Tibidiscis:
Online Service for Personalized
Language Learning

Helena Ahonen-Myka

Tibidiscis Ltd,
Helsinki, Finland
helena.ahonen-myka@tibidiscis.fi
http://www.tibidiscis.fi

Abstract. In this paper, we present a new online service for learning languages. It gives learners the possibility to add learning content that they find relevant, interesting and comprehensible for themselves. The service generates automatically fill-in-the-blanks exercises from the content given, provides feedback and stores the results for monitoring progress. The service makes it easy to maintain and enhance competence of many languages at the same time. We also describe the linguistic processing used to generate the exercises, particularly, how the gap words, i.e. the words that are left out, and the guiding hints in the gaps, are selected.

Keywords: language learning, e-learning, multilingualism, morphological analysis, text analysis

1 Introduction

According to the information page of the European Day of Languages, there are over 200 European languages, 24 official EU languages, about 60 regional/minority languages, and many more spoken by people from other parts of the world [1]. It is widely accepted that foreign language skills provide competitive advantage both for businesses and job seekers [2]. In addition to the needs of international trade, we are now increasingly facing the language learning needs of immigrants: many people are on the move to find work and better life in Europe, and one of the most central prerequisites for an active citizenship is to learn the language(s) of the new home.

Despite this, children are learning fewer and fewer languages, and the range of languages learned is getting narrower. According to the newest Eurostat statistics, the proportion of upper secondary general education students learning two or more foreign languages decreased in most of the EU countries between 2009 and 2014 [3].

Why is that? We think that language learning should be made more attractive and an integrated part of everyday life, such that learners can learn the kind of language they need for their work and other interests, anytime and anywhere.

We have developed and operate an online service Tibidiscis[1], which lets language learners concentrate on language that is relevant to them. In order to offer personalized exercises efficiently, we have developed automatic methods to analyze text given by the learners and to create exercises based on this content.

Our approach makes it easy to maintain and enhance competence of several languages. Even if a learner has only a few minutes to spare, it is easy to go to the learning site and do some exercises.

2   Tibidiscis Language Learning Service

Tibidiscis lets learners learn languages using texts that they have chosen and added themselves. The service generates automatically fill-in-the-blanks exercises from the text given, provides feedback, and saves the learning results for monitoring progress. At the moment, one can learn English, French, German, Spanish, Italian, Swedish, and Finnish. We have also experimented with some Finno-Ugric languages, like Northern Sami and Meadow Mari.

The service splits the given text into paragraphs and shows a paragraph at a time as an exercise, as seen in the example below.

```
Auf ___ anderen Seite ___ sich die Frage, inwieweit es Sinn macht, Bücher
in Katalogen ___zuweisen, die ___ gar keine Bibliothek zugänglich ge_____
werden können.
```

After the learner has submitted the exercise, the service shows the original snippet, with the answers highlighted according to if they are the same as or different from the original word. When the learner goes through the story again later, the gaps are shuffled anew, which makes the exercises always look different and fresh. The learner's attention is invited to varying parts of the text, and simply learning where one word reminds of the next word can be avoided.

```
___ der anderen Seite stellt sich ___ Frage, inwieweit es Sinn _____,
Bücher ___ Katalogen nachzuweisen, die über gar keine Bibliothek
___gänglich gemacht werden können.
```

Traditionally, fill-in-the-blanks exercises focus on one feature of language at a time, for instance, all articles are missing, or all verb forms have to be converted to a correct tense. In the Tibidiscis exercises, however, gaps can correspond to varying features of language, both morphological, syntactic, and semantic.

Languages are best learned in context. The meaning and form of a word is defined by the other words in its neighborhood. Although it is useful to be aware of the basic grammatical structures, language usage is highly idiosyncratic [4]

---

[1] www.tibidiscis.com

and it is impossible to include all the conventions in any dictionary, textbook, or language course.

Paul Nation says about learning vocabulary[5]:

> The two most important conditions supporting learning are spaced repetition and the quality of attention given to items. Quality is increased by recalling what you have met, making varied recalls, analyzing and elaborating on language items, and giving deliberate attention.

Tibidiscis clearly fulfills these conditions. The learner can repeat the exercises any number of times, with shorter or longer intervals, each time recalling different language items. It is also easy to add examples of similar (but varying) language usage situations to the Tibidiscis service. For instance, the learner could add all the emails received where a native speaker proposes a meeting, in order to learn different ways to express time and other details in such emails.

Our approach suits best for intermediate and advance level learners who easily find content they can understand but find hard to produce. However, even a rather restricted competence is enough, as there are many ways to aid comprehension. For instance, it is easier for learners to read a news story about something they already heard about in their own language or know of due to other background knowledge. If a learner reads a story about the British Royal Family, the knowledge of the members of the family and their relationships provides a good starting point for guessing the meaning of the unknown words and structures.

In addition to learning correct word forms or usage of specific language items, like prepositions or determiners of nouns, Tibidiscis exercises encourage deliberate thinking processes. Filling in the blanks can be compared to solving a crossword puzzle: the learner has to actively process the text to find out which words might make the content coherent. On the other hand, there is no negative pressure involved: the learner can always submit the exercise to see the correct answer. There is no penalty for a wrong answer.

3 Linguistic processing

Tibidiscis does not try to cover all the features of a language, but concentrates on features that are reasonable to learn with this kind of a method. The features included differ from language to language, for several reasons. First, the morphological and grammatical systems of languages differ greatly. For learners of English, the use of articles, prepositions, and phrasal expressions is a central part of learning, whereas the tense system of Spanish is more complicated than for English.

In German, expressions often contain a noun (or several) and a verb, like in the examples below. Usually it is rather easy to know which noun should be used, but choosing the correct verb, prepositions, and articles (in gender, number, and case), is harder. Moreover, many nouns and verbs have prefixes (like `nachfolgen` and `verfolgen` in the examples below), the exact meanings of which can only be learned in context.

```
Im Bereich Thermometer sind vorwiegend Produkte entwickelt worden,
die den bisher vermarkteten Produkten in ihrem Lebenszyklus
nachfolgen sollen.

Dabei verfolgen die Kooperationspartner meist ein gemeinsames Ziel
und am Ende profitieren beide Seiten auch wirtschaftlich
von der Kooperation.
```

In general, the focus is on the verbal expressions, as well as on functional features, like articles and prepositions. We do not focus on the nouns as such, as there is more variety within nouns, which may make them too difficult to guess if they are missing. There are also a lot of vocabulary training programs available that are particularly useful for learning nouns. However, as mentioned above, some functional features connected to nouns, like plural forms, cases, articles, or collocated prepositions, are more suitable for learning with Tibidiscis.

Second, the features covered depend on the linguistic information that can be extracted using the analysis tools available. Linguistic processing is used to select *gap words* and *hints*, that is, to decide which words can be left out of the paragraph and whether there should be a guiding hint, e.g. part of the word or its lemma, in the gap, or no hint at all.

Linguistic processing depends on the resources of each language. If there is no analyzer available for a language, we can still collect word forms for closed categories of the language, like pronouns and prepositions, as well as some frequent verbs, and use them as candidates for good gap words. Some languages also have many regular prefixes and suffixes that represent either syntactic or semantic features of the language. For short and frequent words hints are not necessary, whereas a natural hint for a word with a regular prefix (resp. suffix) is the word without the prefix (resp. suffix). For instance, if the word in the gap is `nachfolgen`, the hint could be `folgen`. As `folgen` is a rather frequent verb, also a hint `nach` could be possible.

If there is a morphological analyzer available, we can define that words of certain parts-of-speech are good gap words. For instance, we do not have to list all the prepositions and pronouns but say that all of them can be used as gap words in the exercises. For hints, a base form of the word in question is often available, although due to ambiguous cases, it has to be used with care. The base form is particularly useful, if irregular forms are common, or there are changes to a stem (e.g. consonant gradation, umlauts).

Language learning applications are quite sensitive to the quality of linguistic analyzers, but as we have decided to hold the word in the original text as the correct answer, the only risk caused by a false analysis is that the hints we offer are misleading. This risk can be reduced, if the analyzers are able to tell, whether the analysis is ambiguous, and how confident the analysis result is.

We do not need to cover all the features of the languages, and hence, we can just decide not to use some words in gaps, or give no hint for a word, if the gap or hint would be misleading. It is naturally more efficient for learning the usage

of some word, if the learner has to write it, but even if the learner just sees the word in varying contexts, it can be learned, although slower.

4   Future objectives

Our intention is to add more languages to the Tibidiscis service in the future. Therefore, multilingual and/or language-independent tools are a necessity for us. The linguistic processing has to be fast and automatic, as the learners are able to add their own texts, and hence, the processing cannot be done offline in the background.

We are also planning to have a special attention to multiword expressions. At the moment, multiword expressions (phrases, collocations) are learned rather accidentally, as they usually contain prepositions and other frequent language items that are considered good gap words. It might be more useful to be able to control how such an expression is presented to a learner. For instance, that not too many words are missing at the same time, and that for a fixed expression, a base form of a noun may not be a good hint.

5   Conclusion

We have presented an overview of a new language learning service for personalized and multilingual learning. The service takes content that a learner finds comprehensible and interesting and generates from it fill-in-the-blanks exercises automatically. We have also described how linguistic analysis is used depending on the grammatical systems and resources available for each language.

References

1. European Day of Languages, `http://ec.europa.eu/languages/information/languages-day_en.htm`
2. Study on Foreign Language Proficiency and Employability (Executive Summary). European Commission, Publications Office of the European Union, Luxembourg 2015.
3. Eurostat. Foreign language learning statistics, `http://ec.europa.eu/eurostat/statistics-explained/index.php/Foreign_language_learning_statistics`
4. Kathrin Steyer, Wolfgang Teubert. Deutsch-Französische Übersetzungsplattform. Deutsche Sprache Jg. 25 (1997).
5. Paul Nation. What do you need to know to learn a foreign language? School of Linguistics and Applied Language Studies, Victoria University of Wellington, New Zealand, 2014.