On Automatically Analyzing Learner Language: Interpreting Form and Meaning in Context

Detmar Meurers
Universität Tübingen

Colloquium of the Research Center for English and Applied Linguistics (RCEAL)
Cambridge, February 8, 2011

Overview

- Points of contact: Analyzing learner language and computational linguistics
- Some issues in linguistic modeling of learner language
  - Parts-of-speech as an example
    - sources of evidence
    - nature of categories
  - Which level of analysis?
    - between robustness and representing variation
  - Target hypotheses and error annotation
    - Inter-annotator agreement and gold-standards
    - Comparative fallacy
  - Relevance of the task and learner modeling
- How about analyzing meaning?

CoMiC Automatically evaluating the meaning of learner responses to reading comprehension questions.

Contact Points with Computational Linguistics

- **Learner corpora**: representing, annotating, searching
  - can provide empirical evidence for SLA research
  - can provide insights into typical student needs in FLT

  annotation = off-line analysis
- **Writer's aid tools**: on-line analysis of learner language to provide immediate feedback *aimed at producing text*
- **Language testing**: off-line or on-line analysis to support or automate *assessment of learner abilities*
- **Intelligent Tutoring Systems**: on-line analysis
  - to provide immediate, individualized feedback, e.g.:
    - meta-linguistic feedback in a form-focused activity
    - incidental focus-on-form in a meaning-based activity
    - feedback on meaning (very rare in ITS)
  - to determine progression through pedagogical material

  aimed at supporting language acquisition.

Data in SLA research

An example: Clahsen & Muysken (1986)

- They studied word order acquisition in German by native speakers of Romance languages.
- Stages of acquisition:

1. S (Aux) V O	4. XP V[+fin] S O
2. (AdvP/PP) S (Aux) V O	5. S V[+fin] (Adv) O
3. S V[+fin] O V[-fin]	6. *dass* S O V[+fin]

Stage 2 example: *Früher* *ich kannte* *den Mann*
earlier$_{AdvP}$ I$_S$ knew$_V$ [the man]$_O$

Stage 4 example: *Früher* *kannte* *ich den Mann*
earlier$_{AdvP}$ knew$_{V[+fin]}$ I$_S$ [the man]$_O$

- How is the data characterized?
  - lexical and syntactic categories and functions
  - some acquisition stages are well-formed, others ill-formed

Annotation: Error Annotation and Beyond

- SLA research essentially observes correlations of linguistic properties, whether erroneous or not.
- Yet, the annotation of learner corpora has focused on *errors* made by the learners (cf. Granger 2003; Díaz-Negrillo & Fernández-Domínguez 2006).
- Even where errors are the research focus, their correlation with other linguistic properties is relevant.
- A wide range of linguistic modeling useful for capturing
  - overuse/underuse of particular patterns
  - measures of language development
    - CAF (Wolfe-Quintero et al. 1998; Ortega 2003; Housen & Kuiken 2009; Lu 2010)
    - Criterial Features (Hawkins & Buttery 2009, 2010)

Annotation of Linguistic Properties

- Annotation schemes have been developed for a wide range of linguistic properties, including
  - part-of-speech and morphology
  - syntactic constituency or lexical dependency structures
  - semantics (word senses, coreference), discourse structure
- Each type of annotation typically requires an extensive manual annotation effort → gold standard corpora
- Automatic annotation tools learning from such gold standard annotation are becoming available, but
  - quality of automatic annotation drops significantly for text differing from the gold standard training material
- Interdisciplinary collaboration between SLA & CL crucial to adapt annotation schemes & methods to learner language
  - Surprisingly little research on this (Meunier 1998; de Haan 2000; de Mönnink 2000; van Rooy & Schäfer 2002, 2003).

Annotation quality

- An annotation scheme is only as good as the distinctions it reliably supports making based on available evidence.
  - E.g., particle vs. preposition dropped in Penn Treebank tagset since often not enough evidence available.
  - Note: More classes can be more reliable if they are more coherent (cf. CLAWS7 annotation, followed by mapping to CLAWS5 in BNC Tag Enhancement Project).
- How can high quality gold standards be obtained?
  - Keep only reliably and consistently identifiable distinctions, described in detailed manual, including appendix on hard cases (Voutilainen & Järvinen 1995; Sampson & Babarczy 2003)
  - Annotate corpus several times and independently, then test interannotator agreement (Brants & Skut 1998)
  - Detection of annotation errors through automatic analysis of comparable data recurring in the corpus → DECCA (Dickinson & Meurers 2003a,b, 2005; Boyd et al. 2008)

Linguistically annotating learner language

Parts-of-speech as an example

- The NOCE learner corpus (Díaz-Negrillo 2009)
  - Short essays written by Spanish 1st and 2nd year students of English, annotated with editing and error tags
  - 998 texts, 337.332 tokens (149.256 types)

⇒ How about adding linguistic information?
(Díaz-Negrillo, Meurers, Valera & Wunsch 2010)

- Exploring automatic POS annotation
- What does it mean to POS-annotate learner language?

Automatic POS-Tagging of NOCE

- Used 3 POS taggers trained on WSJ newspaper text, using Penn Treebank tagset (TreeTagger, TnT, Stanford)
- Manually evaluated POS tags assigned by taggers to 10 texts by 10 different participants (1.850 words)
  - TreeTagger: 94.95%
  - TnT Tagger: 94.03%
  - Stanford Tagger: 88.11%

Aspects of a qualitative analysis

- We found lower performance for expressions which do not exist in English (cf. also de Haan 2000; van Rooy & Schäfer 2002).

(1) *I think that university teachs to people . . .* [spelling]

(2) *They can't pay their studies and more over they have to pay a flat . . .* [tokenization]

- But is tagging learner language really just a robustness issue, like adapting taggers to another domain?
- What does it mean to use POS tags developed for native language for the interlanguage of learners?
  - What research questions can such POS tags answer?

Three Sources of Evidence for POS analysis

Lemma/Lexical entry: *of* ⇒ preposition

(3) *drugs can be killer **of** many of ours.*

Morphology: *-ion* ⇒ noun

(4) *but it was a **revolution** in that period*

Distribution: det __ noun ⇒ adjective

(5) *In the **modern** life the people can communicate*

Case 1: Stem-Distribution mismatch

Stem Distribution Morphology

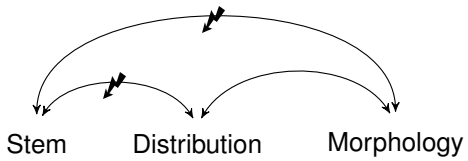
(6) [. . .] *you can find a big **vary** of beautiful beaches* [. . .]

Stem	Distribution	Morphology
verb	noun	?

(7) *RED helped him **during** he was in the prison.*

Stem	Distribution	Morphology
preposition	conjunction	?

Case 2: Stem-Distrib./Stem-Morph. mismatch



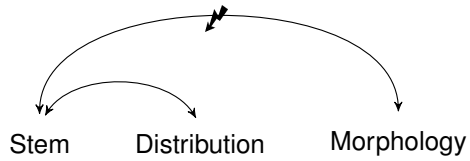
(8) […] *one of the favourite places to visit for many* ***foreigns***.

Stem	Distribution	Morphology
adjective	noun	noun / verb 3rd sg

(9) […] *to be* ***choiced*** *for a job* […]

Stem	Distribution	Morphology
noun / adjective	verb	verb

Case 3: Stem-Morphology mismatch



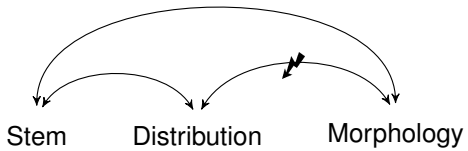
(10) […] *this film is one of the* ***bests*** *ever* […]

Stem	Distribution	Morphology
adjective (noun / verb)	adjective	noun / verb 3rd sg

(11) […] *television, radio are very* ***subjectives*** […]

Stem	Distribution	Morphology
adjective / noun	adjective	noun / verb 3rd sg

Case 4: Distribution-Morphology mismatch



(12) […] *for almost every* ***jobs*** *nowadays* […]

Stem	Distribution	Morphology
noun	noun sg	noun pl / verb 3rd sg

(13) […] *it has* ***grew*** *up a lot specially after 1996* […]

Stem	Distribution	Morphology
verb	verb past participle	verb past tense

Systematic POS for Learner Language

- A single, standard POS tag **fails to systematically identify properties of learner language**.
- Alternative: tripartite POS encoding of
  - distribution, stem, morphology
- Some errors in learner language are **epiphenomena of mismatches** in linguistic encoding.
  - → Identify such errors through linguistic annotation.
- The value of identifying such mismatches systematically is confirmed by recent SLA research (Zyzik & Azevedo 2009)
  - L2 learners have difficulty distinguishing between word classes among semantically related forms
  - Hypothesis: limited ability to interpret syntactic and morphological cues

On the nature of categories for learner language

- Where do the categories abstracted to come from?
- Categories result from generalizations, which require a significant amount of comparable data to be made.
  - requires decision on what constitutes comparable data, which is difficult for a dynamic target such as interlanguage
- Robustness and the level of analysis:
  - In NLP, *robustness* is the ability to *ignore* variation in the realization of a category to be identified.
  - But variation in the realization of a category is an important characteristic of learner language.
    - → Design annotation schemes for learner language to encode minimal observations.
    - → Provide access to those on one level of annotation, with other annotation levels providing robust L2 abstractions.

On the nature of categories for learner language

Comparative fallacy

- "mistake of studying the systematic character of one language by comparing it to another" (Bley-Vroman 1983)
  - extended to include bias towards towards native language (Lakshmanan & Selinker 2001)
- Essentially trying to analyze a "non-canonical variety" using a "robust" version of the canonical grammar.
  - divergences from norm is annotated as errors
  - but: the research question is the issue here, not corpus error annotation as such (Tenfjord et al. 2006)
- Issue more general than language acquisition research:
  - Eurocentrism in field work, e.g., Gil (2001)
  - Variationist sociolinguistics

→ Importance of explicitly defining classes and when an instance is counted as one of the variants.

On the nature of categories for learner language

Aspects of syntactic modeling

- Just like POS categories, syntactic structure is motivated by different types of evidence.
- For analyzing learner language, one can separate:
  - overall topology of a sentence (Hirschmann et al. 2007)
  - chunks and chunk-internal word order (Abney 1997)
  - lexical dependencies
    - canonical, as interface to meaning (MacWhinney 2008; Rosén & Smedt 2010; Ott & Ziai 2010; Hirschmann et al. 2010)
    - non-canonical, separating evidence for morpho-syntactic and semantic relations (Dickinson & Ragheb 2009)

Error annotation

- Error annotation involves (implicitly or explicitly):
  a) Determining what the learner wanted to say (**target**).
  b) Identifying
     i. the **location** of the error, and
     ii. the **type** of the error corresponding to the difference between the learner sentence and the target hypothesis.
  c) **Annotating** the error in the corpus
- Each of these steps can present ambiguity:
  a) multiple possible target hypotheses
  b) i. different locations in which the error can be rooted
     ii. different types of errors a divergence can be attributed to
  c) different ways to mark an error location & type in corpus

Error annotation schemes: Desiderata

Inter-annotator agreement

- An annotation is only relevant and useful if it provides a uniform, reliable index to relevant classes of data.
- Traditionally every researcher develops their own error annotation scheme. (cf. Díaz-Negrillo & Fernández-Domínguez 2006)
- Alarmingly, no studies on which inter-annotator agreement can be reached for which distinctions in error annotation
- No freely available gold standard corpora, so
  - no reliable quantitative evaluation in research
  - no reliable training & evaluation of NLP for error analysis
- Promising progress for some subclasses (det, prep) (e.g., Lee & Seneff 2006; Tetreault & Chodorow 2008; De Felice 2008)
  - but it is important to establish a tool-independent, transparent definition of the markables to be annotated

Target hypotheses

- Fitzpatrick & Seegmiller (2004) report unsatisfactory levels of agreement in determining learner target forms.
  - Keeping the target hypothesis implicit results in error annotation which diverge even more unsatisfactorily.
- Anke Lüdeling has argued for making target hypotheses an explicit part of error annotation (Lüdeling et al. 2005; Hirschmann et al. 2007; Lüdeling 2008).
  - supports alternative targets (and corresponding error annotation), and
  - supports multiple target hypotheses, differing in scope and operations allowed to obtain them
    - e.g., only replacement, omission, etc. to make sentence locally well-formed vs. taking context into account
- If target hypothesis is explicit, one can evaluate reliability of second step, from target hypothesis to error tag.

Difficulty of determining target hypotheses

- What are the target forms for the sentences taken from the Hiroshima English Learners' Corpus (Miura 1998)?

  (14) *I didn't know*
  (15) *I don't know his lives.*
  (16) *I know where he lives.*
  (17) *I know he lived*

  They are taken from a translation task, for the Japanese of

  (18) *I don't know where he lives.*

- How can one obtain a better handle on target hypotheses?
  - Focus on more advanced learners.
  - Take explicit task context into account.
  - Support targets other than fully explicit surface forms.
  - Take more learner strategies into account.
    - Learners often lift material from texts or use mastered chunks instead of trying to express appropriate meaning!

Constraining the search space of interpretation

Importance of activity and learner modeling

- All approaches to modeling learner language, such as
  - *mal*-rules, constraint relaxation, statistical modeling

  must model the space of **well-formed and ill-formed variation** that is possible given
  - a particular activity, and
  - a given learner.
- For example, without task and speaker context, how would you interpret the following?

  (19) *I will not buy this record it is scratched*
  (20) *My hovercraft is full of eels.*

Exemplifying interpretation in context

Monty Python: Hungarian Phrase Book sketch
http://www.youtube.com/watch?v=akbfIkF_1zY

Towards task-specific learner corpora

- Explicit task and learner models included as meta-information in a corpus can provide crucial constraining information for interpreting learner language.
  - E.g., it's easier to infer what a learner wanted to say if one knows the text they are answering questions about.
  - Related to taking strategic competence, task, and L1 into account in learner models of Intelligent Tutoring Systems (Amaral & Meurers 2008).
- Most current learner language corpora consist of essays, yet learners produce language in a wide range of contexts, naturalistic or instructed, e.g.,
  - email and chat messages
  - answering reading or listening comprehension questions
  - asking questions in information gap activities
- To obtain learner corpora which are interpretable and representative, we need language resulting from explicit tasks, in a variety of contexts, including longitudinal data.

Content assessment

Based on joint work with Stacey Bailey, Niels Ott, Ramon Ziai

- Meaningful, contextualized use of language plays a crucial role in second language acquisition – yet the (automatic) analysis has focused on form aspects.
- How can the meaning of sentences and text fragments be analyzed and compared in realistic situations?
  - Realistic situations:
    - differences in situative and world knowledge
    - language not necessarily well-formed
- Two challenges:
  - Which linguistic representations can be robustly identified as basis of a computational approximation of meaning?
  - How can the role of the context be integrated?

⇒ Start by collecting data of authentic language in context.

Collecting authentic data for content assessment

- We want to make the context explicit by collecting data in the setting of a concrete task.
  - To support evaluation of meaning, focus on tasks using information encoded in language, not world knowledge.
  - Current learner corpora typically consist of essay data, so only the essay topic is known
    → contents quite unconstrained and not predictable.
  - Other activities provide more explicit, language-based context with more predictable contents: summarization, reading comprehension, information-gap activities, . . .

⇒ Compile a corpus with answers to reading comprehension questions written by learners of English (Bailey 2008; Bailey & Meurers 2008; Meurers, Ziai, Ott & Bailey 2011).
  - In the CoMiC project, we focus on learners of German (Meurers, Ott & Ziai 2010a).

Loosely restricted reading comprehension

An example

Question: *What are the methods of propaganda mentioned in the article?*

Target: *The methods include use of labels, visual images, and beautiful or famous people promoting the idea or product. Also used is linking the product to concepts that are admired or desired and to create the impression that everyone supports the product or idea.*

Sample Learner Responses:

- *A number of methods of propaganda are used in the media.*
- *Bositive or negative labels.*
- *Giving positive or negative labels. Using visual images. Having a beautiful or famous person to promote. Creating the impression that everyone supports the product or idea.*

CAM-En learner corpus

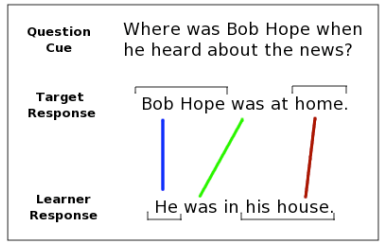
- The corpus was collected in second language classrooms, using the ordinary exercises assigned by the teacher.
  - Teachers also provided target answers and learner answer assessment.
- CAM-En corpus: 566 responses to RC questions from intermediate English as a Second Language students.
  - Development set:
    - 311 responses from 11 students to 47 questions
  - Test set:
    - 255 responses from 15 students to 28 questions

Annotation: Categories for content assessment

- The annotation scheme was developed by analyzing target and learner responses in the development corpus.
- Two graders independently annotated the data:
  - detection (binary): correct vs. incorrect meaning
  - diagnosis (5 codes): correct; missing concept, extra concept, blend, non-answer

  Eliminated responses which graders did not agree on
  - 48 in development set (15%) and 31 in test set (12%)
- Learner responses vary significantly; no full bag-of-word overlap between test set answers and targets.
- On average, 2.7 form errors per sentence.

Basic idea: Comparing responses & targets



- CAM compares target & learner responses in three steps:
  1. **Annotation** uses NLP to enrich the learner and target responses and question text with linguistic information.
  2. **Alignment** maps units in the learner response to units in the target response using the annotated information.
  3. **Diagnosis** analyzes the alignment to label the learner response with a target modification diagnosis code.

Types of alignment

Alignment can involve different types of representation:

Alignment Type	Example Match
Token-identical	**advertising** **advertising**
Lemma-resolved	**advertisement** **advertising**
Spelling-resolved	**campaing** **campaign**
Reference-resolved	**Clinton** **he**
Semantic similarity-resolved	**initial** **beginning**
Specialized expressions	**May 24, 2007** **5/24/2007**

Levels of alignment

Alignment can take place at different levels of representation:

Level	Example	Alignment
Tokens	**The explanation is simple.** **The reason is simple.**	**explanation** **reason**
Chunks	**A brown dog sat in a nice car.** **A nice dog sat in a car.**	**a brown dog** **a nice dog**
Dependency triples	**He knows the doctor.** **John knows him.**	**obj(knows, doctor)** **obj(knows, him)**

NLP tools used

Annotation Task	Language Processing Tool
Sentence Detection, Tokenization, Lemmatization	MontyLingua (Liu 2004)
Lemmatization	PC-KIMMO (Antworth 1993)
Spell Checking	Edit distance (Levenshtein 1966), SCOWL word list (Atkinson 2004)
Part-of-speech Tagging	TreeTagger (Schmid 1994)
Noun Phrase Chunking	CASS (Abney 1996)
Lexical Relations	WordNet (Miller 1995)
Similarity Scores	PMI-IR (Turney 2001; Mihalcea et al. 2006)
Dependency Relations	Stanford Parser (Klein & Manning 2003)

Features used for content assessment

- Diagnosis is based on 14 features:

  # of Overlapping Matches:
  - keyword (head word)
  - target/learner token
  - target/learner chunk
  - target/learner triple

  Nature of Matches:
  - % token matches
  - % lemma matches
  - % synonym matches
  - % similarity matches
  - % sem. type matches
  - match variety

  Semantic error detection

- For combining the evidence, machine learning (TiMBL, Daelemans et al. 2007) worked better than manual rules.

Results

Binary classification	Accuracy
Development Set (leave-one-out testing)	87%
Test Set	88%

Diagnosis with 5 codes	Accuracy
Development Set	87%
Test Set	87%

Form errors don't negatively impact results:

- 68% of **correctly** diagnosed items had form errors.
- 53% of **incorrectly** diagnosed ones did as well.

Related work

- No directly comparable systems, but competitive with results for automatic scoring of native speaker short answers by C-Rater (Leacock & Chodorow 2003; Leacock 2004).
  - Techniques used by essay grading systems (e.g., E-Rater, Burstein et al. 2003; AutoTutor, Graesser et al. 1999) do not generalize well to short (1-2 sentence) responses.
- Related research issues
  - Paraphrase recognition (e.g., Brockett & Dolan 2005; Hatzivassiloglou et al. 1999)
  - Machine translation evaluation (e.g., Banerjee & Lavie 2005; Lin & Och 2004)
  - Essay-based question answering systems (e.g., Deep Read, Hirschman et al. 1999)
  - Automatic grading (e.g., Leacock 2004; Marín 2004)
  - Recognition of Textual Entailment (RTE, Dagan et al. 2006)

Future work

Towards Interpretation in Context

- The reading comprehension question task we are focusing on provides an explicit context in form of
  - the text, and
  - the questions asked about it.
- CAM currently takes this context into account for basic anaphora resolution in the target and learner answers.
- But how about about other aspects of this context?
  - How should information in the answers that is *given* in the question be interpreted?
  - How can the nature of a question and the task strategies it requires be taken into account?

Towards interpretation in context

Treatment of *given* information

- Example from CAM-en:
  - **Cue**: *What was **the** major **moral question raised by the Clinton incident**?*
  - **Target**: ***The moral question raised by the Clinton incident was** whether a politician's person life is relevant to their job performance.*
  - **Response**: ***A basic question for the media is** whether a politician's personal life is relevant to his or her performance in the job.*
- The original CAM simply removed *given* words.
- We are developing a more sophisticated approach to
  - keep sentence intact for deeper processing
  - use the occurrence of *given* information to distinguish between incorrect answers and off-topic answers.

Towards interpretation in context

Question classification

- Comparing the meaning of answers to questions should make use of nature of the questions being answered.
- Features to be investigated include
  - **Learning Goals**: Targeted cognitive skills and knowledge (e.g., Anderson & Krathwohl 2001)
  - **Knowledge Sources**: The implicit/explicit answer source (Irwin 1986; Pearson & Johnson 1978)
  - **Text Type**: The rhetorical structure of the text (Champeau de Lopez et al. 1997)
  - **Answer Type**: The kind of answer expected (Gerbault 1999)
- Results here may also help answer:
  - What are suitable, more fine grained diagnosis categories for content assessment?

Adaptivity of analysis

- Given the high number of form errors in learner data, deep analysis and model construction often is not feasible.
- However, there are patterns for which a dedicated, deep analysis may be possible or even important.
- Patterns to be explored include
  - semantic units expected in the answer (cf. answer typing)
  - specific linguistic constructions identified in the answer which require special treatment (e.g., negation).
  - typical well-formed “islands of compositionality” supporting a deep analysis (e.g., particular NP patterns)
- Adaptively combining shallow & deeper analyses becomes especially important when going from English to languages with richer morphology & freer word order (e.g., German).

Conclusion

- We motivated linguistic annotation to support effective querying for SLA patterns and discussed an approach to the POS analysis of learner language separating
  - lexical, morphological, and distributional information

  Goal: Corpus annotation systematically characterizing language (native-like as well as learner innovations).
- Turning to error annotation, we argued for inter-annotator agreement as crucial for establishing which distinctions are replicable based on the available information.
- We explored the nature of target hypotheses and argued for explicit task and learner modeling to constrain the search space of interpretation.
- Turning to aspects of meaning, we discussed the analysis of answers to reading comprehension questions and research issues we are currently exploring in this context.

Context of this research: Our Background

Analyzing language for learners

- Input enhancement of texts for learners (Meurers et al. 2010b)
- Search engine for language learners (Ott & Meurers 2010)
- Prediction of functional elements (Elghafari, Meurers & Wunsch 2010)

Analyzing learner language

- Intelligent Tutoring System TAGARELA for Portuguese (Amaral & Meurers 2008, 2009, 2011; Amaral et al. 2011)
- Linguistic analysis of NOCE corpus of English written by Spanish learners (Díaz-Negrillo, Meurers, Valera & Wunsch 2010)
- Automatic analysis of learner language (Meurers 2009)
- Word order errors (Metcalf & Meurers 2006b; Boyd & Meurers 2008)
- Content assessment of answers to reading comprehension questions (Bailey & Meurers 2008) → SFB 833 A4 (CoMIC)
  - Longitudinal corpus collection using WELCOME (Meurers, Ott & Ziai 2010a) → KU/OSU collaboration
  - Dependency parsing of learner language (Ott & Ziai 2010)

References

Abney, S. (1996). Partial Parsing via Finite-State Cascades. In *The Robust Parsing Workshop of the European Summer School in Logic, Language and Information (ESSLLI '96)*. Prague, Czech Republic, pp. 1–8. URL http://www.vinartus.net/spa/97a.pdf.

Abney, S. (1997). Partial Parsing via Finite-State Cascades. *Natural Language Engineering* 2, 337–344. URL http://www.vinartus.net/spa/97a.pdf.

Amaral, L., V. Metcalf & D. Meurers (2006). Language Awareness through Re-use of NLP Technology. Pre-conference Workshop on NLP in CALL – Computational and Linguistic Challenges. CALICO 2006. May 17, 2006. University of Hawaii. URL http://purl.org/net/icall/handouts/calico06-amaral-metcalf-meurers.pdf.

Amaral, L. & D. Meurers (2008). From Recording Linguistic Competence to Supporting Inferences about Language Acquisition in Context: Extending the Conceptualization of Student Models for Intelligent Computer-Assisted Language Learning. *Computer-Assisted Language Learning* 21(4), 323–338. URL http://purl.org/dm/papers/amaral-meurers-call08.html.

Amaral, L. & D. Meurers (2009). Little Things With Big Effects: On the Identification and Interpretation of Tokens for Error Diagnosis in ICALL. *CALICO Journal* 27(1). URL http://purl.org/dm/papers/amaral-meurers-09.html.

Amaral, L. & D. Meurers (2011). On Using Intelligent Computer-Assisted Language Learning in Real-Life Foreign Language Teaching and Learning. *ReCALL* 23(1). URL http://purl.org/dm/papers/amaral-meurers-10.html.

Amaral, L., D. Meurers & R. Ziai (2011). Analyzing Learner Language: Towards A Flexible NLP Architecture for Intelligent Language Tutors. *Computer-Assisted Language Learning* 24(1), 1–16. URL http://purl.org/dm/papers/amaral-meurers-ziai-10.html.

Anderson, L. W. & D. Krathwohl (eds.) (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman Publishers.

Antworth, E. L. (1993). Glossing Text with the PC-KIMMO Morphological Parser. *Computers and the Humanities* 26, 475–484. URL http://www.springerlink.com/content/r20w66k70976ur9l/fulltext.pdf.

Atkinson, K. (2004). Spell Checking Oriented Word Lists (SCOWL). URL http://wordlist.sourceforge.net/. Web resource.

Bailey, S. (2008). Content Assessment in Intelligent Computer-Aided Language Learning: Meaning Error Diagnosis for English as a Second Language. Ph.D. thesis, The Ohio State University. URL http://purl.org/net/Bailey-08.pdf.

Bailey, S. & D. Meurers (2008). Diagnosing meaning errors in short answers to reading comprehension questions. In J. Tetreault, J. Burstein & R. D. Felice (eds.), *Proceedings of the 3rd Workshop on Innovative Use of NLP for Building Educational Applications (BEA-3) at ACL'08*. Columbus, Ohio, pp. 107–115. URL http://aclweb.org/anthology/W08-0913.

Banerjee, S. & A. Lavie (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43th Annual Meeting of the Association of Computational Linguistics (ACL-2005)*. URL http://www.cs.cmu.edu/~alavie/papers/BanerjeeLavie2005-final.pdf.

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction

Learner Corpora
- Data in SLA Research
- Corpus annotation
- Annotation Quality

Categories for Learner Language
- Example: Parts-of-speech
- Automatic POS-Tagging
- Three Sources of Evidence
- Mismatching Evidence
- Systematic categories
- Nature of interlang. categories
- Comparative fallacy
- Error annotation
- Target hypotheses
- Activity & learner modeling
- Task-specific learner corpora

Content assessment
- CAM-En learner corpus
- Basic approach
- Realizing the approach
- Results
- Related work
- Future work

Conclusion

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

44 / 44

Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning* 33(1), 1–17.

Boyd, A., M. Dickinson & D. Meurers (2008). On Detecting Errors in Dependency Treebanks. *Research on Language and Computation* 6(2), 113–137. URL http://purl.org/dm/papers/boyd-et-al-08.html.

Boyd, A. & D. Meurers (2008). On Diagnosing Word Order Errors. Poster presented at the CALICO Pre-Conference Workshop on Automatic Analysis of Learner Language. URL http://purl.org/net/calico-workshop-abstracts.html#6.

Brants, T. & W. Skut (1998). Automation of Treebank Annotation. In *Proceedings of New Methods in Language Processing*. Sydney, Australia. URL http://wing.comp.nus.edu.sg/acl/W/W98/W98-1207.pdf.

Brockett, C. & W. B. Dolan (2005). Support Vector Machines for Paraphrase Identification and Corpus Construction. In *Proceedings of the Third International Workshop on Paraphrasing (IWP2005)*. pp. 1–8. URL http://aclweb.org/anthology/I05-5001.

Burstein, J. & M. Chodorow (1999). Automated Essay Scoring for Nonnative English Speakers. In *Proceedings of a Workshop on Computer-Mediated Language Assessment and Evaluation of Natural Language Processing, Joint Symposium of the Association of Computational Linguistics (ACL-99) and the International Association of Language Learning Technologies*. pp. 68–75. URL http://www.ets.org/Media/Research/pdf/erater_acl99rev.pdf.

Burstein, J., M. Chodorow & C. Leacock (2003). Criterion: Online Essay Evaluation: An Application for Automated Evaluation of Student Essays. In *Proceedings of the Fifteenth Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-03)*. Acapulco, Mexico, pp. 3–10. URL http://ftp.ets.org/pub/res/erater_iaai03_burstein.pdf.

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction

Learner Corpora
- Data in SLA Research
- Corpus annotation
- Annotation Quality

Categories for Learner Language
- Example: Parts-of-speech
- Automatic POS-Tagging
- Three Sources of Evidence
- Mismatching Evidence
- Systematic categories
- Nature of interlang. categories
- Comparative fallacy
- Error annotation
- Target hypotheses
- Activity & learner modeling
- Task-specific learner corpora

Content assessment
- CAM-En learner corpus
- Basic approach
- Realizing the approach
- Results
- Related work
- Future work

Conclusion

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

44 / 44

Champeau de Lopez, C., G. Marchi & M. Arreaza-Coyle (1997). A Taxonomy: Evaluating Reading Comprehension in EFL. *English Teaching Forum* 35(2), 30–42. URL http://dosfan.lib.uic.edu/usia/E-USIA/forum/vols/vol35/no2/p30.htm.

Clahsen, H. & P. Muysken (1986). The availability of Universal Grammar to adult and child learners: A study of the acquisition of German word order. *Second Language Acquisition* 2, 93–19. URL http://slr.sagepub.com/cgi/reprint/2/2/93.pdf.

Daelemans, W., J. Zavrel, K. der Sloot & A. van den Bosch (2007). *TiMBL: Tilburg Memory-Based Learner Reference Guide, ILK Technical Report ILK 07-03*. Induction of Linguistic Knowledge Research Group Department of Communication and Information Sciences, Tilburg University, P.O. Box 90153, NL-5000 LE, Tilburg, The Netherlands, version 6.0 ed. URL http://ilk.uvt.nl/downloads/pub/papers/ilk.0703.pdf.

Dagan, I., O. Glickman & B. Magnini (2006). The PASCAL Recognising Textual Entailment Challenge. In J. Q. Candela, I. Dagan, B. Magnini & F. d'Alché Buc (eds.), *Machine Learning Challenges, Evaluating Predictive Uncertainty, Visual Object Classification and Recognizing Textual Entailment, First PASCAL Machine Learning Challenges Workshop, MLCW 2005, Southampton, UK, April 11-13, 2005, Revised Selected Papers*. Springer, vol. 3944 of *Lecture Notes in Computer Science*, pp. 177–190. URL http://u.cs.biu.ac.il/~dagan/publications/RTEChallenge.pdf.

De Felice, R. (2008). Automatic Error Detection in Non-native English. Ph.D. thesis, St Catherine's College, University of Oxford.

de Haan, P. (2000). Tagging non-native English with the TOSCA-ICLE tagger. In Mair & Hundt (2000), pp. 69–79.

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction

Learner Corpora
- Data in SLA Research
- Corpus annotation
- Annotation Quality

Categories for Learner Language
- Example: Parts-of-speech
- Automatic POS-Tagging
- Three Sources of Evidence
- Mismatching Evidence
- Systematic categories
- Nature of interlang. categories
- Comparative fallacy
- Error annotation
- Target hypotheses
- Activity & learner modeling
- Task-specific learner corpora

Content assessment
- CAM-En learner corpus
- Basic approach
- Realizing the approach
- Results
- Related work
- Future work

Conclusion

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

44 / 44

de Mönnink, I. (2000). Parsing a learner corpus. In Mair & Hundt (2000), pp. 81–90.

Díaz-Negrillo, A. (2009). *EARS: A User's Manual*. Munich, Germany: LINCOM Academic Reference Books.

Dickinson, M. & W. D. Meurers (2003a). Detecting Errors in Part-of-Speech Annotation. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL-03)*. Budapest, Hungary, pp. 107–114. URL http://purl.org/dm/papers/dickinson-meurers-03.html.

Dickinson, M. & W. D. Meurers (2003b). Detecting Inconsistencies in Treebanks. In *Proceedings of the Second Workshop on Treebanks and Linguistic Theories (TLT-03)*. Växjö, Sweden, pp. 45–56. URL http://purl.org/dm/papers/dickinson-meurers-tlt03.html.

Dickinson, M. & W. D. Meurers (2005). Detecting Errors in Discontinuous Structural Annotation. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*. pp. 322–329. URL http://aclweb.org/anthology/P05-1040.

Dickinson, M. & M. Ragheb (2009). Dependency Annotation for Learner Corpora. In *Proceedings of the Eighth Workshop on Treebanks and Linguistic Theories (TLT-8)*. Milan, Italy. URL http://jones.ling.indiana.edu/~mdickinson/papers/dickinson-ragheb09.html.

Díaz-Negrillo, A. & J. Fernández-Domínguez (2006). Error Tagging Systems for Learner Corpora. *Revista Española de Lingüística Aplicada (RESLA)* 19, 83–102. URL http://dialnet.unirioja.es/servlet/fichero_articulo?codigo=2198610&orden=72810.

Díaz-Negrillo, A., D. Meurers, S. Valera & H. Wunsch (2010). Towards interlanguage POS annotation for effective learner corpora in SLA and FLT. *Language Forum* URL http://purl.org/dm/papers/diaz-negrillo-et-al-09.html.

Elghafari, A., D. Meurers & H. Wunsch (2010). Exploring the Data-Driven Prediction of Prepositions in English. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)*. Beijing, China, pp. 267–275. URL http://aclweb.org/anthology/C10-2031.

Fitzpatrick, E. & M. S. Seegmiller (2004). The Montclair electronic language database project. In U. Connor & T. Upton (eds.), *Applied Corpus Linguistics: A Multidimensional Perspective*, Amsterdam: Rodopi. URL http://chss.montclair.edu/linguistics/MELD/rodopipaper.pdf.

Gerbault, J. (1999). Towards an analysis of answers to open-ended questions in computer-assisted language learning. In S. Lajoie & M. Vivet (eds.), *Proceedings of AIED*. IOS Press, pp. 686–689.

Gil, D. (2001). Escaping Eurocentrism: Fieldwork as a Process of Unlearning. In P. Newman & M. Ratliff (eds.), *Linguistic Fieldwork*, Cambridge University Press, pp. 102–132.

Graesser, A. C., K. Wiemer-Hastings, P. Wiemer-Hastings & R. Kreuz (1999). AutoTutor: A simulation of a human tutor. *Journal of Cognitive Systems Research* 1, 35–51.

Granger, S. (2003). Error-tagged learner corpora and CALL: A promising synergy. *CALICO Journal* 20(3), 465–480. URL http://purl.org/calico/granger03.pdf.

Hatzivassiloglou, V., J. Klavans & E. Eskin (1999). Detecting Text Similarity over Short Passages: Exploring Linguistic Feature Combinations via Machine Learning. In *Proceedings of Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP'99)*. College Park, Maryland, pp. 203–212. URL http://aclweb.org/anthology/W99-0625.

Hawkins, J. A. & P. Buttery (2009). Using Learner Language from Corpora to Profile Levels of Proficiency – Insights from the English Profile Programme. In

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction

Learner Corpora
Data in SLA Research
Corpus annotation
Annotation Quality

Categories for Learner Language
Example: Parts-of-speech
Automatic POS-Tagging
Three Sources of Evidence
Mismatching Evidence
Systematic categories
Nature of interlang. categories
Comparative fallacy
Error annotation
Target hypotheses
Activity & learner modeling
Task-specific learner corpora

Content assessment
CAM-En learner corpus
Basic approach
Realizing the approach
Results
Related work
Future work

Conclusion

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

44 / 44

Studies in Language Testing: The Social and Educational Impact of Language Assessment, Cambridge: Cambridge University Press.

Hawkins, J. A. & P. Buttery (2010). Criterial Features in Learner Corpora: Theory and Illustrations. *English Profile Journal* .

Hirschman, L., M. Light, E. Breck & J. Burger (1999). Deep Read: A Reading Comprehension System. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL-99)*. College Park, Maryland, pp. 325–332. URL http://www.eecs.berkeley.edu/~nimar/readings/hirschman1999.pdf.

Hirschmann, H., S. Doolittle & A. Lüdeling (2007). Syntactic annotation of non-canonical linguistic structures. In *Proceedings of Corpus Linguistics 2007*. Birmingham. URL http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/neu2/mitarbeiter-innen/anke/pdf/HirschmannDoolittleLuedelingCL2007.pdf.

Hirschmann, H., A. Lüdeling, I. Rehbein, M. Reznicek & A. Zeldes (2010). Syntactic Overuse and Underuse: A Study of a Parsed Learner Corpus and its Target Hypothesis. Presentation given at the Treebanks and Linguistic Theory Workshop.

Housen, A. & F. Kuiken (2009). Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics* 30(4), 461–473. URL http://applij.oxfordjournals.org/content/30/4/461.full.pdf.

Irwin, J. W. (1986). *Teaching Reading Comprehension Processes*. Engelwood Cliffs, New Jersey: Prentice-Hall, Inc.

Klein, D. & C. D. Manning (2003). Accurate Unlexicalized Parsing. In *Proceedings of the 41st Meeting of the Association for Computational Linguistics (ACL 2003)*. Sapporo, Japan, pp. 423–430. URL http://aclweb.org/anthology/P03-1054.

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction

Learner Corpora
Data in SLA Research
Corpus annotation
Annotation Quality

Categories for Learner Language
Example: Parts-of-speech
Automatic POS-Tagging
Three Sources of Evidence
Mismatching Evidence
Systematic categories
Nature of interlang. categories
Comparative fallacy
Error annotation
Target hypotheses
Activity & learner modeling
Task-specific learner corpora

Content assessment
CAM-En learner corpus
Basic approach
Realizing the approach
Results
Related work
Future work

Conclusion

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

44 / 44

Lakshmanan, U. & L. Selinker (2001). Analysing interlanguage: how do we know what learners know? *Second Language Research* 17(4), 393–420. URL http://proxy.lib.ohio-state.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=ufh&AN=7393417&site=ehost-live.

Leacock, C. (2004). Scoring Free-Responses Automatically: A Case Study of a Large-Scale Assessment. *Examens* 1(3). URL http://www.nocheating.org/Media/Research/pdf/erater_examens_leacock.pdf.

Leacock, C. & M. Chodorow (2003). C-rater: Automated Scoring of Short-Answer Questions. *Computers and the Humanities* 37, 389–405. URL http://www.ingentaconnect.com/content/klu/chum/2003/00000037/00000004/05144721?crawler=true.

Lee, J. & S. Seneff (2006). Automatic Grammar Correction for Second-Language Learners. In *INTERSPEECH 2006 – ICSLP*. URL http://groups.csail.mit.edu/sls/publications/2006/IS061299.pdf.

Levenshtein, V. I. (1966). Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. *Soviet Physics Doklady* 10(8), 707–710.

Lin, C.-Y. & F. J. Och (2004). Automatic Evaluation of Machine Translation Quality Using Longest Common Subsequence and Skip-Bigram Statistics. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*. pp. 605–612. URL http://www.mt-archive.info/ACL-2004-Lin.pdf.

Liu, H. (2004). MontyLingua: An End-to-End Natural Language Processor with Common Sense. Software Website. Media Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts. URL http://web.media.mit.edu/~hugo/montylingua/.

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction

Learner Corpora
Data in SLA Research
Corpus annotation
Annotation Quality

Categories for Learner Language
Example: Parts-of-speech
Automatic POS-Tagging
Three Sources of Evidence
Mismatching Evidence
Systematic categories
Nature of interlang. categories
Comparative fallacy
Error annotation
Target hypotheses
Activity & learner modeling
Task-specific learner corpora

Content assessment
CAM-En learner corpus
Basic approach
Realizing the approach
Results
Related work
Future work

Conclusion

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

44 / 44

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics* 15(4), 474–496.

Lüdeling, A. (2008). Mehrdeutigkeiten und Kategorisierung: Probleme bei der Annotation von Lernerkorpora. In P. Grommes & M. Walter (eds.), *Fortgeschrittene Lernervarietäten*, Tübingen: Niemeyer, pp. 119–140.

Lüdeling, A., M. Walter, E. Kroymann & P. Adolphs (2005). Multi-level error annotation in learner corpora. In *Proceedings of Corpus Linguistics*. Birmingham. URL http://www.corpus.bham.ac.uk/PCLC/Falko-CL2006.doc.

MacWhinney, B. (2008). Enriching CHILDES for morphosyntactic analysis. In H. Behrens (ed.), *Corpora in Language Acquisition Research: History, Methods, Perspectives*, Amsterdam and Philadelphia: John Benjamins, vol. 6 of *Trends in Language Acquisition Research*, pp. 165–197. URL http://childes.psy.cmu.edu/grasp/morphosyntax.doc.

Mair, C. & M. Hundt (eds.) (2000). *Corpus Linguistics and Linguistic Theory*. Amsterdam: Rodopi.

Marín, D. R. P. (2004). Automatic Evaluation of Users' Short Essays by Using Statistical and Shallow Natural Language Processing Techniques. Master's thesis, Universidad Autónoma de Madrid. URL http://www.ii.uam.es/~dperez/tea.pdf.

Metcalf, V. & D. Meurers (2006a). Generating Web-based English Preposition Exercises from Real-World Texts. URL http://purl.org/net/icall/handouts/eurocall06-metcalf-meurers.pdf. EUROCALL 2006. Granada, Spain. September 4–7, 2006.

Metcalf, V. & D. Meurers (2006b). When to Use Deep Processing and When Not To – The Example of Word Order Errors. URL

http://purl.org/net/icall/handouts/calico06-metcalf-meurers.pdf. Pre-conference Workshop on NLP in CALL – Computational and Linguistic Challenges. CALICO 2006. May 17, 2006. University of Hawaii.

Meunier, F. (1998). Computer Tools for Interlanguage Analysis: A Critical Approach. In G. Sylviane (ed.), *Learner English on Computer*, London and New York: Addison Wesley Longman, pp. 19–37.

Meurers, D. (2009). On the Automatic Analysis of Learner Language: Introduction to the Special Issue. *CALICO Journal* 26(3), 469–473. URL http://purl.org/dm/papers/meurers-09.html.

Meurers, D., N. Ott & R. Ziai (2010a). Compiling a Task-Based Corpus for the Analysis of Learner Language in Context. In *Proceedings of Linguistic Evidence*. Tübingen, pp. 214–217. URL http://purl.org/dm/papers/meurers-ott-ziai-10.html.

Meurers, D., R. Ziai, L. Amaral, A. Boyd, A. Dimitrov, V. Metcalf & N. Ott (2010b). Enhancing Authentic Web Pages for Language Learners. In *Proceedings of the 5th Workshop on Innovative Use of NLP for Building Educational Applications (BEA-5) at NAACL-HLT 2010*. Los Angeles: Association for Computational Linguistics. URL http://purl.org/dm/papers/meurers-ziai-et-al-10.html.

Meurers, D., R. Ziai, N. Ott & S. Bailey (2011). Integrating Parallel Analysis Modules to Evaluate the Meaning of Answers to Reading Comprehension Questions. *IJCEELL Special Issue on Automatic Free-text Evaluation* URL http://purl.org/dm/papers/meurers-ziai-ott-bailey-11.html.

Meurers, W. D. (2005). On the use of electronic corpora for theoretical linguistics. Case studies from the syntax of German. *Lingua* 115(11), 1619–1639. URL http://purl.org/dm/papers/meurers-03.html.

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction

Learner Corpora
Data in SLA Research
Corpus annotation
Annotation Quality

Categories for Learner Language
Example: Parts-of-speech
Automatic POS-Tagging
Three Sources of Evidence
Mismatching Evidence
Systematic categories
Nature of interlang. categories
Comparative fallacy
Error annotation
Target hypotheses
Activity & learner modeling
Task-specific learner corpora

Content assessment
CAM-En learner corpus
Basic approach
Realizing the approach
Results
Related work
Future work

Conclusion

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

44 / 44

Meurers, W. D. & S. Müller (2009). Corpora and Syntax (Article 42). In A. Lüdeling & M. Kytö (eds.), *Corpus linguistics*, Berlin: Mouton de Gruyter, vol. 2 of *Handbooks of Linguistics and Communication Science*, pp. 920–933. URL http://purl.org/dm/papers/meurers-mueller-09.html.

Mihalcea, R., C. Corley & C. Strapparava (2006). Corpus-based and Knowledge-based Measures of Text Semantic Similarity. In *Proceedings of the National Conference on Artificial Intelligence*. Menlo Park, CA: American Association for Artificial Intelligence (AAAI) Press, vol. 21(1), pp. 775–780. URL http://www.cse.unt.edu/~rada/papers/mihalcea.aaai06.pdf.

Miller, G. (1995). WordNet: a lexical database for English. *Communications of the ACM* 38(11), 39–41. URL http://aclweb.org/anthology/H94-1111.

Miura, S. (1998). Hiroshima English Learners' Corpus: English learner No. 2 (English I & English II). URL http://home.hiroshima-u.ac.jp/d052121/eigo2.html. Last Modified 14 May, 1998.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics* 24(4), 492—-518.

Ott, N. (2009). Information Retrieval for Language Learning: An Exploration of Text Difficulty Measures. ISCL master's thesis, Universität Tübingen, Seminar für Sprachwissenschaft, Tübingen, Germany. URL http://drni.de/zap/ma-thesis.

Ott, N. & D. Meurers (2010). Information Retrieval for Education: Making Search Engines Language Aware. *Themes in Science and Technology Education. Special issue on computer-aided language analysis, teaching and learning: Approaches, perspectives and applications* URL http://purl.org/dm/papers/ott-meurers-10.html.

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction

Learner Corpora
Data in SLA Research
Corpus annotation
Annotation Quality

Categories for Learner Language
Example: Parts-of-speech
Automatic POS-Tagging
Three Sources of Evidence
Mismatching Evidence
Systematic categories
Nature of interlang. categories
Comparative fallacy
Error annotation
Target hypotheses
Activity & learner modeling
Task-specific learner corpora

Content assessment
CAM-En learner corpus
Basic approach
Realizing the approach
Results
Related work
Future work

Conclusion

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

44 / 44

Ott, N. & R. Ziai (2010). Evaluating Dependency Parsing Performance on German Learner Language. In M. Dickinson, K. Müürisep & M. Passarotti (eds.), *Proceedings of the Ninth International Workshop on Treebanks and Linguistic Theories*. vol. 9 of *NEALT Proceeding Series*, pp. 175–186. URL http://www.sfs.uni-tuebingen.de/~rziai/papers/Ott.Ziai-10.pdf.

Pearson, P. D. & D. Johnson (1978). *Teaching Reading Comprehension*. New York: Holt, Rinehart and Winston.

Rosén, V. & K. D. Smedt (2010). Syntactic Annotation of Learner Corpora. In H. Johansen, A. Golden, J. E. Hagen & A.-K. Helland (eds.), *Systematisk, variert, men ikke tilfeldig. Antologi om norsk som andrespråk i anledning Kari Tenfjords 60-årsdag [Systematic, varied, but not arbitrary. Anthology about Norwegian as a second language on the occasion of Kari Tenfjord's 60th birthday]*, Oslo: Novus forlag, pp. 120–132.

Sampson, G. & A. Babarczy (2003). Limits to annotation precision. In *Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora (LINC-03)*. pp. 61–68. URL http://www.sfs.uni-tuebingen.de/~zinsmeis/AnnotCorp05/materials/sampson-barbarczy03.pdf.

Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of the International Conference on New Methods in Language Processing*. Manchester, UK, pp. 44–49. URL http://www.ims.uni-stuttgart.de/ftp/pub/corpora/tree-tagger1.pdf.

Tenfjord, K., J. E. Hagen & H. Johansen (2006). The Hows and Whys of coding categories in a learner corpus (or "How and Why an error-tagged learner corpus is not ipso facto one big comparative fallacy"). *Rivista di psicolinguistica applicata* 6, 93–108.

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction

Learner Corpora
Data in SLA Research
Corpus annotation
Annotation Quality

Categories for Learner Language
Example: Parts-of-speech
Automatic POS-Tagging
Three Sources of Evidence
Mismatching Evidence
Systematic categories
Nature of interlang. categories
Comparative fallacy
Error annotation
Target hypotheses
Activity & learner modeling
Task-specific learner corpora

Content assessment
CAM-En learner corpus
Basic approach
Realizing the approach
Results
Related work
Future work

Conclusion

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

44 / 44

Tetreault, J. & M. Chodorow (2008). The Ups and Downs of Preposition Error Detection in ESL Writing. In *Proceedings of COLING-08*. Manchester, UK. URL http://www.ets.org/Media/Research/pdf/r3.pdf.

Turney, P. (2001). Mining the Web for Synonyms: PMI-IR Versus LSA on TOEFL. In *Proceedings of the Twelfth European Conference on Machine Learning (ECML-2001)*. Freiburg, Germany, pp. 491–502.

van Rooy, B. & L. Schäfer (2002). The Effect of Learner Errors on POS Tag Errors during Automatic POS Tagging. *Southern African Linguistics and Applied Language Studies* 20, 325–335.

van Rooy, B. & L. Schäfer (2003). An Evaluation of Three POS Taggers for the Tagging of the Tswana Learner English Corpus. In D. Archer, P. Rayson, A. Wilson & T. McEnery (eds.), *Proceedings of the Corpus Linguistics 2003 conference Lancaster University (UK), 28 – 31 March 2003*. vol. 16 of *University Centre For Computer Corpus Research On Language Technical Papers*, pp. 835–844.

Voutilainen, A. & T. Järvinen (1995). Specifying a shallow grammatical representation for parsing purposes. In *Proceedings of the 7th Conference of the EACL*. Dublin, Ireland. URL http://portal.acm.org/ft_gateway.cfm?id=977003&type=pdf&coll=GUIDE&dl=GUIDE&CFID=47108142&CFTOKEN=71182750.

Wiemer-Hastings, P., K. Wiemer-Hastings & A. Graesser (1999). Improving an Intelligent Tutor's Comprehension of Students with Latent Semantic Analysis. In S. Lajoie & M. Vivet (eds.), *Artificial Intelligence in Education*, IOS Press, pp. 535–542. URL http://eprints.kfupm.edu.sa/45213/1/45213.pdf.

Wolfe-Quintero, K., S. Inagaki & H.-Y. Kim (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy & Complexity*. Honolulu: Second Language Teaching & Curriculum Center, University of Hawaii at Manoa.

Zyzik, E. & C. Azevedo (2009). Word Class Distinctions in Second Language Acquisition. *SSLA* 31(31), 1–29. URL http://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=3981776.

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction

Learner Corpora
Data in SLA Research
Corpus annotation
Annotation Quality

Categories for Learner Language
Example: Parts-of-speech
Automatic POS-Tagging
Three Sources of Evidence
Mismatching Evidence
Systematic categories
Nature of interlang. categories
Comparative fallacy
Error annotation
Target hypotheses
Activity & learner modeling
Task-specific learner corpora

Content assessment
CAM-En learner corpus
Basic approach
Realizing the approach
Results
Related work
Future work

Conclusion

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

44 / 44