Enhancing Authentic Web Pages for Language Learners

Detmar Meurers, Ramon Ziai, Luiz Amaral, Adriane Boyd, Aleksandar Dimitrov, Vanessa Metcalf, Niels Ott

University of Tübingen
University of Massachusetts Amherst
The Ohio State University
Los Angeles
June 5, 2010

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
  - use of gerunds vs. to-infinitives
  - wh-question formation
  - phrasal verbs

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

Motivation

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 news, up-to-date information, hip stuff
  - pages remain fully contextualized (video, audio, links)

⇒ wide range of potential learning contexts:
  - can supplement regular classroom instruction
  - can support voluntary, self-motivated pursuit of knowledge, i.e., lifelong learning.
  - can foster implicit learning, e.g., for adult immigrants:
    - already functionally living in second language environment, but stagnating in acquisition
    - without access/motivation to engage in explicit learning, but browsing the web for information and entertainment

What should we enhance?

For successful second language acquisition, meaningful, task-based use of the language to be learned is essential.

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

Insights from Second Language Acquisition Research

Our starting point

- At the same time, learners benefit from or may require a so-called focus on form to overcome incomplete or incorrect knowledge (Long 1991; Lightbown 1998).
  - Focus on Form: “an occasional shift of attention to linguistic code features” (Long & Robinson 1998, p. 23).

- 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

Context and related work

ICALL: ILTs and ATICALL
Data-driven learning
Exercise Generation
Reading Support Tools
Summary
Outlook

Evaluation

Evaluating learning outcomes Evaluating the NLP

Context and related work

ICALL: ILTs and ATICALL
Data-driven learning
Exercise Generation
Reading Support Tools
Summary
Outlook

Exercise Generation

Data-driven learning

Evaluation

Evaluating learning outcomes Evaluating the NLP

Context and related work

ICALL: ILTs and ATICALL
Data-driven learning
Exercise Generation
Reading Support Tools
Summary
Outlook

Enhancing Web Pages for Language Learners

Meurers, Ziai, Amaral, Boyd, et al.

Motivation

WERTi

What should we enhance?

How should it be enhanced?

Example activities

Prepositions
Gerunds vs. to-infinitives
Wh-questions
Re-ranking WERTi
WERTi architecture
Pattern-specific NLP

Evaluation

Evaluating learning outcomes Evaluating the NLP

Context and related work

ICALL: ILTs and ATICALL
Data-driven learning
Exercise Generation
Reading Support Tools
Summary
Outlook

Enhancing Authentic Web Pages for Language Learners

Meurers, Ziai,
Amaral, Boyd, et al.

Motivation

WERTi

What should we enhance?

How should it be enhanced?

Example activities

Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions
Re-ranking WERTi
WERTi architecture
Pattern-specific NLP

Evaluation

Evaluating learning outcomes Evaluating the NLP

Context and related work

ICALL: ILTs and ATICALL
Data-driven learning
Exercise Generation
Reading Support Tools
Summary
Outlook

Enhancing Authentic Web Pages for Language Learners

Meurers, Ziai,
Amaral, Boyd, et al.

Motivation

WERTi

What should we enhance?

How should it be enhanced?

Example activities

Prepositions
Phrasal verbs
Gerunds vs. to-infinitives
Wh-questions
Re-ranking WERTi
WERTi architecture
Pattern-specific NLP

Evaluation

Evaluating learning outcomes Evaluating the NLP

Context and related work

ICALL: ILTs and ATICALL
Data-driven learning
Exercise Generation
Reading Support Tools
Summary
Outlook
How should the targeted forms be enhanced?

- WERTi currently offers three types of input enhancement:
  - color highlighting of the pattern or selected parts thereof
  - automatic feedback compares automatic annotation of clicked on form with targeted form
  - automatic feedback compares form entered by learner with form in original text
- This follows standard pedagogical practice (“PPP”):
  a) receptive presentation
  b) presentation supporting limited interaction
  c) controlled practice
  d) (free production)

Prepositions: Practice (FIB)

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 Weis, 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 cows and they definitely moo with a Somerset drawl."

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 was 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?

**Enhancing Web Pages for Language Learners**

**Motivation**

- WERTi
- What should we enhance?
- How should it be enhanced?
- Example activities

**Example activities**

- Prepositions
- Phrasal verbs
- Gerunds vs. to-infinitives
- Wh-questions
- Realizing WERTi

**WERTi architecture**

**Pattern-specific NLP**

**Evaluation**

- Evaluating learning outcomes
- Evaluating the NLP

**Context and related work**

- ICALL: ILTs and ATICALL
- Data-driven learning
- Exercise Generation
- Reading Support Tools

**Summary**

**Outlook**

---

**Phrasal verbs**

**Presentation (Color)**

**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.

---

**Phrasal verbs**

**Presentation + Interaction (Click)**

**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.

---

**Phrasal verbs**

**Presentation + Interaction (Click)**

**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.

---

**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 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.
Gerunds vs. infinitives: Presentation (Color)

"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, adding pressure 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

Gerunds vs. infinitives: Practice (FIB)

"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, adding pressure 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

Wh-questions: Presentation (Color)

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?**

Most illegal drugs cause people to become intoxicated. 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

Wh-questions: Presentation + Interaction (Click)

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?**

Most illegal drugs cause people to become intoxicated. 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
Enhancing Web Pages for Language Learners
Meurers, Ziai, Amaral, Boyd, et al.

Motivation
WERTi
What should we enhance? How should it be enhanced?
Example activities
Propositions
Phrasal verbs
Gerunds vs. to-infinitives

Wh-questions
Realizing WERTi
WERTi architecture
Pattern-specific NLP

Evaluation
Evaluating learning outcomes
Evaluating the NLP
Context and related work
CAL: ETs and ATICALL
Data-driven learning
Exercise Generation
Reading Support Tools
Summary
Outlook

Wh-questions: Practice (FIB)

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? [change]

Realizing WERTi

Guiding ideas behind implementation:
- Reuse existing NLP tools where possible
- Support integration of a range of language patterns

First WERTi prototype (Amaral/Meurers/Metcalf at CALICO 06, EUROCALL 06)
- implemented in Python using NLTK (Bird & Loper 2004), TreeTagger (Schmid 1994)
- integrated into Apache2 webserver using mod_python
- input enhancement targets: determiners and prepositions in Reuters news text
- still available at http://purl.org/icall/werti-v1

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

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

Wh-questions: Presentation + Interaction (Click)

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? [change]

Realizing WERTi

Guiding ideas behind implementation:
- Reuse existing NLP tools where possible
- Support integration of a range of language patterns

First WERTi prototype (Amaral/Meurers/Metcalf at CALICO 06, EUROCALL 06)
- implemented in Python using NLTK (Bird & Loper 2004), TreeTagger (Schmid 1994)
- integrated into Apache2 webserver using mod_python
- input enhancement targets: determiners and prepositions in Reuters news text
- still available at http://purl.org/icall/werti-v1

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

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

Pattern-specific NLP

- 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 | DTT | 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

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

beta version at: http://purl.org/icall/werti-plugin
Evaluating input enhancement techniques

High precision NLP needed 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 must be established by studying what pages learners choose for enhancement.

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
- Performance of the LingPipe POS tagger in WERTi:
  |
  | precision | recall |
  | prepositions | 95.07%  | 95.06%  |
  | determiners  | 97.06%  | 94.07%  |

Contextualizing our work

- NLP has received most attention in ICALL in connection with analyzing learner language to
  - provide feedback to the learner
  - guide learner through material according to performance
  - Note: Uses NLP to process learner language

Evaluating the NLP

- WERTi analyzes native language texts to
  - identify target language categories and forms to make learners aware of them and their context of use.
  - Note: Uses NLP to process well-formed, native language = Authentic Text ICALL (ATICALL)

Related work:
- Data-Driven Learning
- Automatic Exercise Generation
- Reading Support Tools

Evaluating learning outcomes

- Example activities
  - Prepositions
  - Gerunds vs. to-infinitives
  - Wh-questions
  - Identifying WERTi
  - WERTi architecture
  - Pattern-specific NLP

Evaluation

- Evaluating learning outcomes
  - Example activities
  - Prepositions
  - Gerunds vs. to-infinitives
  - Wh-questions
  - Identifying WERTi
  - WERTi architecture
  - Pattern-specific NLP

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, but NLP uses ‘teacher knowledge’ about relevant language properties to make those more prominent to the learner.
Related Work

Automatic Exercise Generation

- Antoniadis et al. (2004) describes plans of MIRTO project to support “gap-filling” and “lexical spotting” exercises in combination with a corpus database.
- VISL project (Bick 2005) offers games and visual presentations to foster knowledge of syntactic forms/rules.
  - KillerFiller produces slot-filler exercises from corpus texts; presented in isolation, in a testing setup.

Glosser-RuG (Nerbonne et al. 1998) supports reading of French texts for Dutch learners
- context-dependent dictionary, morphological analysis, and examples of word use in corpora
- COMPASS project (Breidt & Feldweg 1997): similar to Glosser-RUG, focusing on multi-word lexemes
- ALPHEIOS project (http://alpheios.net): supports lexicon lookup and provides aligned translations
- REAP project (http://reap.cs.cmu.edu) supports learners in searching for texts that are well-suited for providing vocabulary and reading practice (Heilman et al. 2008b).

Summary

- We motivated and discussed an approach providing automatic input enhancement of authentic web pages.
  - NLP identifies relevant linguistic categories and forms.
  - The sentences turned into activities can remain fully contextualized as part of the pages selected by learner.
- Automatic feedback for the practice activities is feasible since the original text is known.
- Next step: Where possible alternatives exist, determine equivalence classes automatically; e.g., for prepositions building on Elghafari, Meurers & Wunsch (2010).
- Web pages are selected by learners based on interests.
- Next step (Ott & Meurers submitted): Develop search engine which takes into account
  - content of interest to learner
  - general readability measures (Petersen 2007; Heilman et al. 2008a; Miltsakaki & Troutt 2008)
  - language properties to be input enhanced

Outlook: Questions to be addressed

- Which language pattern types should be input enhanced?
  - adverb placement
  - tense and aspect
    - while effect is semantic, lexical cues can be identified by NLP ("usually go" vs. "are going tomorrow")
    - passive vs. active
    - . . .
- Which aspect of the patterns should be input enhanced?
  - lexical classes, morphemes
  - contextual clues (optional or obligatory)
- What is the best input enhancement, i.e., highlighting or interaction possibilities
  - for a particular linguistic pattern,
  - given a specific web page with its existing visual design features?
References


