þÿ E v a l u a t i n g   s t u d e n t s   m e t a l i n g u i s t i c   k n o w l e d g e   w

Volodina, Elena

2014-11

Evaluating students' metalinguistic knowledge with Lärka

Elena Volodina[*], Therese Lindström Tiedemann[†]

*Språkbanken, Department of Swedish, University of Gothenburg
Box 200, 40530 Gothenburg, Sweden

[†]Department of Linguistics and Philology, Uppsala University
Box 635, 751 26 Uppsala, Sweden

`elena.volodina@svenska.gu.se, therese.lindstrom_tiedemann@lingfil.uu.se`

1. On corpora in education

Corpora have been used for a long time for research and in the past years they have gained popularity in teaching languages and linguistics. Corpora within language teaching is mainly used in the form of concordances and frequency lists (Hunston, 2002; O'Keeffe et al., 2007). However, corpora can offer a lot more than that to teaching, namely authentic data, grammatical patterns and language features. The latter include lexical, grammatical, morphological features, collocation patterns, semantic features etc. depending upon the linguistic parameters which have been annotated in the corpus.

The advantage of using corpora within the teaching of linguistics in an exercise generator is that learning materials can be customized to the individual needs of learners, courses or syllabus requirements; materials can be reused independent of time and place, automatic generation of teaching materials can save teachers' time on both production and correction of assessment items. In addition, students get instant feedback on each exercise and have the opportunity to get a final summary of their feedback including recommended categories to focus on. Furthermore, such systems automatically provide students with more exercises on categories they find difficult, something which can also mean that doing exercises will seem less tedious.

2. Lärka

Lärka (Lark) is a learning platform designed for learning Swedish: it supports both second language learners of Swedish and students of linguistics (Volodina et al., 2014). Lärka is a modular web-based exercise generator that reuses available annotated corpora and lexical resources, it is adaptable to student levels and allows students to focus on the areas of difficulty.

Lärka reuses Swedish resources from Korp and Karp, Korp being the corpus infrastructure (Borin et al., 2012b) and Karp the lexical infrastructure (Borin et al., 2012a) maintained at Språkbanken. Besides, a text-to-speech module and a readability module are employed to enhance the pedagogical value of the exercises.

For students of linguistics, Lärka offers exercises to train parts of speech (POS), syntactic relations (SYNT) and semantic roles (SEM). For each exercise, the student can choose the mode – diagnosis, self-study, test, or timed test; and he/she can select the categories to focus on. For POS and SYNT two difficulty levels are offered.

3. Evaluation setting

With this evaluation we wanted, primarily, to find out how well our students cope with metalanguage, what their problem areas are and how this compares to international studies (being aware that the material won't be the same and hence the studies will not be absolutely comparable).

In this study we have logged student performance during several linguistic courses offered at the University of Gothenburg and Uppsala University during the Spring Term, 2014. Students were also asked to fill out an online evaluation form.

Students worked with Lärka during 1-3 labs depending on the course. They initially worked with parts of speech, followed by syntactic relations and semantic roles. Each lab started with a diagnostic test, followed by a number of exercise types that students preferred, such as self-studies, tests and timed tests with options to focus on categories that were difficult or problematic. All students were asked to complete a new diagnostic test by the end of the lab to see whether their results had improved.

4. Evaluation results

We have logged 39339 answers during the period February, 10 – August, 26, 2014. The logs are distributed as follows between the exercise types (Table 1).

	Correct	Incorrect	Total
POS, total	15441 (72%)	6080 (29%)	21521
POS, diagnosis	5236 (78%)	1507 (22 %)	6743
SYNT, total	7369 (49%)	7733 (51%)	15101
SYNT, diagnosis	1850 (56%)	1461 (44%)	3311
SEM, total	1621(60%)	1095 (40%)	2716
SEM, diagnosis	449 (68 %)	208 (32%)	657
Total	24431(62%)	14908 (38%)	39339

Table 1. Distribution of logs between exercise types.

Results per category show that students are best at parts of speech, where the error rate is 22.35% in diagnostic mode. Syntactic relations they have less than a 50% chance to get right in the total sample, but a slightly better chance in diagnostic mode. Semantic roles they get wrong 31.66%. Interestingly, they do slightly better in

diagnostic mode for all exercises.

Even though semantic roles tend to be given less class room time and there are large differences between researchers and books as far as the terminology and categories go, students still do better on semantic roles, than syntactic relations.

The survey results confirm that Swedish linguistic students have some trouble with metalinguistic terminology but nowhere near as much as UK students (Alderson et al 1997; Alderson and Hudson 2013, cf also for Spain Corona and Mur-Dueñas 2010). Furthermore, materials used in Lärka are likely to be more difficult than the single sentence used for the British and Spanish studies, and hence it is interesting to see that the error rate is not higher. Instead, Swedish students do a lot better on some categories, e.g. adverbs (Table 2 below).

If we look at POS, students score slightly worse than UK students on nouns, verbs and adjectives, see table 2. However on all other POS they do immensely better than UK students hence refuting the conclusion in Alderson et al (1997) that students can only be assumed to know nouns and verbs. However, to be fair the UK study was done primarily pre-university training in grammar and our students had already had some instruction regarding grammar and metalinguistic terminology, something Alderson et al (2013: 330) have shown can make a difference.

Student behaviour shows that students find encyclopedia resources very useful. The number of times Wikipedia has been opened tops all other resources: 1728 times, followed by Saldo morphological lexicon (1385) and Wiktionary lookups (1020). The visualized syntactic tree has proven to be rather confusing and only in 531 cases did students consult it.

	Noun	Verb	Adj.	Def art	Indef art	Det.	Conj.	Subj.	Prep.	Adv.	Rel. Pron.	Participle	Past part.
Lark log	86	88	71	N/A	N/A	72	78	68	72	64	86	77	N/A
Alderson & al (2013)	93	91	76	30	26	N/A	68	N/A	34	43	27	N/A	27
Corona & al (2010)	99	98	95	80	78	N/A	68	N/A	90	82	88	N/A	72

Table 2. Percentages of correct answers for POS in our diagnostic tests, the UK (2013) and the Spanish (2010) study

5. Future work

The survey comments show that we need to improve our sentence selection algorithm which at the moment selects sentences that are ranked as "good dictionary examples" (GDEX) by our readability module (Pilan et al., 2014). A subset of selected sentences are context-bound, and it may be problematic to make a meaningful analysis of the sentence and its components.

The study has also indicated a need to correlate sentence complexity with the error rate. The assumption is that the beginner students need to be exposed to simpler training contexts, whereby the sentence complexity should grow with students' experience. For that, we need to train our sentence selection algorithm for additional features. We plan to do so using COCTAILL, a newly compiled corpus of coursebooks aimed at learners of Swedish as a second language (Volodina et al., submitted). Deriving an algorithm for identification of sentences understandable at lower levels of language proficiency will supply us with simple sentences (in terms of grammatical complexity).

Another problem outlined by students refers to terminology. One striking example is *determiner*, a part of speech that is widely accepted in English, but for which there is no support in official grammars for Swedish. Still Swedish students score better on determiner (72.49% correct) than UK students do on the definite and indefinite article (29.89% and 26.40%). The automatic corpora annotation flow exploits this POS at the moment, however, this will be changed in the near future as a result of the project Koala and this category will be replaced in accordance with SAG, Swedish Academic Grammar (Teleman et al 1999).

The surprisingly good results for determiners, a category not known to students from their course

literature, have led us to suspect that the selection of distractors for each exercise item might not be an optimal one. In the future, we will consider including all the possible categories, e.g. all 11 parts of speech, as distractors for diagnostic purposes, whereas for the other modes of training (self-study, test and timed test) we will keep the presently exploited list of a maximum of 5 choices.

Extensive use of Wikipedia suggests that students experience a need to consult some encyclopedia for explanation of linguistic notions. Since Wikipedia cannot be expected to contain all the necessary linguistic information, we will explore the possibilities to use other (more reliable) sources of linguistic knowledge, as well as linking additional lexical resources containing explanations of running words in a sentence, and potentially including bilingual resources.

A more controlled experiment where we test students before their first grammar course and after their course is planned for the near future. This experiment is planned to involve a diagnostic test similar to that used in the British and Spanish studies (Cf Bloor 1986; Alderson et al 1997; Alderson & Hudson 2013; Corona & Mur-Dueñas 2010) and will probably be conducted in two-three comparative groups of which one - two will not use Lärka during their course.

On the user interface side, it would be useful to facilitate feedback for teachers by allowing them to request log summaries for a particular session, to get an easy overview of their class and use this to see what they need to revise. Alternatively, the possibility to email a summary of one's results would be a useful feature to explore. Logs could also be studied in detail to look at the improvement over time, or the categories that get confused.

References

Alderson, J. C., C. Clapham & D. Steel. 1997. Metalinguistic knowledge, language aptitude and language proficiency. *Language Teaching Research*. 1.2. pp. 93-121.

Alderson, J. C. & R. Hudson. 2013. The metalinguistic knowledge of undergraduate students of English Language or linguistics. *Language Awareness*. 22.4, pp. 320-337.

Bloor, Thomas. 1986. What do language students know about grammar? *British Journal of Language Teaching*. 24.3, pp. 157-162.

Borin, L., M. Forsberg, L-J. Olsson & J. Uppström. 2012a. The open lexical infrastructure of Språkbanken. *Proceedings of LREC 2012*. Istanbul: ELRA.

Borin, L., M. Forsberg & J. Roxendal. 2012b. Korp – the corpus infrastructure of Språkbanken. *Proceedings of LREC 2012*. Istanbul: ELRA.

Corona, I. & P. Mur-Dueñas. 2010. Getting to grips with grammar: Native vs non-native first-year undergraduates. A contrastive application of the KAL test. Paper presented at the 28th International Conference of the Spanish Society for Applied Linguistics, AESLA 2010, 'Analysing Data, Describing Variation', Vigo, April 15-17.

Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge, UK: Cambridge University Press.

O'Keeffe, A., McCarthy, M. & Carter, R. (2007). *From Corpus to Classroom. Language Use and Language Teaching*. Cambridge: Cambridge University Press.

Pilán, I., E. Volodina & R. Johansson. 2014. Rule-based and machine learning approaches for second language sentence-level readability. *Proceeding of the ACL 2014 9th Workshop on Innovative Use of NLP for Building Educational Applications*.

Teleman, U., S. Hellberg, E. Andersson & L. Holm. 1999. *Svenska Akademiens Grammatik*. Stockholm: Svenska Akademien & Norstedts.

Volodina, E., Pilán, I., Eide, S.R. & Heidarsson, H. (submitted). You get what you annotate: a pedagogically annotated corpus of coursebooks for Swedish as a Second Language.

Volodina, E., I. Pilán, L. Borin & T. Lindström Tiedemann. 2014. A flexible language learning platform based on language resources and web services. *Proceedings of LREC 2014*. Reykjavik, Iceland: ELRA.