

Semantic Analysis of Learner Sentences

Levi King, Dissertation Prospectus

Research from the field of second language studies indicates that second language learners benefit from communicative and task based learning methods [Celce-Murcia, 1991, 2002, Larsen-Freeman, 2002, Ellis, 2006]. In the field of intelligent computer assisted language learning (ICALL), many applications overlook these findings and focus instead on grammar instruction and correction or menu based choices, approaches which are less beneficial to learners [Bailey and Meurers, 2008, Amaral and Meurers, 2011]. The thesis proposed here is motivated by this disconnect. The research will focus on assessment of the appropriateness of English non-native speakers' (NNSs) responses to a picture description task (PDT) by comparing them to native speakers' (NSs) responses via an evaluation system constructed with existing language resources and natural language processing (NLP) tools. By varying the degree to which responses are restricted by the PDT, this work will examine the effect of the elicitation task on automatic semantic assessment and thus explore the limits of current NLP tools for related ICALL and testing uses.

Background. This work sits at the intersection of NLP, ICALL and language testing, and related work from these fields will be examined. My own previous work on this task used a markedly different method than what is proposed here; it used only one type of PDT and parsed responses with an existing tool, then used custom rules for extracting *subject-verb-object* semantic triples and attempted to match these NNS triples against NS triples (see King and Dickinson [2013, 2014]). I will present the findings from this past work, such as the need to explore more restricted tasks and to move from rule-based to probabilistic approaches.

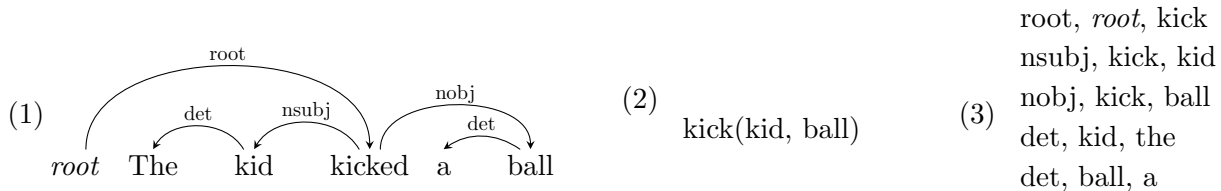
Research Questions. In this work I will explore the following research questions, which primarily relate to NNS variation, how it compares to NS variation and the tools necessary to automatically analyze contextual meaning in the face of such variation.

1. Are the responses of intermediate and advanced L2 English learners sufficiently *similar* to those of NSs to allow automatic evaluation based on a collection of NS responses? Do learners demonstrate significant overlap with native-like usage in a PDT setting?
2. In the constrained visual environment of a PDT, what are appropriate response representations for the purpose of providing meaning-oriented analysis? In other words, which linguistic components are crucial and which are superfluous?
3. Can NLP tools and language resources be successfully integrated to form a content analysis system for open response language learning tasks?
4. What does it mean for a response to be appropriate for this task and how can this be captured with an annotation scheme?

Data. In the current work, PDTs serve not only as a research tool but as a proxy for language use in visual settings, extending the impact of this work beyond ICALL and second language testing and into many areas of NLP where contextual NNS language may require

processing, such as dialog systems, machine translation and gaming. PDT responses will be collected primarily from ESL students through an arrangement with the English Language Improvement Program (ELIP) at Indiana University. The PDT will include items which rely on verbal or visual cues for targeted elicitation as well as less restricted versions of the items, allowing for an evaluation of the approach itself across task settings. I will also develop an annotation scheme for responses (in order to evaluate the system’s performance).

Method. First, to reduce spelling errors, responses will be preprocessed with existing tools. A major focus of the research will be establishing representations of NNS sentences and a corresponding gold standard (GS) derived from NS responses with which the NNS responses can be compared. Previous attempts used an existing tool [de Marneffe et al., 2006, Klein and Manning, 2003] to obtain a dependency parse (1), which provides labeled grammatical relations such as *subject* between word pairs, and extracted individual words into a semantic triple (2). Among other potential representations, the current work will instead convert these parses into lists of dependencies (3). This richer representation captures linguistic information not utilized in the previous work.



Unlike other ICALL systems that offer menu-based activities or simply analyze a user’s grammar, the current system and its GS will need to handle novel responses and focus on content over form. The GS for a given PDT item will be a collection of representations of individual NS responses to that item. In addition, I will explore different possibilities for automatically improving coverage of NNS responses. For example, existing synonym lists could be used to expand the dependencies in (3) by replacing *kid* with *child*.

I will explore various methods of comparing the NNS responses to the GS in order to find the method or methods that most consistently score responses in agreement with the human annotation. These methods will consider the frequencies of words and dependencies in the GS and NNS response. Comparisons will also use techniques from tasks like automatic indexing to upweight the most important words and dependencies.

This work will bring insights about the nature of NNS and NS variation in visual settings to the fields of language testing, ICALL and second language studies. More broadly, it will demonstrate methods for meaning-based analysis of contextual sentences, particularly when given an incomplete GS and potential errors in spelling and grammar. These findings will be valuable in tasks like image captioning, information retrieval and conversational agents.

References

- Luiz Amaral and Detmar Meurers. On using intelligent computer-assisted language learning in real-life foreign language teaching and learning. *ReCALL*, 23(1):4–24, 2011. URL <http://purl.org/dm/papers/amaral-meurers-11.html>.
- Kevin Atkinson. GNU Aspell. <http://aspell.net>, 1998.
- Stacey Bailey and Detmar Meurers. Diagnosing meaning errors in short answers to reading comprehension questions. In *The 3rd Workshop on Innovative Use of NLP for Building Educational Applications (ACL08-NLP-Education)*, pages 107–115, Columbus, OH, 2008. <http://aclweb.org/anthology-new/W/W08/W08-0913.pdf>.
- Marianne Celce-Murcia. Grammar pedagogy in second and foreign language teaching. *TESOL Quarterly*, 25:459–480, 1991.
- Marianne Celce-Murcia. Why it makes sense to teach grammar through context and through discourse. In Eli Hinkel and Sandra Fotos, editors, *New perspectives on grammar teaching in second language classrooms*, pages 119–134. Lawrence Erlbaum, Mahwah, NJ, 2002.
- Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC 2006*, Genoa, Italy, 2006.
- Rod Ellis. Task-based research and language pedagogy. *Language Teaching Research*, 4(3): 193–220, 2000.
- Rod Ellis. Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, 40:83–107, 2006.
- Katrina Forbes-McKay and Annalena Venneri. Detecting subtle spontaneous language decline in early Alzheimer’s disease with a picture description task. *Neurological Sciences*, 26(4): 243–254, 2005.
- Levi King and Markus Dickinson. Shallow semantic analysis of interactive learner sentences. In *Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 11–21, Atlanta, Georgia, June 2013. URL <http://www.aclweb.org/anthology/W13-1702>.
- Levi King and Markus Dickinson. Leveraging known semantics for spelling correction. In *Proceedings of the Third Workshop on NLP for Computer-Assisted Language Learning*, pages 43–58, Uppsala, Sweden, November 2014. URL <http://www.aclweb.org/anthology/W14-3504>.
- Dan Klein and Christopher D. Manning. Accurate unlexicalized parsing. In *Proceedings of ACL-03*, Sapporo, Japan, 2003.
- Dom Lachowicz. Enchant. <http://abisource.com/projects/enchant>, 2003.

- Diane Larsen-Freeman. Teaching grammar. In Diane Celce-Murcia, editor, *Teaching English as a second or foreign language*, pages 251–266. Heinle & Heinle, Boston, 3rd edition, 2002.
- Swapna Somasundaran and Martin Chodorow. Automated measures of specific vocabulary knowledge from constructed responses (‘Use these words to write a sentence based on this picture’). In *Proceedings of the Ninth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 1–11, Baltimore, Maryland, June 2014. URL <http://www.aclweb.org/anthology/W14-1801>.
- Swapna Somasundaran, Chong Min Lee, Martin Chodorow, and Xinhao Wang. Automated scoring of picture-based story narration. In *Proceedings of the 10th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 42–48, 2015.