SweLLex: second language learners' productive vocabulary

Elena Volodina[1], **Ildikó Pilán**[1], **Lorena Llozhi**[1], **Baptiste Degryse**[2], **Thomas François**[2],[3]
[1] University of Gothenburg, Sweden
[2] Université catholique de Louvain, Belgium
[3] FNRS Post-doctoral Researcher
elena.volodina@svenska.gu.se

Abstract

This paper presents a new lexical resource for learners of Swedish as a second language, SweLLex, and a know-how behind its creation. We concentrate on L2 learners' *productive* vocabulary, i.e. words that they are actively able to produce, rather than the lexica they comprehend (*receptive* vocabulary). The proposed list covers productive vocabulary used by L2 learners in their essays. Each lexical item on the list is connected to its frequency distribution over the six levels of proficiency defined by the Common European Framework of Reference (CEFR) (Council of Europe, 2001). To make this list a more reliable resource, we experiment with normalizing L2 word-level errors by replacing them with their correct equivalents. SweLLex has been tested in a prototype system for automatic CEFR level classification of essays as well as in a visualization tool aimed at exploring L2 vocabulary contrasting receptive and productive vocabulary usage at different levels of language proficiency.

1 Introduction

The results of the Survey of Adult Competencies (PIAAC, 2013), where literacy as a skill has been assessed among the adult population (16-65 years) has shown that on average Sweden scored among the top 5 countries out of the 23 OECD participants. However, the national Swedish report claims that the difference between the average literacy levels of native (L1) born citizens compared to citizens with an immigrant (L2) background is the largest observed among all participating countries (OECD, 2013, p.6). The low literacy population in Sweden has three times higher risk of being unemployed or reporting poor health. The results of the survey point to an acute need to support immigrants and other low-literacy groups in building stronger language skills as a way of getting jobs and improving their lifestyle (SCB, 2013, p.8).

A way of addressing the needs of immigrants as well as L2 teachers would be to provide an extensive amount of self-study materials for practice. This could be achieved through the development of specific algorithms, but they generally heavily rely on linguistic resources, such as descriptions of vocabulary and grammar scopes per each stage of language development, or (to avoid level-labeling) at least a predefined sequenced presentation of vocabulary and grammar so that automatic generation of learning materials would follow some order of increasing complexity. To do this, as a first step, we need to examine *reading materials* used in L2 courses versus *essays* written during such courses, to study what constitutes L2 learners' lexical and grammatical competence at various levels of proficiency.

Our study has addressed one sub-problem among those outlined above, namely, a descriptive list of productive vocabulary based on a corpus of L2 learner essays. We have combined corpus linguistics methods, computational linguistics methods and empiric analysis to secure a resource that could be used both for L2 research as well as for teaching and assessment purposes. As a preliminary step, we have tested two methods of normalization of L2 word-level errors to see how that would improve the

Elena Volodina, Ildikó Pilán, Lorena Llozhi, Baptiste Degryse and Thomas François 2016. SweLLex: Second language learners' productive vocabulary. *Proceedings of the joint workshop on NLP for Computer Assisted Language Learning and NLP for Language Acquisition at SLTC 2016*. Linköping Electronic Conference Proceedings 130: 76–84.

quality of automatic annotation and the quality of the list itself. The resource is not perfect; a number of iterations for its improvement would be needed, complemented with pedagogical experiments. However, this is a pilot study that helps us analyze and improve the methodology, find out its weaknesses and strengths and decide on the paths to take ahead.

The result of the study is a browsable inventory of Swedish L2 productive vocabulary with frequency distributions across CEFR levels. It is possible to browse the resource in parallel with its sister resource for L2 receptive vocabulary, SVALex (François et al., 2016).

Below, we provide a short survey of lexical resources for second language learners (Section 2), present our experiments on normalization (Section 3.2), describe the resulting list (Section 4) and conclude by outlining future perspectives (Section 5).

2 Background

In developing L2 courses as well as designing L2 tests, considerations about which vocabulary to teach or assess are critical. According to the findings within L2 research, to cope with reading comprehension tasks, a learner should understand 95-98% of the text vocabulary (Laufer and Ravenhorst-Kalovski, 2010). But which vocabulary should be taught, and in which order?

Attempts to outline lexical items to concentrate on in L2 context date back to Thorndike (1921). Several approaches have been used since then to identify relevant vocabulary for L2 learners, such as relying on expert intuitions (Allén, 2002), combining statistical insights with expert judgments (Hult et al., 2010), and lately estimating frequencies from corpus-based sources where several variations can be found: domain-specific lists (Coxhead, 2000), general purpose vocabulary (West, 1953), word family frequencies (Coxhead, 2000), and lately sense-based lists (Capel, 2010; Capel, 2012).

Most of the lists above, however, do not reflect the order in which vocabulary should be taught or tested for L2 learners, or at which level. An attempt to cover that need was made in the English Vocabulary Profile (Capel, 2010; Capel, 2012). For Swedish, an effort to list receptive vocabulary useful for L2 learners was made in the European Kelly project (Kilgarriff et al., 2014) and recently in the SVALex list (François et al., 2016). While Kelly list is based on web-texts whose primary target readers are first language speakers; SVALex is based on the reading comprehension texts used in coursebooks aimed at L2 learners. Both lists, thus, cover receptive vocabulary, i.e. vocabulary that L2 learners can understand when exposed to it while reading or listening. To complement the receptive repertoire with the productive one, we have explored L2 learner essays.

3 Method

3.1 Source corpus

It is natural that any vocabulary list would reflect the corpus it is based on. It is thus important to know what constitutes the source corpus, in our case the SweLL corpus. SweLL (Volodina et al., 2016b) is a corpus consisting of essays written by learners of Swedish as a second language, aged 16 or older. It has been collected at three educational establishments and covers the six CEFR levels: A1 (beginner), A2, B1, B2, C1 and C2 (near-native proficiency). However, C2 is heavily underrepresented. Table 1 summarizes the distribution of essays (and sentences and tokens) across the 6 CEFR levels.

Level	Nr. essays	Nr. sent	Nr. tokens
A1	16	247	2084
A2	83	1727	18349
B1	75	2005	29814
B2	74	1939	32691
C1	89	3409	60455
C2	2	46	694
Total	**339**	**9 373**	**144 087**

Table 1: Number of essays, sentences, and tokens per CEFR level in the SweLL corpus.

The SweLL corpus contains a number of variables associated with the essays, including:

- *learner variables*: age at the moment of writing, gender, mother tongue (L1), education level, duration of the residence stay in Sweden;
- *essay-related information*: assigned CEFR level, setting of writing (exam/classroom/home), access to extra

materials (e.g. lexicons, statistics), academic term and date when the essays have been written, essay title, and depending upon the subcorpus - topics (SpIn, TISUS, SW1203), genre (TISUS, SW1203), and grade (TISUS).

Another important characteristics of a corpus that influences a word list derived from it is text topics. In SweLL, the major part of the essays have been annotated for topics, with often several topics assigned to the same essay. The topics are presented in Table 2 in decreasing frequency order.

Topic	Nr essays
health and body care	117
personal identification	97
daily life	60
relations with other people	31
free time, entertainment	19
places	16
arts	15
travel	15
education	9
family and relatives	7
economy	4

Table 2: Number of essays per topic

Since the corpus is rather small, there is a bias towards the dominating topics, something that we intend to overcome in future updates of the list.

3.2 L2 text normalization

Standard corpus annotation follows a number of steps, including tokenization, PoS-tagging, lemmatization and syntactic parsing. A project dealing with learner language requires handling of texts exhibiting a great amount of deviation from standard Swedish. While texts with normative Swedish can be relatively accurately annotated with existing automatic methods, annotating learner language with the same tools is error-prone due to various (and often overlapping) orthographic, morphological, syntactic and other types of errors, e.g.:

- segmentation problems: “jag har två kompisar som hete S och P de är från Afghanistan också jag älskar de för att när jag behöver hjälp de hjälpar gärna mig och jag också hjälpa de.”
- misspelling variations: “sommern”, “kultor”
- unexpected morphological forms and agreement errors: “Min drömar”
- word order errors: “Jag bara studera 4 ämne i skolan och på fritiden träna jag på gym”

To tackle that problem, an extra step is often added to the annotation process before a standard annotation pipeline is applied, where deviating forms are rewriten to fit into the accepted norms of the language. That step is often referred to as *normalization* (Megyesi et al., 2016; Wisniewski et al., 2013; Dickinson and Ragheb, 2013). Previous error-normalization approaches include, among others, finite state transducers (Antonsen, 2012) and a number of systems, mostly hybrid, created within the CoNLL Shared Task on grammatical error correction for L2 English (Ng et al., 2014).

A more practical reason for our normalization experiments is based on the fact that after the initial collection of raw frequencies for SweLLex, we noticed that there were 4,308 unique tokens which were not assigned a lemma during the linguistic annotation. Figure 1 shows the distribution of non-lemmatized items across all levels of proficiency.

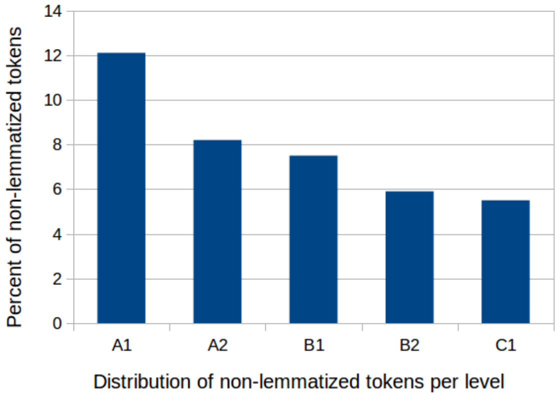

Figure 1: Percent of non-lemmatized items per level, %

We examined a selection of the non-lemmatized words (about 1000 tokens) and split those into five categories. Table 3 shows some examples of the five categories, including correct spelling and English translation where applicable.

Category	Example (correct)	Eng
Misspelling	fotbol (fotboll)	football
Compound	arbetsstress	job stress
Hyphenation	för-söka (försöka)	to attempt
Foreign word	opportunity	
Acronym	fö (för övrigt)	moreover

Table 3: Examples of word entries that failed to match against SALDO morphology lexicon, by category

Level	Correct/total
A1	7/20
A2	13/20
B1	13/20
B2	15/20
C1	16/20

Table 4: Number of correctly returned suggestions per level

To reduce the number of non-lemmatized items, especially in cases of misspellings and hyphenation, we experimented with two normalization approaches at the word level: pure Levenshtein distance, and LanguageTool's output combined with candidate ranking strategies. Our hypothesis has been that normalization should take care of the word-level anomalies of learner language replacing them with a standard variant, so that the automatic annotation in the next step would be more accurate.

Approach 1: Levenshtein distance

As the first strategy for normalization we experimented with pure Levenshtein distance (LD) as implemented in NLTK (Bird, 2006)[1]. LD is a measure for the distance between two strings. In our case, this was the difference between the (possibly) misspelled word and the (probable) target word. Output suggestions were based on SALDO-morphology lexicon (Borin et al., 2013), a full-form lexicon where all inflected forms are listed alongside their base forms and parts of speech. As such, in the cases where the word form was not present in SALDO, we chose the word form in SALDO morphology to which the original word form in our source had the shortest LD, selecting the first suggestion with the shortest edit distance. Suggestions had to start with the same letter, based on the assumption that a misspelled word is likely to start with the same letter as its corresponding correct lemma (Rimrott and Heift, 2005).

Analysis of 20 randomly selected corrections per level has shown that apart from level A1, LD performed quite well at the other levels (see Table 4).

Zooming into the observed cases, we could see that our LD-based algorithm returns the right lemma in those cases where the edit distance equals 1. Those cases include:

(1) substitution of one misspelled letter, e.g.: ursprang*[2] → ursprung (*origin*);

(2) deletion of an extra letter, e.g.: sekriva* → skriva (*to write*), naman* → namn (*name*);

(3) insertion of one missing letter, i.e. sammanfata* → sammanfatta (*summarize*).

However, when multiple misspellings occur in a word, the performance of LD is rather poor. Also, whenever a word is very short there will likely be many lemmas that have a Levenshtein distance of 1 from the token, and the returned suggestion is often incorrect.

In cases where the first letter is misspelled (e.g. andå* → ändå, *anyway*) our LD-based algorithm fails to return a correct lemma.

Our analysis shows that Levenshtein distance is applicable to normalization of writing at more advanced levels of language proficiency, whereas at the earlier stages it should be complemented by a more complex approach, for example candidate ranking based on word co-occurrence measures as described below.

Approach 2: LanguageTool & candidate ranking

The second type of error normalization was based on LanguageTool[3] (LT) (Naber, 2003), an open-source rule-based proof-reading program available for multiple languages. This tool detects not only spelling, but also some grammatical errors (e.g. inconsistent gender use in inflected forms).

As a first step, we identified errors and a list of one or more correction suggestions, as well as the *context*, i.e. the surrounding tokens for the er-

[1] http://www.nltk.org/

[2] An asterisk (*) is added to (potentially erroneous) word forms not found in the SALDO-morphology lexicon.

[3] www.languagetool.org

ror within the same sentence. When more than one correction candidate was available, as an additional step, we made a selection based on *Lexicographers' Mutual Information* (LMI) scores (Kilgarriff et al., 2004). LMI measures the probability of two words co-occurring together in a corpus and it offers the advantage of balancing out the preference of the Mutual Information score for low-frequency words (Bordag, 2008).

The choice of a correction candidate was based on assuming a positive correlation between a correction candidate co-occurring with a context word and that word being the correct version of the learner's intended word. We checked LMI scores for each LT correction candidate and the lemma of each available noun, verb and adjective in the context based on a pre-compiled list of LMI scores. We have created this list using a Korp API (Borin et al., 2012) and a variety of modern Swedish corpora totaling to more than 209 million tokens. Only scores for noun-verb and noun-adjective combinations have been included with a threshold of LMI ≥ 50. When available, we select the correction candidate maximizing the sum of all LMI scores for the context words. In the absence of LMI scores for the pairs of correction candidates and context words, the most frequent word form in Swedish Wikipedia texts is chosen as a fallback. Once correction candidates are ranked, each erroneous token identified by LanguageTool is replaced in the essays by the top ranked correction candidate.

In absence of L2 Swedish learner data with error annotations, we performed a small manual evaluation. We checked 114 randomly chosen corrections obtained with the approach described above, out of which 84 were correct, corresponding to 73.68% accuracy. Table 5 shows the amount of corrected tokens per CEFR level. Some of the corrections concerned stylistic features such as inserting a space after punctuation, which was especially common at higher CEFR levels, thus a higher error percentage at B2 and C1 levels is not necessarily an indication of less grammatical texts.

The final variant of SweLLex was derived from a version of the essays normalized with the second approach.

	# tokens	**% tokens**
A1	204	9.7
A2	1118	6.0
B1	1650	5.5
B2	3526	10.8
C1	7511	12.4

Table 5: Amount of corrected tokens per CEFR level

3.3 Frequency estimation

Each entry in the final list is a base form (lemma) and its part of speech. An entry can also be a multi-word expression (MWE) which is identified during the annotation process by matching potential MWEs to entries in SALDO. Further, each entry is associated with its dispersed frequency in the corpus as a total, frequency at each level of proficiency, as well as for each individual writer ID. Besides, we have connected each writer ID to their mother tongues and have thus a possibility to analyze vocabulary per level and L1.

To estimate frequencies, we used the same formula as for SVALex list (François et al., 2016) to ensure comparability between the two resources aimed at the same language learner group. The frequency formula takes into consideration dispersion of vocabulary items across all learners in the corpus (learner IDs), i.e. it compensates for any influences introduced due to overuse of specific vocabulary by an individual learner (Francis and Kucera, 1982). Dispersion has become a standard approach to frequency estimations, e.g. in projects such as English Vocabulary Profile and FLELex (Capel, 2010; Capel, 2012; François et al., 2016).

4 Description of the resource

The resulting list contains in total 6,965 items. Despite the fact that SweLLex has been generated from a normalized SweLL corpus (Volodina et al., 2016b), about 1490 items could not be lemmatized. In 526 cases it is due to compounds which are not present in SALDO, the rest are the items that haven't been identified by LanguageTools. The statistics below is provided for the rest of the list, i.e. excluding the non-lemmatized items. We compare SweLLex statistics with two other resources, SVALex and English Vocabulary Profile (EVP), to see:

Lev	#items	#new	#MWE	#hapax	Doc.hapax examples	#SVALex	#EVP
A1	398	398	15	0	-	1,157	601
A2	1,327	1,038	82	12	*i kväll* ”tonight”	2,432	925
B1	2,380	1,542	206	36	*fylla år* ”have birthday”	4,332	1,429
B2	2,396	959	264	58	*fatta beslut* ”make a decision”	4,553	1,711
C1	3,566	1,545	430	152	*sätta fingret* ”put a finger on sth”	3,160	N/A
C2	145	7	12	1	*i bakhuvudet* ”in mind”	N/A	N/A

Table 6: Distribution of SweLLex entries per CEFR level, including the nr. of items, new items, multi-words expressions, and nr. of document hapaxes per level. We also provide the number of new items for SVALex and EVP (Capel (2014)) for comparison

.

(1) trends between productive lists across two languages, Swedish & English (SweLLex versus EVP)

(2) and productive-receptive relation within the same language (SweLLex versus SVALex).

Table 6 shows that the number of new items per level follows the same pattern as in the English Vocabulary Profile with (almost) comparable numbers at all levels except for B2, where the number of new items in SweLLex is twice as little as in the EVP resource. A hypothetical reason for that could be that we have essays on a very limited number of topics at B2 level (and levels above), which constraints learners from using more varied vocabulary. Since numbers at C1 and C2 levels are not available for EVP, we cannot trace this trend at these levels. However, it would be interesting to see whether the tendency will change once we have collected essays on more varied topics from these levels.

The trend in the receptive resource shows that the number of items increases almost twofold between A1 and A2 in both lists. However, between A2 and B1 students are exposed to many more items than they are able to use actively in writing, at least if we rely on the numbers in SweLLex and SVALex. At B2 we have a low point trend in SweLLex even in comparison to receptive vocabulary, which indirectly supports our previous hypothesis that essays at B2 level have too few topics, influencing (and limiting) the type of vocabulary that students use in their essays. At C2 level we have only 2 essays, which makes the numbers non-representative for analysis.

We can also see that the number of MWEs is growing steadily between levels and can be viewed as one of the most stable (and probably reliable) characteristics of increasing lexical complexity between levels, despite essay topic variation per level.

A document hapax means that the item has been used in one document only in the whole corpus. Document hapaxes are potential candidates for being excluded from central vocabulary at that level. We can, however, see from hapax examples that they can be very good items to keep on the list, covering such words as *tonight, make a decision*, etc. Decisions on how to treat document hapaxes should follow a more pedagogical approach.

A look at the ten most frequent words per level shows that the most frequent word at A1 and A2 levels is the pronoun *jag* (Eng: “I”), which denotes that during the earlier levels, students gradually learn how to talk about their daily lives and people they associate with. This is also apparent from the most used nouns: *skola* (Eng: “school”) and *kompis* (Eng: “friend”). At level A2, we see that more pronouns, *han* and *vi* (Eng: “he” and “we”), are included among the top ten words. This indicates that learners are starting to refer to other people more frequently.

At the intermediate levels (B1 and B2), *jag* is no longer the top frequent word, but rather *vara* (Eng: “[to] be”). From this, we can assume that language at these levels becomes more about describing things and probably moves beyond the personal life prevalent at the A levels. Moreover, the verb *ha* (Eng: “have”) is introduced among the most frequent words at the B levels. In Swedish, *ha* is also used as an auxiliary verb in order to form perfect tenses. As such, the high frequencies of this word may be because the students are more acquainted with additional tenses.

An interesting addition to note at the C1

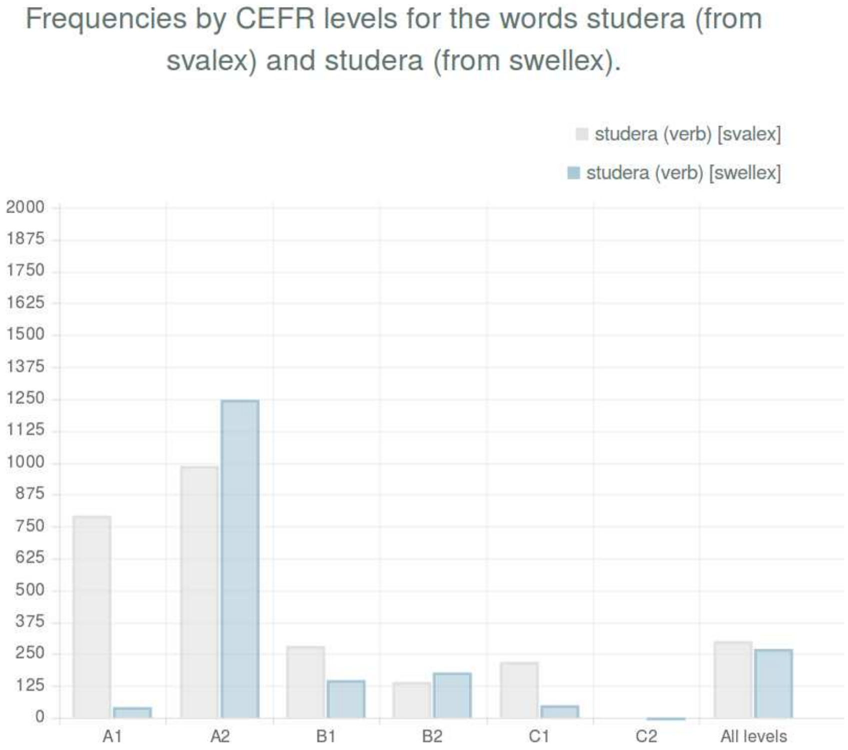

Figure 2: Distribution of the verb *studera*, Eng. ”to study”, in receptive and productive resources (screen capture from the website)

level is the presence of the lemma *som* (Eng: “who/which/as/that”). This is a clear indication that students have reached a relatively proficient language level, being able to frequently construct subordinate clauses. These are only a few examples of the most frequent words at each level, but they already show the students’ language progress. Our list gives a potential to explore further lexical patterns related to vocabulary progress.

Availability of resources of the two kinds - covering receptive and productive vocabulary - makes it possible to contrast receptive and productive distributions. Initially, we matched the two resources to look into the overlaps and possible SweLLex items that are not present SVALex. This yielded the results shown in Table 7.

Resource	#items	#overlaps	#missing
SVALex	15,861	3,591	3,226
SweLLex	6,965	3,591	12,060

Table 7: Comparison between SVALex and SweLLex lists

As we can see, SVALex is an extensive vocabulary list, almost twice the size of SweLLex. Consequently, it is not surprising that 12,060 entries present in SVALex are missing from SweLLex. On the other hand, there are 3,226 entries in SweLLex which are not present in SVALex. Analysis of those items is left for future work, but from the initial inspection, those consist mostly of the non-lemmatized items (e.g. due to learner errors) and compounds.

A more interesting insight can be gained by inspecting distribution profiles of different items. Hypothetically, learners are first exposed to an item through reading, and afterwards start using it productively in writing at a later level. Figure 2 supports this trend. However, words can be expected to show different trends, something that can be explored in the browsable interface for the two resources[4].

[4] http://cental.uclouvain.be/svalex/

5 Conclusions

The work presented is only the first step towards a comprehensive description of the productive vocabulary scope used by L2 Swedish learners at different proficiency levels. We have looked into the lexical scope learners demonstrate at various levels productively; two normalization methods at the word-level in the context of L2 writing; initial comparison between receptive and productive vocabulary. The method of creating SweLLex needs to be complemented by deeper empiric analysis and pedagogical evaluation; extended by more advanced normalization procedures.

There are multiple directions for future work, including mapping SweLLex distributions to single levels (ongoing work); identifying core versus peripheral vocabulary (must-know vs good-to-know lexical competence); merging SVALex, SweLLex and Kelly-list into a common resource; incorporating SweLLex into real-life applications and tools aimed at L2 learners of Swedish. Another future research direction consists in finding a way to automatically normalize errors stretching over two or more words, as well as at the syntactic level, something that is planned to be addressed within L2 infrastructure efforts (Volodina et al., 2016a).

References

Sture Allén. 2002. *Våra viktiga ord.* Liber, Sweden.

Lene Antonsen. 2012. Improving feedback on L2 misspellings-an FST approach. In *Proceedings of the SLTC 2012 workshop on NLP for CALL; Lund; 25th October; 2012*, number 080, pages 1–10. Linköping University Electronic Press.

Steven Bird. 2006. NLTK: the natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions*, pages 69–72. Association for Computational Linguistics.

Stefan Bordag. 2008. A comparison of co-occurrence and similarity measures as simulations of context. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 52–63. Springer.

Lars Borin, Markus Forsberg, and Johan Roxendal. 2012. Korp - the corpus infrastructure of Språkbanken. In *LREC*, pages 474–478.

Lars Borin, Markus Forsberg, and Lennart Lönngren. 2013. SALDO: a touch of yin to WordNet's yang. *Language Resources and Evaluation*, 47(4):1191–1211.

A. Capel. 2010. A1–B2 vocabulary: insights and issues arising from the English Profile Wordlists project. *English Profile Journal*, 1(1):1–11.

A. Capel. 2012. Completing the English Vocabulary Profile: C1 and C2 vocabulary. *English Profile Journal*, 3:1–14.

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Press Syndicate of the University of Cambridge.

Averil Coxhead. 2000. A new academic word list. *TESOL quarterly*, 34(2):213–238.

M. Dickinson and M. Ragheb. 2013. *Annotation for Learner English Guidelines, v. 0.1. Technical report.* Indiana University, Bloomington.

W. Francis and H. Kucera. 1982. *Frequency analysis of English usage.* Houghton Mifflin Company, Boston, MA.

Thomas François, Elena Volodina, Ildikó Pilán, and Anaïs Tack. 2016. SVALex: a CEFR-graded Lexical Resource for Swedish Foreign and Second Language Learners. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, may.

Ann-Kristin Hult, Sven-Göran Malmgren, and Emma Sköldberg. 2010. Lexin - a report from a recycling lexicographic project in the North. In *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010).*

Adam Kilgarriff, Pavel Rychly, Pavel Smrz, and David Tugwell. 2004. Itri-04-08 the Sketch Engine. *Information Technology*, 105:116.

A. Kilgarriff, F. Charalabopoulou, M. Gavrilidou, J. B. Johannessen, S. Khalil, S. J. Kokkinakis, R. Lew, S. Sharoff, R. Vadlapudi, and E. Volodina. 2014. Corpus-based vocabulary lists for language learners for nine languages. *Language resources and evaluation*, 48(1):121–163.

B. Laufer and G.C. Ravenhorst-Kalovski. 2010. Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a foreign language*, 22(1):15–30.

Beáta Megyesi, Jesper Näsman, and Anne Palmér. 2016. The Uppsala Corpus of Student Writings: Corpus Creation, Annotation, and Analysis. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016).*

Daniel Naber. 2003. A rule-based style and grammar checker. Master's thesis, Bielefeld University, Bielefeld, Germany.

Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Hendy Susanto, and Christopher Bryant, editors. 2014. *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*. Association for Computational Linguistics, Baltimore, Maryland.

OECD. 2013. *OECD Skills Outlook 2013. First Results from the Survey of Adult Skills.*

PIAAC. 2013. *Survey of Adult Skills (PIAAC).*

Anne Rimrott and Trude Heift. 2005. Language learners and generic spell checkers in CALL. *CALICO journal*, pages 17–48.

SCB. 2013. *Tema utbildning, rapport 2013:2, Den internationella undersökningen av vuxnas färdigheter.* Statistiska centralbyrån.

E.L. Thorndike. 1921. *The teacher's word book*. Teachers College, Columbia University, New York.

Elena Volodina, Beata Megyesi, Mats Wirén, Lena Granstedt, Julia Prentice, Monica Reichenberg, and Gunlög Sundberg. 2016a. A Friend in Need? Research agenda for electronic Second Language infrastructure. In *Proceedings of SLTC 2016, Umeå, Sweden*.

Elena Volodina, Ildikó Pilán, Ingegerd Enström, Lorena Llozhi, Peter Lundkvist, Gunlög Sundberg, and Monica Sandell. 2016b. SweLL on the rise: Swedish Learner Language corpus for European Reference Level studies. *LREC 2016, Slovenia*.

Ma West. 1953. *A General Service List of English Words*. London: Longman, Green and Co.

Katrin Wisniewski, Karin Schöne, Lionel Nicolas, Chiara Vettori, Adriane Boyd, Detmar Meurers, Andrea Abel, and Jirka Hana. 2013. MERLIN: An online trilingual learner corpus empirically grounding the European reference levels in authentic learner data. In *ICT for Language Learning 2013, Conference Proceedings, Florence, Italy. Libreriauniversitaria. it Edizioni*.