including but not limited to language education and pedagogy, language assessment and evaluation, first and second language acquisition and attrition, research on language cognition and neurolinguistics, discourse analysis, text analysis, sociolinguistics, language policy and planning, language and technology, and translation and interpretation.

The range of meanings associated with Applied Linguistics can be described as follows, with 1 representing the narrowest meaning, and 4 representing the broadest meaning:

1. emphasis on *applied*: the pursuit of improved language learning and language teaching methods.
2. emphasis on *linguistics*: research in subfields of linguistics that investigate language acquisition and use (e.g., first language acquisition, second language acquisition, sociolinguistics, corpus linguistics, forensic linguistics, language assessment).
3. emphasis on interdisciplinarity: all of the above, plus areas of research in other fields (e.g., psychology, speech pathology, literature, computer science, information science) that develop and use linguistic measures as indicators of other phenomena or that concentrate on the understanding or improvement of language-related knowledge and abilities.
4. liberal definition: the pursuit of any questions that are language-related and not purely

theoretical, and/or the pursuit of practical applications that are language-related.

In my own use of the term, the scope of Applied Linguistics is somewhere between 2 and 3. I see Applied Linguistics as being part of the larger field of Linguistics, but I also see it as extending to other disciplines that share with Linguistics a common understanding of the nature of language, and which rely on tools and concepts from the field of Linguistics in their research designs, analysis, and interpretations. I recognize the value of purely practical pursuits, and also the value of language-related research that is not grounded in the conventions of the field of Linguistics, but in my mind, Applied Linguistics proper involves empirical research and is grounded in theories and conventions of the field of Linguistics.

Concerning the intersection (or areas of overlap) between Applied Linguistics and Applied Natural Language Processing (ANLP), I see this as including areas of empirical investigation on language acquisition and use that involve the examination and analysis of human language samples with the aid of computer technology, and which are grounded in theories and conventions of the field of Linguistics.

Among the more prominent areas of overlap between Applied Linguistics and ANLP are the areas of research known as corpus linguistics, learner corpus linguistics, stylometry, and lexical richness research. These are the areas of research I will draw attention to in the present chapter. The purpose of this chapter is to describe some of the issues and questions that are being addressed in these areas of research, and particularly to characterize the challenges that researchers continue to face in these areas regarding lexical deployment. The bulk of the chapter is then devoted to describing the nature of these problems and suggestions for possible solutions.

Related Content

Formal Semantics for Metamodel-Based Domain Specific Languages

Paolo Arcaini, Angelo Gargantini, Elvinia Riccobeneand Patrizia Scandurra (2014). *Computational Linguistics: Concepts, Methodologies, Tools, and Applications (pp. 297-323).*

www.irma-international.org/chapter/formal-semantics-for-metamodel-based-domain-specific-languages/108727

Learning Words from Experience: An Integrated Framework

Annette M. E. Hendersonand Mark A. Sabbagh (2014). *Computational Linguistics: Concepts, Methodologies, Tools, and Applications (pp. 1705-1727).*

www.irma-international.org/chapter/learning-words-from-experience/108801

The Impact of Deep Learning on the Semantic Machine Learning Representation

Abdul Kader Saiodand Darelle van Greunen (2021). *Advanced Concepts, Methods, and Applications in Semantic Computing (pp. 22-49).*

www.irma-international.org/chapter/the-impact-of-deep-learning-on-the-semantic-machine-learning-representation/271119

Lip Modelling and Segmentation

A. Caplier, S. Stillittano, C. Bouvierand P. Y. Coulon (2009). *Visual Speech Recognition: Lip Segmentation and Mapping (pp. 70-127).*

www.irma-international.org/chapter/lip-modelling-segmentation/31065

Predicting Credit Rating Migration Employing Neural Network Models

Michael D'Rosarioand Calvin Hsieh (2020). *Natural Language Processing: Concepts, Methodologies, Tools, and Applications (pp. 65-82).*

www.irma-international.org/chapter/predicting-credit-rating-migration-employing-neural-network-models/239931