Bridging the Gap between Intentional and Incidental Vocabulary Acquisition

Oliver Streiter, Judith Knapp, Leonhard Voltmer
European Academy of Bolzano, Italy
{ostreiter;jknapp;lvoltmer}@eurac.edu

Daniel Zielinski.
University of Saarland, Germany
d.zielinskis@mx.uni-saarland.de.

The Dilemma of Vocabulary Acquisition

A dilemma in vocabulary acquisition is the antagonistic advantages of the commonly distinguished .intentional. and .incidental. vocabulary acquisition. Intentional vocabulary acquisition is memorizing straightforwardly term after term with their respective translations from a list. Intentional learning is quick and therefore usually preferred by learners, but it is also superficial. Learners encounter vocabulary in an isolated, often infinitive form and remain incapable of using it correctly in context. Moreover intentionally learned vocabulary sinks faster into oblivion. Didactically recommendable vocabulary acquisition exposes learners comprehensively to every term, embedding it deeply and solidly in the mental lexicon [1, 10]. Beneficial is also personalized vocabulary acquisition on authentic texts [6, 9, 13, 17].

Incidental vocabulary acquisition, namely through contextual deduction in target language reading, meets these recommendations. Learners encounter terms together with syntactic information, which helps using the accurate words in an idiomatic way. Vocabulary in context often appears repeatedly under different aspects and hence engrains in the learners. minds. Unfortunately it takes long until enough vocabulary for fluent conversations is incidentally gathered [3]. Problematic is, that deduction works best when new terms are mostly surrounded by familiar vocabulary [6]. In fact, with more than 5 to 10 new terms presented simultaneously, our retention capacity declines [12].

Gymn@zilla Combines annotated reading with term lists

Gymn@zilla (http://www.eurac.edu/gymnazilla) addresses this dilemma by dynamically annotating authentic text with definitions, translations, pictures, and other descriptive information. When learners access local and Internet documents through Gymn@zilla, server-side processing of texts dynamically adds links from every term to the corresponding entry in an open learning resource. Gymn@zilla employs stemming tools to match inflected word forms with dictionary entries. Learners receive linguistically enriched documents with their original link structure, so that they need only to move their mouse over a term to check it, or continue browsing.

Annotated reading is considered as a valuable feature in language learning [2, 11, 16] and implemented in several reading systems [4, 14, 15, 17]. However, all these systems have a closed, annotated corpus and a closed set of dictionary entries. None of them combine Internet browsing with annotation links to local and on-line dictionaries.
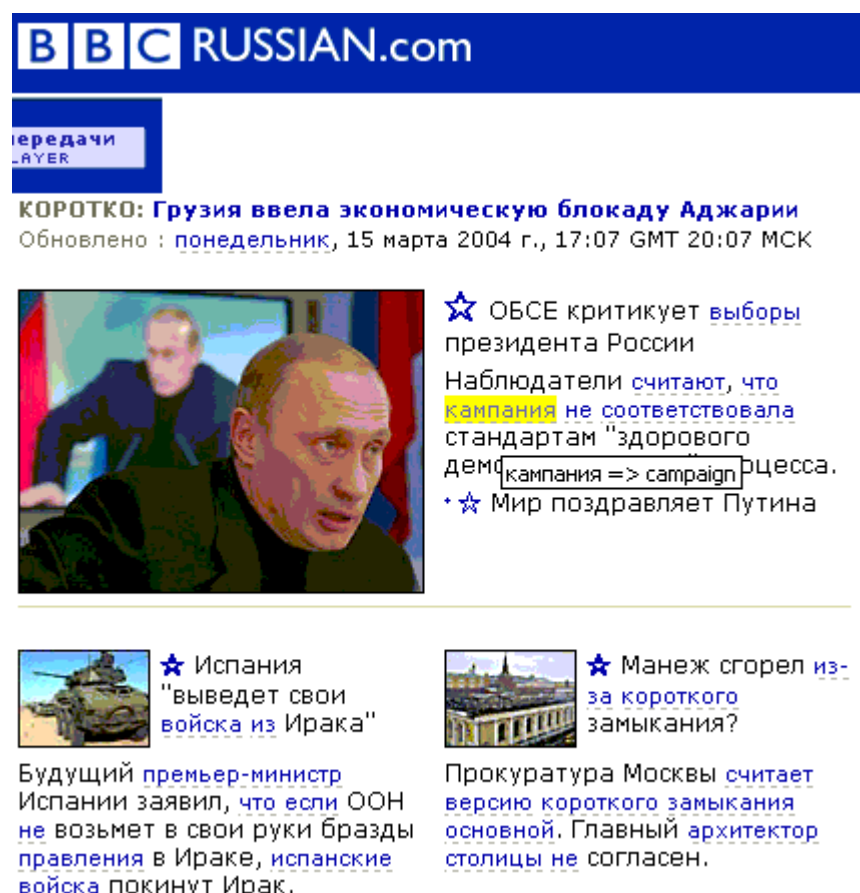
For intentional vocabulary acquisition, learners can collect terms and their translations from several sites in their personal word list by clicking on term links. In this way vocabulary acquisition occurs both incidentally by reading texts annotated with dictionary information and studying word lists extracted from this text. Depending on the underlying learning resource, Gymn@zilla exposes learners abundantly to new vocabulary in up-to-date, personalized contexts and activates the mental lexicon in several steps and on several levels.

Dynamic Interactivity

Traditional classroom teaching uses gap-filling and multiple choice quizzes for decades. Their usefulness is generally accepted regardless of the applied methodology. Quizzes can combine to an integrated vocabulary acquisition environment [13]. The main advantage of electronic learning material over traditional paper material is interactivity [18]. Authoring tools allow language teachers to manually create electronic true-

material is interactivity [18]. Authoring tools allow language teachers to manually create electronic true-false, multiple choice, matching, gap-filling, spelling, or sentence generation quizzes [5]. Learning environments exploit multimedia features [6, 9] and gap-filling quizzes for grammar training and even sentence formation [7, 19]. The effectiveness of such quizzes especially for weaker students has been shown in [8]. Figure 1 to 3 shows the transition from annotated reading via the construction of a word list to the interactive quizzes. To our knowledge, no other system offers interactive practice on annotated internet texts in similar ways.
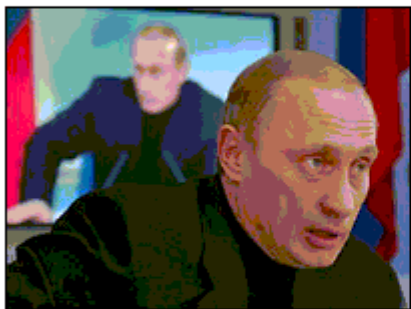
Figure 1: Annotated reading with Gymn@zilla.

Figure 2: Vocabulary List with Gymn@zilla.

выборы:: выборы => election, elections

ОБСЕ критикует **выборы** президента России Ирака"

считают (считать):: считать => believe, compute, consider, count, deem, judge, numerate, reckon, regard, repute

Наблюдатели **считают** , что кампания не соответствовала стандартам "здорового демократического"

кампания:: кампания => campaign

Наблюдатели считают , что **кампания** не соответствовала стандартам "здорового демократического"

премьер-министр:: премьер-министр => premier, prime minister

Будущий **премьер-министр** Испании заявил , что если ООН не возьмет в свои руки бразды правления в Ираке , испанские войска будут выведены из страны .

правления (правление):: правление => board, board of administration, government, regimen, reign

Будущий премьер-министр Испании заявил , что если ООН не возьмет в свои руки бразды **правления** в Ираке , испанские войска будут выведены из страны .

Figure 3: Interactive quiz with Gymn@zilla.

1) ОБСЕ критикует | выборы ▼ | президента России Ирака"

2) Наблюдатели | ▼ | , что кампания не соответствовала стандартам здорового демок

- правление
- пралвение
- правелния
- правления
- премеьр-министр
- премьер-министр
- считать
- счтиать
- считюат
- считают
- выобры
- выборы
- камапния
- кампания

3) Наблюдатели | ▼ | не соответствовала стандартам здорового демок

4) Будущий | | ании заявил , что если ООН не возьмет в свои руки бразды правлени кие войска будут выведены из страны .

5) Будущий прем | | нии заявил , что если ООН не возьмет в свои руки бразды | | , испанские войска будут выведены из страны .

(1) (2) (3) (4) (5)

The Gymn@zilla project and underlying technology

Gymn@zilla has been developed within the LOGOS-GAIAS project. It supports browsing a local text repository and the Internet by dynamically creating and annotating HTML pages with open dictionaries resources. Gymn@zilla is written in Perl. It is an on-line application running on a Linux web server . not a browser. Both components, Perl and GNU/Linux, guarantee the usage of free and powerful modules. The processing of web pages in real-time and generating exercises from it is a complex task, which involves the following steps: (1) mirroring of web pages, (2) linguistic processing and (3) generation of exercises.

1. Mirroring of web pages is done by using Perl.s LWP modules. All Hyperlinks in a web page pointing to other text documents are rewritten to Gymn@zilla.s URL in order to allow continuous browsing with Gymn@zilla. Links to multimedia documents such as audio, video and graphic files are preserved. In a next step encodings other than utf-8 are converted to utf-8. Documents in formats other than html such as *.doc, *.ps or *.pdf are converted to well formed xhtml by the use of GNU-tools.

step encodings other than utf-8 are converted to utf-8. Documents in formats other than html such as *.doc, *.ps or *.pdf are converted to well formed xhtml by the use of GNU-tools.

2. Once converted, the documents language is guessed before starting natural language processing. In order to annotate the text with linguistic information the text is first segmented into its tokens. Stemming is then done by the use of pattern matching techniques. According to the user.s preferences the text is then annotated with translations and terminological information from on-line dictionaries and terminological databases. The annotation is done by insertion of -tags with advanced link titles containing the linguistic information which will show up as a tooltip when the user moves the mouse over a word. With the help of a Javascript function link titles can be formatted like html-documents so that they may contain images and links to further information sources. Information can thus be structured from general to specific.

3. Each user in Gymn@zilla is associated with a session where history information is stored in order to memorize words seen by the user. This information is then used to make editable word lists and to generate cloze texts or other exercises for training.

Future steps in the development of Gymn@zilla comprise benchmarking, the expansion of linguistic resources integration of automatic document classification and the integration of a morpho-syntactic parser in order to improve linguistic analysis linguistic annotation.

References

1. [1] Jean Aitchison. Words in the Mind: An Introduction to the Mental Lexicon. Blackwell Publishers Ltd, Oxford, UK, third edition, 2003.
2. [2] Dorothy M. Chun. L2 reading on the Web: Strategies for Accessing Information in Hypermedia. Computer Assisted Language Learning, 14(5):367.403, 2001.
3. [3] Tom Cobb. Breath and depth of lexical acquisition with hands-on concordancing. Computer Assisted Language Learning, 12(4):345.360, October 1999.
4. [4] Luigi Colazzo and Marco Costantino. Multi-user hypertextual didactic glossaries. International Journal of Artificial Intelligence in Education, 9(1.2), 1998.
5. [5] Robert Godwin-Jones. Language testing tools and technologies. Language Learning & Technology, 5(2):8.12, 2001.
6. [6] Peter J.M. Groot. Computer assisted second language vocabulary acquisition. Language Learning & Technology, 4(1):60.81, May 2000. http://llt