

# Gymn@zilla: Language learning with the Internet

**Oliver Streiter, Judith Knapp, and Leonhard Voltmer**  
**European Academy Bozen-Bolzano**  
**{ostreiter, jknapp, lvoltmer}@eurac.edu**

Language learning with free internet resources is a seminal but underestimated approach. Research in Computational Linguistics routinely uses the internet's myriad of texts for term extraction, generation of language models and corpus building. Yet the internet is in itself a corpus even according to the more restrictive definition of [Kilgarriff and Grefenstette, 2003]: "A *corpus* is a collection of texts when considered as an object of language or literary study". Gymn@zilla sets out to exploit the Internet for corpus-based language learning in all its aspects.

Using corpora for language learning is a broad practice. Frequency analyses and concordance tables on large electronic text collections help choosing lexical concepts, important grammar and example sentences and allow the creation of personalized dictionaries [Cobb, 1999; Lamy and Goodfellow, 1999]. In teaching language for special purposes (LSP) corpora are exploited to extract terms and create glossaries [Bowker and Pearson, 2002].

We present the browser-like application Gymn@zilla for contextualized learning using the internet as an authentic language corpus. Gymn@zilla combines annotated reading in the Internet (top-down, incidental learning) with the possibility to explicitly memorize and interactively practice unknown items on personalized lists and quizzes (bottom-up, intentional learning). It supports also the classical NLP technologies morphological analysis, tagging, term extraction and frequency analysis.

Gymn@zilla is used like any browser. The program accesses the addressed web-page, identifies its language and encoding and performs a simple word-form stemming of text. The stemmed words and expressions are then linked to their respective translations in external dictionaries. The linked lemma is marked up as html tool-tip to provide an immediate translation aid even without following the external link (see figure 1).

Clicking on a word triggers two actions. First, the complete explanations of the external lexicon are opened. Second, the word, its context and its translation are added to a personal glossary. The learner can edit the vocabulary in his personal dictionary and use it for intentional vocabulary acquisition, as opposed to incidental vocabulary acquisition by annotated reading of the web-page. The interactive and spontaneous switching from one method to the other is a unique feature of Gymn@zilla.

The combination approach is motivated by the constructivist approach, according to which users actively create their own knowledge. In Gymn@zilla learners are not limited to follow pre-elaborated content from lesson to lesson, but they can instead browse, read, practice and study intentionally during and after the language course. A Gymn@zilla learner manages herself learning environment and progress.

Figure 1: Step one, annotated reading with Gymn@zilla



Last, the learner can create interactive quizzes from the personal glossary. Gymn@zilla can automatically create clozes and offer inflected, uninflected and misspelled forms to fill the gaps. Permutations of inflected word forms and spelling errors help train formal aspects of language like functional collocates (French: “répondre à”). The semantic and functional aspects of the language are stressed when the proposed solutions are near synonyms or other thematically related lexemes. Gymn@zilla will offer this option in the near future.

## Bibliography

- Bowker, L. and Pearson, J. 2002. *Working with Specialized Language – A practical guide to using corpora* London: Routledge.
- Cobb, T. 1999. “Breath and depth of lexical acquisition with hands-on concordancing.” *Computer Assisted Language Learning*, 12(4):345–360.
- Kilgarriff, A. and Grefenstette, G. 2003. “Introduction to the Special Issue on the Web as Corpus” in *Computational Linguistics* 29, 3: 333-347.
- Lamy, M. and Goodfellow, R. 1999. “Reflective conversation” in the virtual language classroom. *Language Learning & Technology*, 2(2):43–61.