Enhancing Authentic Texts for Language Learners

Detmar Meurers
Universität Tübingen

based on joint work with Adriane Boyd, Ramon Ziai, Luiz Amaral,
Aleksandar Dimitrov, Vanessa Metcalf and Niels Ott

Workshop on the Automatic Analysis of Learner Language III (AALL 2011)
Victoria, Canada – May 18, 2011

UNIVERSITÄT
TÜBINGEN

Motivation

- For second language acquisition, contextualized meaningful use of the language to be learned is crucial.
- At the same time, learners benefit from a **focus on form** to overcome incomplete or incorrect knowledge.
    - *focus on form*: "an occasional shift of attention to linguistic code features" (Long & Robinson 1998)
- There is no learning without awareness, but awareness without input is not sufficient (Schmidt 1995).
- Strategies highlighting the salience of language forms and categories are referred to as **input enhancement** (Sharwood Smith 1993).

⇒ Let's use NLP to provide automatic input enhancement for language learners! → WERTi

---

WERTi: Working with English Real Text

- Provide learners of English (ESL) with input enhancement for any web pages they are interested in.
- good for **learner motivation**:
    - learners can choose material based on their interests
    - includes up-to-date information, news, hip stuff
    - pages remain fully contextualized (audio, video, links)
- wide range of potential **learning contexts**:
    - can supplement *traditional*, *distance*, or *individualized* instruction
    - can contribute to the voluntary, self-motivated pursuit of knowledge → *lifelong learning*
    - can support *implicit learning* for adult immigrants:
        - already functionally living in second language environment, but stagnating in acquisition
        - without access or motivation to engage in explicit learning, but browsing the web for information and entertainment

---

What language properties should we enhance?

- A wide range of linguistic features can be relevant for awareness, incl. morphological, syntactic, semantic, and pragmatic information (Schmidt 1995).
- We focus on enhancing language patterns which are well-established difficulties for ESL learners:
    - determiner and preposition usage
    - noun countability
    - use of gerunds vs. *to*-infinitives
    - phrasal verbs
    - *wh*-question formation
    - passive voice

    NLP identifying other patterns can be integrated as part of a flexible NLP architecture.

How should the targeted forms be enhanced?

- WERTi offers three types of input enhancement:
  a) color **highlighting** of the pattern or selected parts thereof
  b) support **clicking** interaction with automatic color feedback
     - The automatic feedback compares automatic annotation of selected form with original text
  c) support **fill-in-the-black** and **multiple choice** practice with automatic color feedback

- This follows standard pedagogical practice ("PPP")
  a) receptive presentation
  b) presentation supporting limited interaction
  c) controlled practice
  d) (free production)

Prepositions: Presentation (Color)

E-mail this to a friend | Printable version

Cows also 'have regional accents'

Cows have regional accents like humans, language specialists have suggested.

They decided to examine the issue *after* dairy farmers noticed their cows had slightly different moos, depending on which herd they came *from*.

John Wells, Professor of Phonetics at the University of London, said regional twangs had been seen before in birds.

The farmers *in* Somerset who noticed the phenomenon said it may have been the result of the close bond *between* them and their animals.

Farmer Lloyd Green, *from* Glastonbury, said: "I spend a lot of time *with* my ones and they definitely moo *with* a Somerset drawl.

Cows moo **with** a regional twang
🔊 Listen: Cow moo recordings

SEE ALSO
"'Accent' confirms unique species | Highlands and Islands
"Brain bug changes woman's accent
10 Jul 06 | Staffordshire
"What makes you local?
18 Feb 05 | Magazine

RELATED INTERNET LINKS
"University of London phonetics department
The BBC is not responsible for the content of external internet links

TOP UK STORIES
"Warnings *before* Potters Bar crash
"Gaza ship Britons 'to head home'
"Walcott left out of England squad
News feeds

MOST POPULAR STORIES NOW
MOST READ | MOST READ
"Doubt cast on Israel raid account

Source: http://news.bbc.co.uk/2/hi/5277090.stm

Prepositions: Practice (FIB)

E-mail this [ ] a friend | Printable version

Cows also 'have regional accents'

Cows have regional accents like humans, language specialists have suggested.

They decided to examine the issue [ ] after dairy farmers noticed their cows had slightly different moos, depending on which herd they came [ ].

John Wells, Professor of Phonetics at the University [ ] London, said regional twangs had been seen before in birds.

The farmers [ ] Somerset who noticed the phenomenon said it may have been the result of the close bond between them and their animals.

Farmer Lloyd Green, from Glastonbury, said: "I spend a lot of time [ ] my ones and they definitely moo with a Somerset drawl.

Cows moo [ ] a regional twang
🔊 Listen: Cow moo recordings

SEE ALSO
"'Accent' confirms unique species
15 Aug 06 | Highlands and Islands
"Brain bug changes woman's accent
10 Jul 06 | Staffordshire
"What makes you local?
18 Feb 05 | Magazine

RELATED INTERNET LINKS
"University [ ] London phonetics department
The [ ] content of external internet sites

TOP UK STORIES
"Warnings [ ] Potters Bar crash
"Gaza ship Britons 'to head home'
"Walcott left out [ ] England squad
News feeds

MOST POPULAR STORIES NOW

Source: http://www.bbc.co.uk/2/hi/5277090.stm

Prepositions: Practice (Multiple Choice)

E-mail this to a friend | Printable version

Cows also 'have regional accents'

Cows have regional accents like humans, language specialists have suggested.

They decided to examine the issue [ ] dairy farmers noticed their cows had slightly different moos, depending on which herd they came from.

John Wells, Professor [ ] Phonetics **at** the University [ ] London, said regional twangs had been seen before in birds.

The farmers [ ] Somerset who noticed the phenomenon said it may have been the result of the close bond [ ] them and their animals.

Farmer Lloyd Green, [around] Glastonbury, said: "I spend a lot [ ] time with my ones and they definitely moo with a Somerset drawl.

Cows moo [ ] a regional twang
🔊 Listen: Cow moo recordings

SEE ALSO
"'Accent' confirms unique species
15 Aug 06 | Highlands and Islands
"Brain bug changes woman's accent
10 Jul 06 | Staffordshire
"What makes you local?
18 Feb 05 | Magazine

RELATED INTERNET LINKS
"University of London phonetics department
The BBC is not responsible for the content of external internet sites

TOP UK STORIES
"Major manhunt for Afghan soldier
"Unemployment dips to 2.47 million
"PM condemns sympathy for Moat
News feeds

MOST POPULAR STORIES NOW
MOST READ | MOST READ
"Experts claim to discover 'root cause' of male baldness

Source: http://news.bbc.co.uk/2/hi/5277090.stm

Prepositions: Presentation + Interaction (Click)

Car-free cities: an idea with legs

Car-free neighbourhoods are no unrealistic utopia – they exist all over Europe

'Not anti-car, just pro-choice' ... a cyclist in Vauban, Germany. Photograph: Sipa Press/Rex Features

A quarter of households in Britain – more in the larger cities, and a majority in some inner cities – live without a car. Imagine how quality of life would improve for cyclists and everyone else if traffic were removed from areas where people could practically choose to live without cars. Does this sound unrealistic, utopian? Did you know many European cities are already doing it?

Posted by Steve Melia Thursday 29 October 2009 08.00 GMT
guardian.co.uk

Life and style
Cycling

Environment
Ethical and green living
Travel and transport

Series
Bike blog

More from Green living blog on

Life and style
Cycling

Environment
Ethical and green living
Travel and transport

Source: http://www.guardian.co.uk/environment/green-living-blog/2009/oct/29/car-free-cities-neighbourhoods

---

Prepositions: Presentation + Interaction (Click)

Car-free cities: an idea with legs

Car-free neighbourhoods are no unrealistic utopia – they exist all over Europe

'Not anti-car, just pro-choice' ... a cyclist in Vauban, Germany. Photograph: Sipa Press/Rex Features

A quarter of households in Britain – more in the larger cities, and a majority in some inner cities – live without a car. Imagine how quality of life would improve for cyclists and everyone else if traffic were removed from areas where people could practically choose to live without cars. Does this sound unrealistic, utopian? Did you know many European cities are already doing it?

Posted by Steve Melia Thursday 29 October 2009 08.00 GMT
guardian.co.uk

Life and style
Environment

Series

More from Green living blog on

Life and style

Environment

More blogposts

Source: http://www.guardian.co.uk/environment/green-living-blog/2009/oct/29/car-free-cities-neighbourhoods

---

Phrasal verbs: Presentation (Colorize)

Laugh Lines

Funny Stuff From All Over

May 6, 2010, 11:14 AM

Letterman: 'They Don't Like Immigrants'

Monologue | Wednesday night on *"The Late Show With David Letterman"* on CBS: You folks been following the big British Petroleum oil spill in the Gulf of Mexico? I'm telling you, British Petroleum has **put** more birds **in** oil than Colonel Sanders.

I was thinking about this. Here's what I **came up** with. Now, in Arizona, you know about the new immigration law, where if you don't look like you belong there, they can **run** you **out of** the state? And they've got patrol cars driving around, **pulling up** to people, saying: "You don't look like you belong here. **Get out!**" So the deal is, in Arizona, they don't like immigrants. And I was thinking, well, that's odd, because right across the river there in California, they elected one governor.

E-mail This   Print   Share   Twitter   Sign in to Recommend

Source: http://laughlines.blogs.nytimes.com/2010/05/06/letterman-they-dont-like-immigrants/

Phrasal verbs: Practice (Fill-in-the-blank)

Laugh Lines

Funny Stuff From All Over

May 6, 2010, 11:14 AM

Letterman: 'They Don't Like Immigrants'

Monologue | Wednesday night on *"The Late Show With David Letterman"* on CBS: You folks been following the big British Petroleum oil spill in the Gulf of Mexico? I'm telling you, British Petroleum has **put** more birds [?] oil than Colonel Sanders.

I was thinking about this. Here's what I **came up** with. Now, in Arizona, you know about the new immigration law, where if you don't look like you belong there, they can **run** you [?] the state? And they've got patrol cars driving around, **pulling up** to people, saying: "You don't look like you belong here. **Get** **on** [?] !" So the deal is, in Arizona, they don't like immigrants. And I was thinking, well, that's odd, because right across the river there in California, they elected one governor.

E-mail This   Print   Share   Twitter   Sign in to Recommend

Source: http://laughlines.blogs.nytimes.com/2010/05/06/letterman-they-dont-like-immigrants/

Enhancing Authentic Texts for Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should we enhance it?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT TÜBINGEN

"The government says it is expanding access to university, but they are actually blocking people's aspirations and betraying a generation."

The government was forced to cap student numbers after discovering a £200m black hole in the university financing budget at the end of last year. Labour was accused of abandoning its pledge to expand higher education, addingpressure to a growing debate about how to fund the growing number of young people who want to do a degree. The government is due to announce a review of student finance.

The massive increase in applicants has put a strain on the university system this year, with one university forced to convert single bedrooms in halls into doubles, and others putting students up in hotels.

Source: http://www.guardian.co.uk/education/2009/oct/14/30000-miss-university-place

---

"The government says it is expanding access to university, but they are actually blocking people's aspirations and betraying a generation."

The government was forced to cap student numbers after [____] ? (discover) a £200m black hole in the university financing budget at the end of last year. Labour was accused of [____] ? (abandon) its pledge to expand higher education, addingpressure to a growing debate about how to fund the growing number of young people who want [____] ? (do) a degree. The government is due to announce a review of student finance.

The massive increase in applicants has put a strain on the university system this year, with one university forced to convert single bedrooms in halls into doubles, and others putting students up in hotels.

Source: http://www.guardian.co.uk/education/2009/oct/14/30000-miss-university-place

---

If someone takes drugs, they can become addictive depending on the drug. Overdoses typically happen with cocaine, opioids, benzos, especially mixing benzos and opioids (Xanex, Valium, or Klonopin).

Why do people use illegal drugs? | subject | [change]

Most illegal drugs cause people to become intoxicated[needs proving]. The slang term for this experience is "getting stoned" or "getting high." When a drug user is intoxicated, they may feel strange, happy, dizzy, or weird. Some drugs such as marijuana and hashish often make users feel sleepy and relaxed. Some drug users have feelings that they are floating or dreaming. Drugs such as LSD make people feel intensely; they make one see and feel things like never before, and think things about the world they would normally not. Some say it increases knowledge and creates wisdom. Other drugs such as Crystal Meth make users feel excited and happy and full of energy.

Source: http://simple.wikipedia.org/wiki/Illegal_drugs

If someone takes drugs, they can become addictive depending on the drug. Overdoses typically happen with cocaine, opioids, benzos, especially mixing benzos and opioids (Xanex, Valium, or Klonopin).

Why do people use illegal drugs? | subject | [change]

Most illegal drugs cause people to become intoxicated[needs proving]. The slang term for this experience is "getting stoned" or "getting high." When a drug user is intoxicated, they may feel strange, happy, dizzy, or weird. Some drugs such as marijuana and hashish often make users feel sleepy and relaxed. Some drug users have feelings that they are floating or dreaming. Drugs such as LSD make people feel intensely; they make one see and feel things like never before, and think things about the world they would normally not. Some say it increases knowledge and creates wisdom. Other drugs such as Crystal Meth make users feel excited and happy and full of energy.

Source: http://simple.wikipedia.org/wiki/Illegal_drugs

Realizing WERTi

Enhancing Authentic Texts for Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should it be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT TÜBINGEN

17/31

- Guiding ideas behind implementation:
  - Reuse existing NLP tools where possible
  - Support integration of a range of language patterns
- First WERTi prototype: http://purl.org/icall/werti-v1 (Amaral/Meurers/Metcalf at CALICO 2006, EUROCALL 2006)
  - implemented in Python using NLTK (Bird & Loper 2004), TreeTagger (Schmid 1994)
  - integrated into Apache2 web-server using mod_python
  - targeted determiners and prepositions in Reuters news
- How can we flexibly support integration of a wider range of language patterns using heterogeneous set of NLP?
  - → integrate NLP into UIMA-based architecture on server

WERTi architecture
Second prototype

Enhancing Authentic Texts for Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should it be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT TÜBINGEN

18/31

- reimplementation in Java (Dimitrov/Ziai/Ott)
- Tomcat servlet
- idea behind architecture:
  - use same core processing
  - demand-driven pattern-specific NLP
- input enhancement targets:
  - determiners
  - prepositions
  - gerunds vs. to-infinitives
  - tense in conditionals
  - wh-questions

Server
Fetch web page
UIMA
Identify text in HTML page
Tokenization
Sentence Boundary Detection
POS Tagging
Pattern-specific NLP
Browser
Colorize | Click | Practice

WERTi architecture
Current prototype: http://purl.org/icall/werti

Enhancing Authentic Texts for Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should it be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT TÜBINGEN

19/31

- To support sites requiring login, cookies, or dynamically generated text, move fetching of web page and text identification to client. → Firefox plugin (Adriane Boyd)

Server
Tokenization
Sentence Boundary Detection
POS Tagging
Pattern-specific NLP
Browser
Fetch web page
Identify text in DOM
Colorize | Click | Practice

Pattern-specific NLP

Enhancing Authentic Texts for Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should it be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT TÜBINGEN

20/31

- UIMA-based architecture (Ferrucci & Lally 2004)
  - each NLP tool **annotates** the input
    - OpenNLP tools, LingPipe tagger, TreeTagger, Constraint Grammar CG 3
  - UIMA data repository is common to all components (Götz & Suhre 2004)
- We use available pre-trained models for
  - TreeTagger with PennTreebank tagset
  - LingPipe Tagger with Brown tagset
  - OpenNLP tools (tokenizer, sentence detector, tagger, chunker)
- Specify input enhancement targets
  - in terms of standard annotation schemes
    - e.g., identify determiners via AT|DT|DTI|DTS|DTX using Brown tagset
  - using constraint-grammar rules (CG 3 compiler), e.g.:
    - 101 rules for gerunds vs. to-infinitives
    - 126 rules for wh-question patterns

Slide 1 (21/31)

Evaluating input enhancement techniques
Does input enhancement improve learning outcomes?

- Improving learning outcomes is the overall goal of WERTi and visual input enhancement in general.

- While some studies show an improvement in learning outcomes, the study of visual input enhancement sorely needs more experimental studies (Lee & Huang 2008).

- WERTi can systematically produce visual input enhancement for a range of language properties
  → Supports real-life **foreign language teaching studies** under a wide range of parameters.
  → Supports **lab-based experiments** to evaluate when input enhancement succeeds in making learners notice enhanced properties (eye tracking, ERP).

Enhancing Authentic Texts for Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should it be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
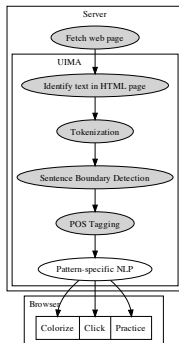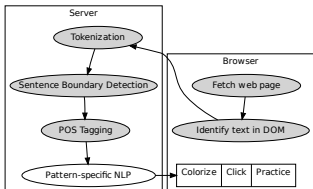Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT TÜBINGEN

21 / 31

Slide 2 (22/31)

Evaluating input enhancement techniques
Is the NLP adequate for automatic input enhancement?

- Automatic visual input enhancement requires reliable identification of the relevant classes using NLP.
  - Note: Precision of identification of specific classes relevant, not overall quality of POS-tagging or parsing!

- Problem 1: Often no established gold standard available for the language classes to be enhanced.

- Problem 2: Realistic test set should be based on pages chosen for enhancement by real learners.

Enhancing Authentic Texts for Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should it be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT TÜBINGEN

22 / 31

Slide 3 (23/31)

Evaluating input enhancement techniques
Evaluating determiner and preposition identification

- Evaluation of preposition and determiner identification using BNC Sampler Corpus
  - high quality CLAWS-7 annotation provides gold standard for preposition and determiner classes
  - relatively broad representation of English, and prepositions and determiners occur frequently in general

- Performance of the LingPipe POS tagger in WERTi:

	precision	recall
prepositions	95.07%	90.52%
determiners	97.06%	94.07%

Enhancing Authentic Texts for Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should it be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT TÜBINGEN

23 / 31

Slide 4 (24/31)

Related perspective
Data-Driven Learning

- One can view automatic input enhancement as an enrichment of Data-Driven Learning (DDL).
  - DDL is an "attempt to cut out the middleman [the teacher] as far as possible and to give the learner direct access to the data" (Boulton 2009, p. 82, citing Tim Johns)

- WERTi:
  - learner stays in control and directly accesses data,
  - but NLP uses 'teacher knowledge' about relevant language properties to make those more prominent to the learner.

Enhancing Authentic Texts for Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should it be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT TÜBINGEN

24 / 31

Related approaches

Enhancing
Authentic Texts for
Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should it be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT
TÜBINGEN

25 / 31

▶ Automatic Exercise Generation:
  ▶ MIRTO (Antoniadis et al. 2004)
  ▶ KillerFiller in VISL (Bick 2005)
  ▶ ClozeFox (Colpaert & Sevinc, cf.
    https://wiki.mozilla.org/Education/Projects/JetpackForLearning/Profiles/ClozeFox)

▶ Reading Support Tools:
  ▶ COMPASS (Breidt & Feldweg 1997)
  ▶ Glosser-RuG (Nerbonne et al. 1998)
  ▶ REAP (Heilman et al. 2008)
  ▶ ALPHEIOS (http://alpheios.net)

Some research issues
Nature of the reference for automatic feedback

Enhancing
Authentic Texts for
Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should it be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT
TÜBINGEN

26 / 31

▶ The automatic feedback for interaction and practice currently is based on the original text as gold standard.

▶ Where alternative correct answers exist, one needs to determine equivalence classes automatically.
  ▶ For prepositions, a data driven method could build on Elghafari, Meurers & Wunsch (2010).
  ▶ For passives, alternative word orders must be considered.

▶ For some practice enhancements supporting responses beyond the lexical level, specialized rules may need to replace extensional solution matching.

Research issues
Supporting users in choosing web pages

Enhancing
Authentic Texts for
Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should it be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT
TÜBINGEN

27 / 31

▶ In principle, any user-selected web page is enhanced.
  ▶ Users typically use standard Internet search engines (Google) to obtain candidate pages on a topic of interest.

▶ This works well for frequent targets (prep, det, . . .), but it does not ensure sufficient representation and balance of occurrence for other targets (questions, passives, . . .).

▶ A language aware search engine is needed to support retrieval and ranking based on
  ▶ content of interest to learner
  ▶ global readability
  ▶ language properties to be enhanced
→ LAWSE (Ott & Meurers 2010)

Research issue
Targets of enhancement

Enhancing
Authentic Texts for
Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should it be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT
TÜBINGEN

28 / 31

▶ Which language pattern types should be input enhanced?
  ▶ e.g., adverb placement, tense and aspect
    ▶ while tense/aspect involves complex semantic distinctions, lexical cues can be identified by the NLP ("*usually* go" vs. "are going *tomorrow*")

▶ Which aspects of language should be enhanced?
  ▶ targets: lexical classes, morphemes, syntactic patterns
  ▶ contextual clues for targets (optional or obligatory)

▶ How is it best determined which of the target instances on a page should be enhanced for practice?

▶ What is the best input enhancement
  ▶ for a particular linguistic pattern,
  ▶ given a specific web page with its existing visual design features (colors, fonts)?

Research outlook

How should different use cases be taken into account?

- How can automatic input enhancement best support
  - traditional classroom teaching, distance education, individualized instruction
  - lifelong learning, immigrant information needs?
- Where teachers are involved,
  - what aspects should we given them control over?
  - what information should they be able to access and track?
  Should WERTi offer test or exercise generation modes with explicit teacher control?
- For foreign language teaching, explicit meta-linguistic information and dictionary lookup would be useful.
- For immigrants satisfying information needs, translation dictionaries and automatic translation could be useful,
  - whereas translation is generally viewed as problematic in current foreign language teaching.

Enhancing
Authentic Texts for
Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should the be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT
TÜBINGEN

29/31

Research issues

Learner modeling

- Currently, the system does not take learner properties into account or keep track of previous system interaction.
  - Where should learner modeling, i.e., information about the learner and their interaction history be integrated?
  - Such records could also be used for SLA research and to improve the system.
- While keeping learners in control is well-motivated, they might appreciate integrating suggestions for web pages and enhancements from peers or social networks.

Enhancing
Authentic Texts for
Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should the be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT
TÜBINGEN

30/31

Conclusion

- We motivated and discussed an approach providing automatic input enhancement of authentic web pages.
  - The sentences turned into activities can remain fully contextualized as part of the pages selected by learner.
  - NLP identifies relevant linguistic categories and forms.
- NLP analysis offers interesting opportunities in the context of language learning
  - analyzing **language for learners**
    → automatic input enhancement
  - analyzing **learner language**
    → immediate feedback on form and contents in ITS
- Interdisciplinary collaboration integrating
  - linguistic modeling and NLP,
  - Foreign Language Teaching practice, and
  - Second Language Acquisition research
  is crucial for sustainable progress in this field.

Enhancing
Authentic Texts for
Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should the be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT
TÜBINGEN

31/31

References

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.

Antoniadis, G., S. Echinard, O. Kraif, T. Lebarbé, M. Loiseau & C. Ponton (2004). NLP-based scripting for CALL activities. In L. Lemnitzer, D. Meurers & E. Hinrichs (eds.), *Proceedings of eLearning for Computational Linguistics and Computational Linguistics for eLearning, International Workshop in Association with COLING 2004.*. Geneva, Switzerland: COLING, pp. 18–25. URL
http://aclweb.org/anthology/W04-1703.

Bick, E. (2005). Grammar for Fun: IT-based Grammar Learning with VISL. In P. Juel (ed.), *CALL for the Nordic Languages*, Copenhagen: Samfundslitteratur, Copenhagen Studies in Language, pp. 49–64. URL
http://beta.visl.sdu.dk/pdf/CALL2004.pdf.

Bird, S. & E. Loper (2004). NLTK: The Natural Language Toolkit. In *Proceedings of the ACL demonstration session*. Barcelona, Spain: Association for Computational Linguistics, pp. 214–217. URL
http://aclweb.org/anthology/P04-3031.

Boulton, A. (2009). Data-driven Learning: Reasonable Fears and Rational Reassurance. *Indian Journal of Applied Linguistics* 35(1), 81–106.

Breidt, E. & H. Feldweg (1997). Accessing Foreign Languages with COMPASS. *Machine Translation* 12(1–2), 153–174. URL

Enhancing
Authentic Texts for
Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should the be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT
TÜBINGEN

31/31

Enhancing
Authentic Texts for
Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should it be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT
TÜBINGEN

31/31

http://www.springerlink.com/content/v833605061168351/fulltext.pdf. Special Issue on New Tools for Human Translators.

Doughty, C. & J. Williams (eds.) (1998). *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press.

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.

Ferrucci, D. & A. Lally (2004). UIMA: An architectural approach to unstructured information processing in the corporate research environment. *Natural Language Engineering* 10(3–4), 327–348. URL http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=252253&fulltextType=RA&fileId=S1351324904003523.

Götz, T. & O. Suhre (2004). Design and implementation of the UIMA Common Analysis System. *IBM Systems Journal* 43(3), 476–489. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.90.5824&rep=rep1&type=pdf.

Heilman, M., L. Zhao, J. Pino & M. Eskenazi (2008). Retrieval of Reading Materials for Vocabulary and Reading Practice. In *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications (BEA-3) at ACL'08*. Columbus, Ohio: Association for Computational Linguistics, pp. 80–88. URL http://aclweb.org/anthology/W08-0910.

Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (ed.), *Perspectives on Pedagogical Grammar*, Cambridge: Cambridge University Press, pp. 293–313.

Lee, S.-K. & H.-T. Huang (2008). Visual Input Enhancement and Grammar Learning: A Meta-Analytic Review. *Studies in Second Language Acquisition* 30, 307–331.

Lightbown, P. M. (1998). The importance of timing in focus on form. In Doughty & Williams (1998), pp. 177–196.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, C. Kramsch & R. Ginsberg (eds.), *Foreign language research in cross-cultural perspective*, Amsterdam: John Benjamins, pp. 39–52.

Long, M. H. & P. Robinson (1998). Focus on form: Theory, research, and practice. In Doughty & Williams (1998), pp. 15–41.

Metcalf, V. & D. Meurers (2006). 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.

Nerbonne, J., D. Dokter & P. Smit (1998). Morphological Processing and Computer-Assisted Language Learning. *Computer Assisted Language Learning* 11(5), 543–559. URL http://url.let.rug.nl/nerbonne/papers/call_fr.pdf.

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

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.

Enhancing
Authentic Texts for
Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should it be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT
TÜBINGEN

31/31

Schmidt, R. (1995). Consciousness and foreign language: A tutorial on the role of attention and awareness in learning. In R. Schmidt (ed.), *Attention and awareness in foreign language learning*, Honolulu: University of Hawaii Press, pp. 1–63.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition* 15, 165–179.

Enhancing
Authentic Texts for
Language Learners

Detmar Meurers

Motivation
Input Enhancement
What should we enhance?
How should it be enhanced?

Example activities
Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions

Realizing WERTi
First prototype
Architecture of Java version
Architecture of Plugin version
Pattern-specific NLP

Towards evaluation
Evaluating learning outcomes
Evaluating the NLP

Related work

Research issues
Automatic feedback
Language-aware search
Targets of enhancement
Different use cases
Learner modeling

Conclusion

UNIVERSITÄT
TÜBINGEN

31/31