

On Automatically Analyzing Learner Language

Detmar Meurers
Universität Tübingen

Learner Corpus Research 2011
"20 years of learner corpus research: looking back, moving ahead"
Louvain-la-Neuve, 15. September 2011

On Automatically Analyzing Learner Language
Detmar Meurers

Introduction

Why Analyze Learner Language?

Contact Points with CL

Learner Corpora

Data in SLA Research

Corpus annotation

Linguistic Annotation

Annotation

Case Study

Automatic POS-Tagging

Three Sources of Evidence

Mismatching Evidence

Categories for Learner Language

Systematic POS for Learner Language

On the nature of interlanguage categories

Comparative fallacy

Variation and robustness

Synthetic annotation

Target hypotheses

Task-specific learner corpora

Comparing Meaning in Context

Conclusion

UNIVERSITÄT TÜBINGEN

1 / 27

Overview

- Motivations behind analyzing learner language and points of contact with computational linguistics
- Linguistic modeling of learner language
 - Which linguistic categories for learner language?
 - sources of evidence: a case study on parts-of-speech
 - comparative fallacy
 - Which level of analysis?
 - between representing variation and robustness
 - role of target hypotheses
 - Importance of explicit task and learner modeling

On Automatically Analyzing Learner Language
Detmar Meurers

Introduction

Why Analyze Learner Language?

Contact Points with CL

Learner Corpora

Data in SLA Research

Corpus annotation

Linguistic Annotation

Annotation

Case Study

Automatic POS-Tagging

Three Sources of Evidence

Mismatching Evidence

Categories for Learner Language

Systematic POS for Learner Language

On the nature of interlanguage categories

Comparative fallacy

Variation and robustness

Synthetic annotation

Target hypotheses

Task-specific learner corpora

Comparing Meaning in Context

Conclusion

UNIVERSITÄT TÜBINGEN

2 / 27

Why Analyze Learner Language?

Second Language Acquisition (SLA)

- SLA research is aimed at understanding *how second languages are acquired* (and how language works)
 - empirical basis: analysis of learner data, ...
- SLA research also studies *instructional intervention*
 - targeting different aspects of *language*,
 - used in different types of *tasks*,
 - supporting different kinds of *feedback*, and
 - the *sequencing* of what is to be acquired.

On Automatically Analyzing Learner Language
Detmar Meurers

Introduction

Why Analyze Learner Language?

Contact Points with CL

Learner Corpora

Data in SLA Research

Corpus annotation

Linguistic Annotation

Annotation

Case Study

Automatic POS-Tagging

Three Sources of Evidence

Mismatching Evidence

Categories for Learner Language

Systematic POS for Learner Language

On the nature of interlanguage categories

Comparative fallacy

Variation and robustness

Synthetic annotation

Target hypotheses

Task-specific learner corpora

Comparing Meaning in Context

Conclusion

UNIVERSITÄT TÜBINGEN

3 / 27

Why Analyze Learner Language?

Foreign Language Teaching (FLT)

- adapt, advance, and test effectiveness of intervention methods in teaching practice
- analysis of learner language data helps *document and advance understanding of student abilities and needs*

On Automatically Analyzing Learner Language
Detmar Meurers

Introduction

Why Analyze Learner Language?

Contact Points with CL

Learner Corpora

Data in SLA Research

Corpus annotation

Linguistic Annotation

Annotation

Case Study

Automatic POS-Tagging

Three Sources of Evidence

Mismatching Evidence

Categories for Learner Language

Systematic POS for Learner Language

On the nature of interlanguage categories

Comparative fallacy

Variation and robustness

Synthetic annotation

Target hypotheses

Task-specific learner corpora

Comparing Meaning in Context

Conclusion

UNIVERSITÄT TÜBINGEN

4 / 27

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
- ▶ *Intelligent Tutoring Systems*: on-line analysis aimed at supporting language acquisition
 - ▶ 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)
 - ▶ determine progression through pedagogical material
- ▶ *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

(cf. Meurers 2012)

Learner Data in SLA Research

An example: Clahsen & Muysken (1986)

- ▶ They studied the acquisition of German word order by native speakers of Romance languages.
- ▶ Stages of acquisition:
 1. S (Aux) V O
 2. (AdvP/PP) S (Aux) V O
 3. S V[+fin] O V[-fin]
 4. XP V[+fin] S O
 5. S V[+fin] (Adv) O
 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.
- ▶ The annotation of learner corpora, however, has mostly focused on errors made by the learners.
- ▶ Even when learner errors are the research focus, their correlation with other linguistic properties is relevant.
- ▶ Linguistic annotation also is important for capturing
 - ▶ measures and characteristics of language development
 - ▶ Complexity, Accuracy & Fluency (Housen & Kuiken 2009) → Lu (2010)
 - ▶ Critical Features (Hawkins & Buttery 2009, 2010) → Alexopoulou et al. (2011)
 - ▶ overuse/underuse of linguistic material (Wiersma et al. 2011, Hirschmann, Lüdeling, Rehbein, Reznicek & Zeldes 2010)

⇒ What is involved in linguistically annotating learner corpora (automatically)?

Annotation of Linguistic Properties

- ▶ Annotation schemes for native language corpora have been developed for a wide range of linguistic properties:
 - ▶ part-of-speech, morphology
 - ▶ syntactic constituency, lexical dependency structures
 - ▶ semantics (word senses, coreference), discourse structure
- ▶ An annotation scheme is only as good as the distinctions it reliably supports making based on evidence in corpus.
 - ▶ E.g., particle vs. preposition dropped in PTB tagset
 - ▶ More classes can actually be more reliable if they are more coherent in terms of their observable properties.
 - ▶ cf. BNC Tag Enhancement Project (CLAWS7 → CLAWS5)
- ▶ Each type of annotation typically requires an extensive manual annotation effort → gold standard corpora

Quality of Annotation

- ▶ How can high quality annotation be obtained?
 - ▶ Keep only reliably and consistently identifiable distinctions
 - ▶ described in detailed manual (Voutilainen & Järvinen 1995;
 - ▶ including appendix on hard cases Sampson & Babarczy 2003)
 - ▶ Annotate corpus several times and independently, then test interannotator agreement (Artstein & Poesio 2009)
 - ▶ Detect annotation errors through automatic analysis of comparable data recurring in the corpus (<http://decca.osu.edu>)
- ▶ Automatic annotation tools which can be trained on such gold standard annotation are available.
 - ▶ But quality of automatic annotation drops significantly for text differing from the gold standard training material.
- ▶ Are the annotation scheme distinctions for native language appropriate and sufficient for learner language research?

On Automatically Analyzing Learner Language

Detmer Meurers

Introduction

Why Analyze Learner Language?
Contact Points with CL

Learner Corpora

Data in SLA Research
Corpus annotation
Linguistic Annotation

Annotation Case Study

Automatic POS-Tagging
Three Sources of Evidence
Mismatching Evidence

Categories for Learner Language

Systematic POS for Learner Language
On the nature of interlanguage categories
Comparative fallacy
Variation and robustness
Synthetic annotation
Target hypotheses
Task-specific learner corpora
Comparing Meaning in Context

Conclusion

UNIVERSITÄT TUBINGEN

9 / 27

Case study on part-of-speech annotating NOCE (Díaz Negrillo, Meurers, Valera & Wunsch 2010)

- ▶ 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
- ⇒ How about adding linguistic annotation?
 - ▶ Explored automatic part-of-speech (POS) annotation
 - ▶ What does it mean to POS-annotate learner language?

On Automatically Analyzing Learner Language

Detmer Meurers

Introduction

Why Analyze Learner Language?
Contact Points with CL

Learner Corpora

Data in SLA Research
Corpus annotation
Linguistic Annotation

Annotation Case Study

Automatic POS-Tagging
Three Sources of Evidence
Mismatching Evidence

Categories for Learner Language

Systematic POS for Learner Language
On the nature of interlanguage categories
Comparative fallacy
Variation and robustness
Synthetic annotation
Target hypotheses
Task-specific learner corpora
Comparing Meaning in Context

Conclusion

UNIVERSITÄT TUBINGEN

10 / 27

Automatic POS-Tagging of NOCE

Setup

- ▶ Used 3 POS taggers trained on WSJ newspaper text, using Penn Treebank tagset
 - ▶ TreeTagger, TnT tagger, Stanford tagger
- ▶ Tagged the error-annotated section of NOCE

Results

- ▶ Manually evaluated POS tags assigned by taggers to 10 texts by 10 different participants (1.850 words)
- ▶ Accuracy of automatically assigned tags
 - ▶ TreeTagger: 94.95%
 - ▶ TnT Tagger: 94.03%
 - ▶ Stanford Tagger: 88.11%
- ▶ What do these results mean?

On Automatically Analyzing Learner Language

Detmer Meurers

Introduction

Why Analyze Learner Language?
Contact Points with CL

Learner Corpora

Data in SLA Research
Corpus annotation
Linguistic Annotation

Annotation Case Study

Automatic POS-Tagging
Three Sources of Evidence
Mismatching Evidence

Categories for Learner Language

Systematic POS for Learner Language
On the nature of interlanguage categories
Comparative fallacy
Variation and robustness
Synthetic annotation
Target hypotheses
Task-specific learner corpora
Comparing Meaning in Context

Conclusion

UNIVERSITÄT TUBINGEN

11 / 27

Aspects of a qualitative analysis

- ▶ We found lower performance for expressions which do not exist in English (cf. de Haan 2000; van Rooy & Schäfer 2002).
 - ▶ Spelling
 - (1) *I think that university **teachs** to people [...]*
 - (2) *They can't pay their studies and **more over** they have to pay a flat [...]*
 - ▶ 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?
 - ▶ Which research questions can "native POS" tags answer?

On Automatically Analyzing Learner Language

Detmer Meurers

Introduction

Why Analyze Learner Language?
Contact Points with CL

Learner Corpora

Data in SLA Research
Corpus annotation
Linguistic Annotation

Annotation Case Study

Automatic POS-Tagging
Three Sources of Evidence
Mismatching Evidence

Categories for Learner Language

Systematic POS for Learner Language
On the nature of interlanguage categories
Comparative fallacy
Variation and robustness
Synthetic annotation
Target hypotheses
Task-specific learner corpora
Comparing Meaning in Context

Conclusion

UNIVERSITÄT TUBINGEN

12 / 27

Three Sources of Evidence for POS analysis

Lemma/Lexical entry:

- (3) *I was surprised by the word **of** the day.*

of ⇒ preposition

Morphology:

- (4) *There is a lot of **construction** going on here.*

-ion ⇒ noun

Distribution:

- (5) *The old **man** left.*

adj __ verb ⇒ noun

On Automatically Analyzing Learner Language
Dietmar Meurers

Introduction
Why Analyze Learner Language?
Contact Points with CL

Learner Corpora
Data in SLA Research
Corpus annotation
Linguistic Annotation

Annotation Case Study
Automatic POS-Tagging

Three Sources of Evidence
Mismatching Evidence

Categories for Learner Language

Systematic POS for Learner Language
On the nature of interlanguage categories

Comparative fallacy
Variation and robustness
Synoptic annotation
Target hypotheses
Task-specific learner corpora
Comparing Meaning in Context

Conclusion

UNIVERSITÄT TUBINGEN

13 / 27

Case 1: Stem-Distribution mismatch



- (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	?

On Automatically Analyzing Learner Language
Dietmar Meurers

Introduction
Why Analyze Learner Language?
Contact Points with CL

Learner Corpora
Data in SLA Research
Corpus annotation
Linguistic Annotation

Annotation Case Study
Automatic POS-Tagging
Three Sources of Evidence

Mismatching Evidence

Categories for Learner Language

Systematic POS for Learner Language
On the nature of interlanguage categories

Comparative fallacy
Variation and robustness
Synoptic annotation
Target hypotheses
Task-specific learner corpora
Comparing Meaning in Context

Conclusion

UNIVERSITÄT TUBINGEN

14 / 27

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 3 rd sg

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

Stem	Distribution	Morphology
noun / adjective	verb	verb

On Automatically Analyzing Learner Language
Dietmar Meurers

Introduction
Why Analyze Learner Language?
Contact Points with CL

Learner Corpora
Data in SLA Research
Corpus annotation
Linguistic Annotation

Annotation Case Study
Automatic POS-Tagging
Three Sources of Evidence

Mismatching Evidence

Categories for Learner Language

Systematic POS for Learner Language
On the nature of interlanguage categories

Comparative fallacy
Variation and robustness
Synoptic annotation
Target hypotheses
Task-specific learner corpora
Comparing Meaning in Context

Conclusion

UNIVERSITÄT TUBINGEN

15 / 27

Case 3: Stem-Morphology mismatch



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

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

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

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

On Automatically Analyzing Learner Language
Dietmar Meurers

Introduction
Why Analyze Learner Language?
Contact Points with CL

Learner Corpora
Data in SLA Research
Corpus annotation
Linguistic Annotation

Annotation Case Study
Automatic POS-Tagging
Three Sources of Evidence

Mismatching Evidence

Categories for Learner Language

Systematic POS for Learner Language
On the nature of interlanguage categories

Comparative fallacy
Variation and robustness
Synoptic annotation
Target hypotheses
Task-specific learner corpora
Comparing Meaning in Context

Conclusion

UNIVERSITÄT TUBINGEN

16 / 27

Case 4: Distribution-Morphology mismatch



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

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

On Automatically Analyzing Learner Language

Demter Meurers

Introduction
Why Analyze Learner Language?
Contact Points with CL

Learner Corpora
Data in SLA Research
Corpus annotation
Linguistic Annotation

Annotation Case Study
Automatic POS-Tagging
Three Sources of Evidence
Mismatching Evidence

Categories for Learner Language
Systematic POS for Learner Language
On the nature of interlanguage categories
Comparative fallacy
Variation and robustness
Syntactic annotation
Target hypotheses
Task-specific learner corpora
Comparing Meaning in Context

Conclusion
UNIVERSITÄT TUBINGEN

17 / 27

Systematic POS for Learner Language

- ▶ A single POS tag from a standard native tagset *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 are shown to have
 - difficulty distinguishing between word classes among semantically related lexical forms
 - limited ability to interpret syntactic and morphological cues
- ▶ Which type of categories are useful for learner language?

On Automatically Analyzing Learner Language

Demter Meurers

Introduction
Why Analyze Learner Language?
Contact Points with CL

Learner Corpora
Data in SLA Research
Corpus annotation
Linguistic Annotation

Annotation Case Study
Automatic POS-Tagging
Three Sources of Evidence
Mismatching Evidence

Categories for Learner Language
Systematic POS for Learner Language
On the nature of interlanguage categories
Comparative fallacy
Variation and robustness
Syntactic annotation
Target hypotheses
Task-specific learner corpora
Comparing Meaning in Context

Conclusion
UNIVERSITÄT TUBINGEN

18 / 27

On the nature of categories for learner language

- ▶ *Comparative fallacy* is “the mistake of studying the systematic character of one language by comparing it to another.” (Bley-Vroman 1983, p. 6)
 - extended to include bias 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 annotated as errors
 - But note: the research question is the issue here, not corpus error annotation as such (Tenford et al. 2006).
- ▶ Issue more general than language acquisition research:
 - Eurocentrism in field work (Gil 2001)
 - Variationist sociolinguistics:
 - Importance of defining variation to be studied and when an instance is counted as one of the variants.

On Automatically Analyzing Learner Language

Demter Meurers

Introduction
Why Analyze Learner Language?
Contact Points with CL

Learner Corpora
Data in SLA Research
Corpus annotation
Linguistic Annotation

Annotation Case Study
Automatic POS-Tagging
Three Sources of Evidence
Mismatching Evidence

Categories for Learner Language
Systematic POS for Learner Language
On the nature of interlanguage categories
Comparative fallacy
Variation and robustness
Syntactic annotation
Target hypotheses
Task-specific learner corpora
Comparing Meaning in Context

Conclusion
UNIVERSITÄT TUBINGEN

19 / 27

On the nature of categories for learner language

Between representing variation and robustness

- ▶ Where do linguistic categories come from?
 - Categories result from generalizations, which *require a significant amount of comparable data to be made.*
- ▶ How fine grained should they be?
 - In *FLP*, *robustness* is the ability to *ignore* variation in the realization of a category to be identified.
 - Any kind of robustness is based on the assumption of an *intended target!*
 - The category system used must be sufficiently fine grained for the *variation we want to identify and analyze.*
- ▶ Some conjectures:
 - Pre-theoretic classes close to the empirical observations are best-suited for annotation of the emergent, individual nature of interlanguage.
 - To provide access to the right level of abstraction for a range of research questions → multiple levels of annotation

On Automatically Analyzing Learner Language

Demter Meurers

Introduction
Why Analyze Learner Language?
Contact Points with CL

Learner Corpora
Data in SLA Research
Corpus annotation
Linguistic Annotation

Annotation Case Study
Automatic POS-Tagging
Three Sources of Evidence
Mismatching Evidence

Categories for Learner Language
Systematic POS for Learner Language
On the nature of interlanguage categories
Comparative fallacy
Variation and robustness
Syntactic annotation
Target hypotheses
Task-specific learner corpora
Comparing Meaning in Context

Conclusion
UNIVERSITÄT TUBINGEN

20 / 27

On the nature of categories for learner language

Consequences for syntactic annotation

- ▶ Idea: break down constituency in terms of
 - overall topology of a sentence (Hirschmann et al. 2007)
 - chunks and chunk-internal word order (Abney 1997)
 - dependency
- ▶ What is the empirical basis of dependency analysis?
 - dissociation of morphological, syntactic, and semantic dependencies (cf. also Meaning Text Theory, Mel'čuk 1988)
- ▶ Some work on dependency analysis of learner language:
 - surface-evidence based (Dickinson & Ragheb 2009)
 - goal: fine-grained record of morphological & syntactic evid.
 - canonical dependencies (MacWhinney 2008; Rosén & Smead 2010; Ott & Zial 2010; Hirschmann et al. 2010)
 - goal: robustly abstract away from learner specific forms
 - e.g., in CoMiC: robust construction of LRS semantics for comparing the meaning of answers to reading comprehension questions (Hahn & Meurers 2011)

Target hypotheses

- ▶ Targets are assumed for any kind of robust classification.
 - ▶ Target hypotheses should be explicit part of annotation.
 - (Lüdeling et al. 2005; Hirschmann et al. 2007; Lüdeling 2008).
 - ▶ Fitzpatrick & Seegmiller (2004): unsatisfactory levels of agreement in determining learner targets for error annot.
 - But keeping the target hypothesis implicit results in annotations which diverge even more unsatisfactorily.
- ⇒ Target hypotheses should be made explicit for any annotation robustly generalizing over variation.
- ▶ Which type of target hypotheses support reliable annotation of which distinctions?
 - ▶ Which evidence is needed to reliably determine such target hypotheses?

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.*

- Cannot be determined just by the learner sentences alone, without task information.

Difficulty of determining target hypotheses

- ▶ How can one obtain a better handle on target hypotheses?
 - take explicit task context into account
 - support targets other than fully explicit surface forms
 - target = minimal commitment required to support annotation
 - focus on more advanced learners
 - take learners and *learner strategies* into account
 - Learners, e.g., lift material from text or use known L2 chunks instead of trying to express appropriate meaning.
 - Example from CREG Corpus (Meurers et al. submitted)

Text: ... **Im Sommerurlaub, fahre ich manchmal nach Eutin.** ...
In the summer vacation, travel I sometimes to Eutin.

Question: **Welche 2 Städte besucht Heike im Urlaub?**
"Which 2 cities does Heike visit during her vacation?"

Student: **Fahre ich manchmal nach Eutin.**
Travel I sometimes to Eutin.

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.
 - = taking task, strategic competence, and L1 into account in learner models of Tutoring Systems (Amaral & Meurers 2008)
 - ▶ clear connection to language testing research
 - ▶ 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 corpora which are interpretable & representative of learner language, we need more language produced in a wide range of explicit task contexts.

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction

Why Analyze Learner Language?

Contact Points with CL

Learner Corpora

Data in SLA Research

Corpus Annotation

Linguistic Annotation

Annotation Case Study

Automatic POS-Tagging

Three Sources of Evidence

Mentorship Evidence

Categories for Learner Language

Systematic POS for Learner Language

On the nature of interlanguage categories

Comparative telecity

Variation and robustness

Synthetic annotation

Target hypotheses

Task-specific learner corpora

Comparing Meaning in Context

Conclusion

UNIVERSITÄT TUBINGEN

25 / 27

Using task-specific learner corpora

Comparing Meaning in Context Project (<http://purl.org/icall/comic>)

- ▶ Task-based corpora can also support an investigation of aspects such as meaning, information structure, . . .
- ▶ In collaboration with Nina Vyatkina (KU) and Kathy Corl (OSU) we are compiling a large corpus of answers to reading comprehension questions
 - ▶ written by US college students learning German
 - ▶ Text + Questions + Target Answers + Learner Answers, graded as adequate or inadequate response
- ▶ Makes it possible to study variation in forms used by language learners to realize the same meaning.
 - ▶ Supports research on information structuring (given/focus).
- ▶ On the practical side: Automatic content assessment of reading comprehension answers, currently with 84.6% accuracy (Meurers et al. 2011) on balanced test set.

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction

Why Analyze Learner Language?

Contact Points with CL

Learner Corpora

Data in SLA Research

Corpus Annotation

Linguistic Annotation

Annotation Case Study

Automatic POS-Tagging

Three Sources of Evidence

Mentorship Evidence

Categories for Learner Language

Systematic POS for Learner Language

On the nature of interlanguage categories

Comparative telecity

Variation and robustness

Synthetic annotation

Target hypotheses

Task-specific learner corpora

Comparing Meaning in Context

Conclusion

UNIVERSITÄT TUBINGEN

26 / 27

Conclusion

- ▶ We motivated linguistic annotation supporting 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.
 - ▶ The granularity of the annotation needed depends on the research question → multi-level annotation
 - ▶ Robust annotation depends on target hypotheses, which should be made explicit to obtain replicable annotation.
 - ▶ Inter-annotator agreement studies are crucial for establishing which distinctions can reliably be identified given information in corpus and its meta-information.
 - ▶ Explicit task and learner modeling can help constrain the search space of interpretation.
- ▶ Interdisciplinary collaboration between SLA & CL crucial to adapt annotation schemes & methods to learner language.

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction

Why Analyze Learner Language?

Contact Points with CL

Learner Corpora

Data in SLA Research

Corpus Annotation

Linguistic Annotation

Annotation Case Study

Automatic POS-Tagging

Three Sources of Evidence

Mentorship Evidence

Categories for Learner Language

Systematic POS for Learner Language

On the nature of interlanguage categories

Comparative telecity

Variation and robustness

Synthetic annotation

Target hypotheses

Task-specific learner corpora

Comparing Meaning in Context

Conclusion

UNIVERSITÄT TUBINGEN

27 / 27

References

- Aarts, J. & S. Granger (1998). Tag Sequences in Learner Corpora: a Key to Interlanguage Grammar and Discourse. In S. Granger (ed.), *Learner English on Computer*, London; New York: Longman, pp. 132–141.
- Abney, S. (1997). Partial Parsing via Finite-State Cascades. *Natural Language Engineering* 2, 337–344. URL: <http://www.vinarus.net/spa/97a.pdf>.
- Alexopoulos, T., H. Yannikoudakis & T. Briscoe (2011). From discriminative features to learner grammars: a data driven approach to learner corpora. Slides of talk given at Second Language Forum Research, Maryland. URL: http://people.pwf.com/ac.uk/ta259/SLRF10_dora.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: University. 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: Using 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* 26(3), 580–591. 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), 4–24. 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>.
- Arstein, R. & M. Poessio (2009). Survey Article: Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics* 34(4), 1–42. URL: <http://www.mitpressjournals.org/doi/abs/10.1162/col.07.034-R2>.
- Bailey, S. & D. Meurers (2008). Diagnosing meaning errors in short answers to reading comprehension questions. In J. Tetreault, J. Burstein & R. D. Feline (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>.

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction

Why Analyze Learner Language?

Contact Points with CL

Learner Corpora

Data in SLA Research

Corpus Annotation

Linguistic Annotation

Annotation Case Study

Automatic POS-Tagging

Three Sources of Evidence

Mentorship Evidence

Categories for Learner Language

Systematic POS for Learner Language

On the nature of interlanguage categories

Comparative telecity

Variation and robustness

Synthetic annotation

Target hypotheses

Task-specific learner corpora

Comparing Meaning in Context

Conclusion

UNIVERSITÄT TUBINGEN

27 / 27

Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning* 33(1), 1-17. URL <http://online.library.wiley.com/doi/10.1111/j.1467-1770.1983.tb09893.x.pdf>.

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.usp.edu.au/pubs/W98/W98-1207.pdf>.

Clahsen, H. & P. Myusken (1986). The availability of universal grammar to adult and child learners - a study of the acquisition of German word order. *Second Language Research* 2(2), 93-119. URL <http://slr.sagepub.com/content/2/2/93.abstract>.

Covington, M. A., C. He, C. Brown, L. Naqj & J. Brown (2006). *How complex is that sentence? A proposed revision of the Rosenberg and Abbeduto D-Level Scale*. Computer Analysis of Speech for Psychological Research (CASPR) Research Report 2006-01. The University of Georgia, Artificial Intelligence Center, Athens, GA. URL <http://www.ai.ua.edu/caspr/2006-01-Covington.pdf>.

de Haan, P. (2000). Tagging non-native English with the TOSCA-ICLE tagger. In C. Mair & M. Hundt (eds.), *Corpus Linguistics and Linguistic Theory*. Amsterdam: Rodopi, pp. 69-79.

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-0303.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 (ACL05)*, pp. 323-329. 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>.

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

Diaz Negrello, 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.

Diaz Negrello, A., D. Meurers, S. Valera & H. Wunsch (2010). Towards interlanguage POS annotation for effective learner corpora in SLA and FLT. *Language Focus* 39(1-2), 139-154. URL <http://purl.org/dm/papers/diaz-negrello-et-al-09.html>. Special Issue on Corpus Linguistics for Teaching and Learning. In Honour of John Sinclair.

Eghafari, 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. Seagrimmer (2004). *The Montreal electronic language database project*. In U. Connor & T. Upton (eds.), *Applied Corpora Linguistics: A Multidimensional Perspective*. Amsterdam: Rodopi. URL <http://ohs.montclair.edu/linguistics/MELD/widpaper.pdf>.

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.

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>.

Hahn, M. & D. Meurers (2011). On deriving semantic representations from dependencies: A practical approach for evaluating meaning in learner corpora. In *Proceedings of the Intern. Conference on Dependency Linguistics (DEPLING 2011)*. Barcelona.

Hawkins, J. A. & P. Buttery (2009). Using Learner Language from Corpora to Profile Levels of Proficiency - Insights from the English Profile Programme. In *Studies in Language Testing: The Social and Educational Impact of Language Assessment*. Cambridge: Cambridge University Press.

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

Hirschmann, H., S. Dootliffe & A. Lideling (2007). Syntactic annotation of non-canonical linguistic structures. In *Proceedings of Corpus Linguistics 2007*. Birmingham. URL http://www.linguistik.hu-berlin.de/institut/professoren/korpuslinguistik/meld/mirthe/better_innen_aranke.pdf#search=DotliffeLudelingCL2007.pdf.

Hirschmann, H., A. Lideling, 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://ajppj.oxfordjournals.org/content/30/4/461.full.pdf>.

Krivánek, J. & D. Meurers (2011). Comparing Rule-Based and Data-Driven Dependency Parsing of Learner Language. In *Proceedings of the Intern. Conference on Dependency Linguistics (DEPLING 2011)*. Barcelona.

Lakshmanan, U. & L. Selinker (2001). Analysing interlanguage: how do we know what learners know? *Second Language Research* 17(4), 393-420. URL <http://provy.ilo.ohio-state.edu/~lsl/te/ehost-live://search.ebscohost.com/login.aspx?direct=true&db=uih&AN=7939417&site=ehost-live>.

Lu, X. (2008). Automatic measurement of syntactic complexity using the revised developmental scale. In *Proceedings of the 21st International Florida Artificial Intelligence Research Society Conference (FLAIRS-08)*. Coconut Grove, FL: AAAI Press.

Lu, X. (2009). Automatic measurement of syntactic complexity in child language acquisition. *International Journal of Corpus Linguistics* 14(1), 3-28.

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 M. Walter & P. Grommes (eds.), *Förgeschritte in Lernerverwaltung: Korpuslinguistik und Zweitspracherwerb*. Tübingen: Max Niemeyer Verlag, pp. 119-140.

Lüdeling, A., M. Walter, E. Kroymann & P. Adolphs (2005). Multi-level error annotation in learner corpora. In *Proceedings of Con Linguistics*. Birmingham, UK. 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://chldes.psy.cmu.edu/grasp/morphosyntax.doc>.

Mačuk, I. (1988). *Dependency Syntax: Theory and Practice*. State University of New York Press. URL <http://books.google.com/books?id=dq29vrtAaC4pg-PR13n0ts-ZcCbImEAt7g&dq=Dependency%20syntax%20Theory&2nd=Practices&pg=PR13#v=onepage&q&f=false>.

Metcalf, V. & D. Meurers (2006a). Generating Web-based English Preposition Exercises from Real-World Texts. URL <http://purl.org/net/calico/handouts/eurocal06-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/calico/handouts/calico06-metcalf-meurers.pdf>. Pre-conference Workshop on NLP in CALL - Computational and Linguistic Collocations. CALICO 2006. May 17, 2006. University of Hawaii.

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. (2012). Natural Language Processing and Language Learning. In C. A. Chapelle (ed.), *Encyclopedia of Applied Linguistics*. Oxford: Blackwell. URL <http://purl.org/dm/papers/meurers-12.html>. To appear.

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

Meurers, D., N. Ott & R. Ziai (submitted). Creation and Analysis of a Reading Comprehension Exercise Corpus: Towards Evaluating Meaning in Context. In T. Schmidt & K. Wörner (eds.), *Multilingual Corpora and Multilingual Corpus Analysis*. Amsterdam: Benjamins, Hamburg Studies in Multilingualism (HSM). URL <http://purl.org/dm/papers/meurers-ott-ziai-draft.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 & J. Kopp (2011). Evaluating Answers to Reading Comprehension Questions in Context: Results for German and the Role of Information Structure. In *Proceedings of the TextReit 2011 Workshop on Textual Entailment*. Edinburgh, Scotland: UK: Association for Computational Linguistics, pp. 1-9. URL <http://www.aclweb.org/anthology/W11-2401>.

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

Meurers, W. D. & S. Müller (2009). *Corpora and Syntax* (Article 42). In A. Lideling & M. Krüß (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>.

Mura, S. (1998). *Hiroshima English Learners' Corpus: English learner No. 2 (English & 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.

Ortega, L. & H. Byrnes (eds.) (2008). *The longitudinal study of advanced L2 capacities*. Routledge.

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* 3(1-2), 9-30. URL <http://purl.org/dm/papers/ott-meurers-10.html>.

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction
Why Analyze Learner Language?
On the nature of interlanguage categories

Learner Corpora
Data in SLA Research
Corpus-based Linguistic Annotation

Annotation Case Study
Automatic POS-tagging
Three Sources of Evidence
Merchandising Evidence

Categories for Learner Language
Systematic POS for Learner Language
On the nature of interlanguage categories

Comparative Method
Variation and robustness
Syntactic annotation
Target hypothesis
Task-specific learner corpora
Comparing Meaning in Context

Conclusion

UNIVERSITÄT TüBINGEN

27/27

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction
Why Analyze Learner Language?
On the nature of interlanguage categories

Learner Corpora
Data in SLA Research
Corpus-based Linguistic Annotation

Annotation Case Study
Automatic POS-tagging
Three Sources of Evidence
Merchandising Evidence

Categories for Learner Language
Systematic POS for Learner Language
On the nature of interlanguage categories

Comparative Method
Variation and robustness
Syntactic annotation
Target hypothesis
Task-specific learner corpora
Comparing Meaning in Context

Conclusion

UNIVERSITÄT TüBINGEN

27/27

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction
Why Analyze Learner Language?
On the nature of interlanguage categories

Learner Corpora
Data in SLA Research
Corpus-based Linguistic Annotation

Annotation Case Study
Automatic POS-tagging
Three Sources of Evidence
Merchandising Evidence

Categories for Learner Language
Systematic POS for Learner Language
On the nature of interlanguage categories

Comparative Method
Variation and robustness
Syntactic annotation
Target hypothesis
Task-specific learner corpora
Comparing Meaning in Context

Conclusion

UNIVERSITÄT TüBINGEN

27/27

On Automatically Analyzing Learner Language

Detmar Meurers

Introduction
Why Analyze Learner Language?
On the nature of interlanguage categories

Learner Corpora
Data in SLA Research
Corpus-based Linguistic Annotation

Annotation Case Study
Automatic POS-tagging
Three Sources of Evidence
Merchandising Evidence

Categories for Learner Language
Systematic POS for Learner Language
On the nature of interlanguage categories

Comparative Method
Variation and robustness
Syntactic annotation
Target hypothesis
Task-specific learner corpora
Comparing Meaning in Context

Conclusion

UNIVERSITÄT TüBINGEN

27/27

- Ott, N. & R. Zhai (2010). Evaluating Dependency Parsing Performance on German Learner Language. In M. Dickinson, K. Müllnseip & M. Passaroti (eds.), *Proceedings of the Ninth International Workshop on Treebanks and Linguistic Theories*. vol. 9 of NEALT Proceeding Series, pp. 175–186. URL: <http://hdl.handle.net/10062/15960>.
- Pienemann, M. (1998). *Language Processing and Second Language Development: Processability Theory*. Amsterdam: John Benjamins.
- Rosenberg, S. & L. Abbeduto (1987). Indicators of linguistic competence in the peer group conversational behavior of mildly retarded adults. *Applied Psycholinguistics* 8, 19–32. URL: <http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=3607480&fulltextType=FA&fileId=S014271640000047>.
- 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.sls.uni-tuebingen.de/~zms/ies/AnnotCorp05/materials/sampson_babarczy03.pdf.
- Scarborough, H. S. (1990). Index of Productive Syntax. *Applied Psycholinguistics* 11(1), 1–22. URL: <http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=2747016>.
- 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.
- 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.
- Voutilainen, A. & T. Järvinen (1995). Specifying a shallow grammatical representation for parsing purposes. In *Proceedings of the 7th Conference of the EAACL*. Dublin, Ireland. URL: http://portal.acm.org/ft_gateway.cfm?id=977003&type=pdf&coll=GUIDE&dl=GUIDE&CFID=47108142&CFTOKEN=71182750.
- Vygotsky, L. S. (1986). *Thought and Language*. Cambridge, MA: MIT Press.
- Wiersma, W., J. Nerbonne & T. Lauttamus (2011). Automatically Extracting Typical Syntactic Differences from Corpora. *Literary and Linguistic Computing* 26(1), 107–124. URL: <http://llc.oxfordjournals.org/content/26/1/107.abstract>.
- 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

Detlev Meurers

Introduction

Why Analyze Learner Language?

Contact Points with CL

Learner Corpora

Diets in SLA Research

Corpus annotation

Linguistic Annotation

Annotation

Case Study

Automatic POS-Tagging

Three Sources of Evidence

Mismatching Evidence

Categories for Learner Language

Systematic POS for Learner Language

On the nature of interlanguage categories

Comparative fallacy

Iteration and robustness

Syntactic annotation

Target hypotheses

Task-specific learner corpora

Comparing Meaning in Context

Conclusion

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN

UNIVERSITÄT TUBINGEN