

# Profiles for Swedish as a Second Language: Lexis, Grammar, Morphology

Elena Volodina<sup>1</sup>, David Alfter<sup>1</sup> and Therese Lindström Tiedemann<sup>2</sup>

<sup>1</sup>University of Gothenburg, Sweden

<sup>2</sup>University of Helsinki, Finland

## Abstract

This article gives a short introduction to the Swedish Second Language Profile, a tool that visualizes language in Swedish learner corpora from different angles, such as vocabulary, grammar and morphology. The tool is aimed at research on Second Language Acquisition, development of NLP models, teaching of Swedish as a second language, automatic approaches for second language teaching and learning, and at a number of other fields.

## Keywords

Second Language Profile for Swedish (SweL2P), Lexical profile, Grammatical profile, Morphological profile

## 1. Introduction

Learner corpus researchers, NLP researchers, as well as Digital Humanities and Social Sciences in general rely on access to various data sets for empirical analysis, statistical insights and/or for model building. However, interpretation of data is a non-trivial task and there is a need for data visualization tools [1, 2, 3, 4]. One such attempt is the Swedish Second Language (L2) profile<sup>1</sup> - a project setting up the first digital tool allowing users to explore written Swedish learner language from a linguistic point of view.

The aim of the project was to describe the learning paths of Swedish as a second language with a focus on vocabulary and grammar. To do this, we analysed two text collections - course books and learner essays - for various patterns and their statistics at different levels of proficiency, and organized them into so-called profiles (section 2).

Access to the profiles facilitates studies into, among others, which vocabulary or grammar is central for a certain level of proficiency, which linguistic features can be discriminative at different levels, in which order the various linguistic aspects are introduced into reading materials and into writing, and many others. The practical application of the profiles can take many different forms. For Intelligent Computer-Assisted Language Learning they facilitate, for example, generation of learning materials (exercises, test items, etc.) of appropriate linguistic complexity for given

---

*Huminfra Conference 2024, Gothenburg, 10–11 January 2024.*

✉ elena.volodina@svenska.gu.se (E. Volodina); david.alfter@gu.se (D. Alfter);

therese.lindstromtiedemann@helsinki.fi (T. Lindström Tiedemann)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

<sup>1</sup>The tool is a by-product of the research project *Development of lexical and grammatical competences in immigrant Swedish*, <https://spraakbanken.gu.se/en/projects/l2profiles>

Multi Word Expressions [🔗](#)

---

Search word  Word Class  Saldo Word Class  Type 1: Syntactic cont...  Type 2: Lexical catego...  Type 3: Verbal subcat...

A1 A2 B1 B2 C1 C2

Tables Graphs Statistics Download Clear all

Extend columns in the table  Show only first occurrence [Tables - description](#) [Filters - description](#)

Figure 1: Filters for Multiword Expressions (Lexical profile)

| CEFR level <a href="#">↑↓</a> | Word <a href="#">↑↓</a>   | Lemgram <a href="#">↑↓</a> | Sense <a href="#">↑↓</a> | Word Class <a href="#">↑↓</a> | Saldo Word Class <a href="#">↑↓</a> | Receptive <a href="#">↑↓</a> <a href="#">🔗</a> | Productive <a href="#">↑↓</a> <a href="#">🔗</a> |
|-------------------------------|---------------------------|----------------------------|--------------------------|-------------------------------|-------------------------------------|--|---|
| A1                            | <a href="#">jeans</a>     | jeans..nn.1                | jeans..1                 | Noun (NN)                     | Noun (nn)                           | 0.77 (2)                                       | 0.00 (0)  |
| A1                            | <a href="#">kläder</a>    | kläder..nn.1               | kläder..1                | Noun (NN)                     | Noun (nn)                           | 3.09 (8)                                       | 7.32 (3)  |
| A1                            | <a href="#">shorts</a>    | shorts..nn.1               | shorts..1                | Noun (NN)                     | Noun (nn)                           | 1.16 (3)                                       | 0.00 (0)  |
| A2                            | <a href="#">makaroner</a> | makaroner..nn.1            | makaroner..1             | Noun (NN)                     | Noun (nn)                           | 0.44 (3)                                       | 0.00 (0)  |

Figure 2: Table view for Sen\*Lex (Lexical profile), filtered for 'Always plural' nouns

levels; or to classify learner texts according to proficiency levels reached using information of the vocabulary or grammar scope per level as input features. For Second Language Acquisition (SLA), it is possible to study the relation between vocabulary acquisition and previous morphological (derivational) knowledge; the order of introduction of various grammatical structures at different levels, and many others.

As such, the users can be researchers, developers, language assessors, course book writers, teachers or learners. The fact that any filtering allows download of a filtered subset of structures makes us believe that such lists will be used on their own as spin-off resources.

## 2. SweL2P - overview

The Swedish Second Language Profile (SweL2P) features the following profiles:

- A lexical profile, organized into subprofiles by words, multiword expressions, adjectival declensions and adjectival & adverbial structure;
- A grammatical profile, including noun patterns and verb patterns;
- A morphological profile, organized into word families and morpheme families.

### 2.1. Source data

SweL2P is empirical in nature since it is based on data from two corpora: **COCTAILL** – a corpus consisting of coursebooks used for teaching Swedish to L2 students [5]; and **SweLL-pilot** – a corpus of learner essays written by L2 learners of Swedish [6]. Texts in both corpora have

Lexical profile | **Grammatical profile** | Morphological profile

Verb patterns

Pattern (description) | Pattern (structure) | Tense | Mood | Voice | Form

A1 | A2 | B1 | B2 | C1 | C2

Extend columns in the table  
 Show only first occurrence  
 Tables - descr

Tense  
 Mood  
 Voice  
 Form

Total rows : 18

| CEFR level | Pattern (description)         | Pattern (structure) | Tense            | Mood       | Voice            | Receptive | Productive |
|------------|-------------------------------|---------------------|------------------|------------|------------------|-----------|------------|
| A2         | Past tense (preterite) - .... | Simple              | Past (Preterite) | Indicative | Passive          | 0.03 (4)  | 0.04 (1)   |
| A2         | Perfect tense - s-passive.... | Complex             | Perfect          | Indicative | Passive/Deponent | 0.01 (2)  | 0.00 (0)   |
| A2         | Perfect tense - s-passive.... | Complex             | Perfect          | Indicative | Passive          | 0.01 (2)  | 0.00 (0)   |
| B1         | Perfect tense - deponent .... | Complex             | Perfect          | Indicative | Deponent         | 0.04 (7)  | 0.00 (0)   |

**Figure 3:** Table view for Verb Patterns (Grammatical profile)

been graded with CEFR [7] levels by experts, starting from A1 (beginner) to C1 (advanced), the C2 level being largely absent. COCTAILL has been used to get an approximation of the vocabulary and grammar L2 learners meet when reading, and therefore what they are expected to understand **receptively**. SweLL-pilot has been used to get an approximation of the vocabulary and grammar L2 learners are able to produce actively when writing, and therefore represents learners' **productive** abilities.

The two corpora have been used to create a sense-disambiguated word list, Sen\*Lex [8, 9, 10], as the main input for the Lexical profile. Sen\*Lex was subsequently manually enriched with morphological analysis giving rise to the CoDeRoMo resource [11] of which an updated version has been used as the main input for the Morphological profile. Furthermore, both corpora have been semi-automatically parsed for verb and noun patterns that currently constitute the core of the Grammatical profile.

## 2.2. Annotation

To prepare the data, we combined **automatic** processing [12, 13, 14] and **manual** annotation [11, 15, 16]. To support manual annotation and visualization of the process, the tool LEGATO [17] was implemented and a range of resources for Swedish were reused to inter-link the lexicographic and other information for each item on the Sen\*Lex list [18, 19, 20].

Manual annotation was used for: (1) classification of multiword expressions (MWEs) into subtypes [15] giving rise to a MWE subprofile under the Lexical profile; (2) morphological analysis of lemmas and word-formation patterns [11, 21] creating the basis for the Morpheme family and Word family under the Morphological profile; and (3) setting up regular expressions to extract noun and verb patterns from linguistically annotated texts used for Verb and Noun patterns under the Grammatical profile [16, 22].

## 2.3. User Interface

The User Interface for browsing the SweL2P<sup>2</sup> has some features that are shared by all its modules and subprofiles, such as a filter for CEFR levels, a possibility to enter a search item (except in the grammatical profile) and an option to see frequencies and samples from receptive/productive data for one's search. Some other features are specific for a (sub)profile in question.

**Filters** appear at the top of the page, providing a set of filters for each subprofile, e.g. *MWE Types 1-3* in Figure 1. The resource **can be explored** using several views: Table, Graphical and Statistical. The **Table** view (Figure 2 and 3) lists all items with associated information. Columns contain descriptive information, among others, a clickable word (e.g. 'jeans' in Figure 2) that opens a link to an entry with this item at <https://svenska.se/> and manual morphological analysis, as well as clickable receptive and productive (relative and absolute) frequencies that open a corpus search tool **Korp** [23] containing hits with those items. The range of the columns depends on the profile – notably, for the *Grammatical profile*, 'pattern' is listed instead of 'word' (Figure 3). **Graphical** and **Statistical** views summarize the statistics and distribution of various features for the current selection in the two sources, receptive and productive. For example, Figure 4 shows a graph view over the use of the Past tense (Sw. preteritum) across levels and a related table with the statistics; while Figure 5 demonstrates an excerpt from the statistics over the *Morpheme family* subprofile. In Figure 5, instead of graphs and distributions, we rather see counts in terms of types, tokens and type-token ratios per filter category so that we can study a statistical break-down of each selection contrasting receptive and productive competences.

The entire dataset or filtered data selection can be **downloaded** from the website.

## 3. Profiles

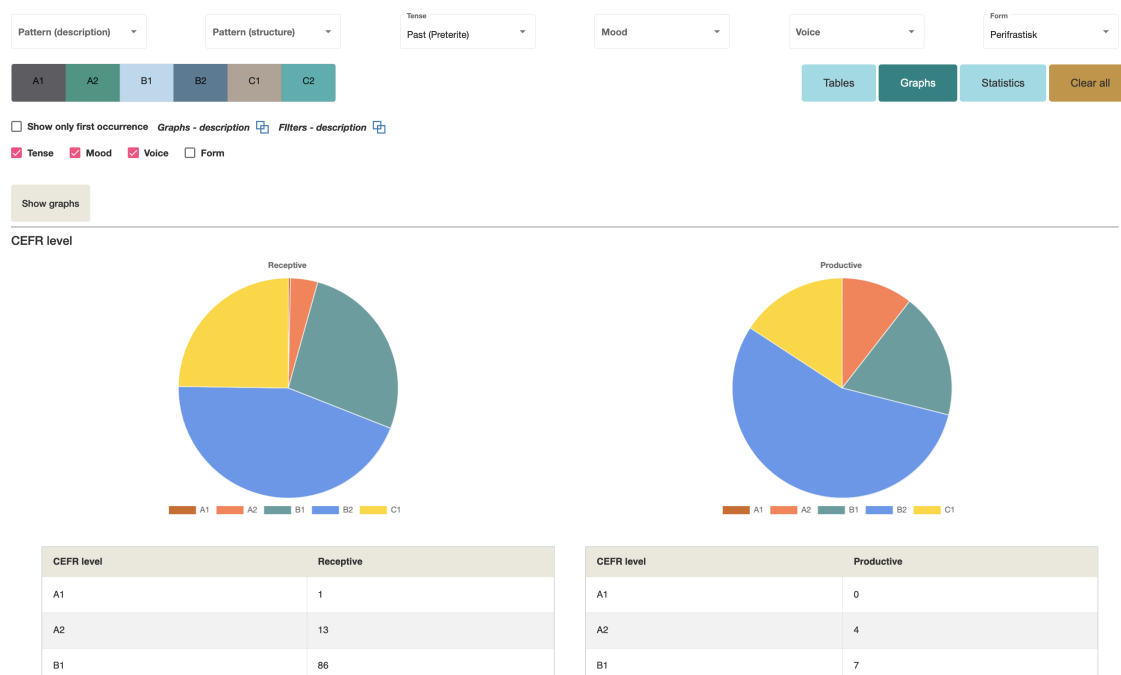
SweL2P covers three distinct areas of learner language – lexis, grammar and morphology.

### 3.1. Lexical profile

The *Lexical profile* includes four subprofiles, each representing a subsection of *Sen\*Lex*. **Adjectival declension** features one specialized filter for *declension* with four declensions: 1st, 2nd, suppletive and indeclinable. Through it, we can see how many of the adjectives in the two corpora that belong to the different declensions. The assignment of items into the different declensions was made semi-automatically. Similarly, the related subsection **Adjectival and Adverbial structure** presents adjectives and adverbs by types of *comparison* (morphological, periphrastic, both or unclassified) and *regularity* (regular, regular with umlaut, irregular). The items were semi-automatically categorized. **Multiword Expressions** allow the user to filter MWEs for (1) *syntactic construction* (contiguous or non-contiguous); (2) *lexical sub-categories*, including some which are closely related to parts of speech (POS), e.g. nominal, adjectival and verbal; (3) *verbal sub-categories*, with 4 types, e.g. particle verbs. MWEs were identified automatically, but the classification of MWEs into subgroups was done manually [15].

---

<sup>2</sup><https://spraakbanken.gu.se/larkalabb/svlp>, login: demo



**Figure 4:** Graphical view over Verb patterns (Grammatical profile): Statistics of (simple) past tense (Sw. preteritum) over CEFR levels

| CEFR level | Receptive (Type) | Productive (Type) | Receptive (Token) | Productive (Token) | Receptive (TTR) | Productive (TTR) |
|------------|------------------|-------------------|-------------------|--------------------|-----------------|------------------|
| A2         | 5153             | 715               | 16395             | 1870               | 31.43           | 38.24            |
| B1         | 10365            | 1189              | 27296             | 2835               | 37.97           | 41.94            |
| B2         | 10961            | 1318              | 22606             | 2778               | 48.49           | 47.44            |

| Morpheme Category   | Receptive (Type) | Productive (Type) | Receptive (Token) | Productive (Token) | Receptive (TTR) | Productive (TTR) |
|---------------------|------------------|-------------------|-------------------|--------------------|-----------------|------------------|
| Derivational prefix | 1941             | 337               | 5166              | 620                | 37.57           | 54.35            |
| Root morpheme       | 21448            | 3109              | 70004             | 8954               | 30.64           | 34.72            |
| Derivational suffix | 9133             | 1491              | 23780             | 3246               | 38.41           | 45.93            |

**Figure 5:** Statistical view (Morpheme family)

**Sen\*Lex** features a list of lexical items from the two corpora, ordered by sense-disambiguated lemmings [8, 9, 10] (see an example in Figure 2). The three specific filters, all automatically assigned to the items but manually checked and sometimes corrected, include: (1) Noun declension; (2) Gender; and (3) Conjugation.

| Verb patterns  | Noun patterns   |
|--|---|
| <i>Pattern</i> – 40 unique patterns, e.g. Imperative   | <i>Pattern</i> – 28 head patterns, 81 unique patterns, e.g. Nsg (Jag har körkort) |
| <i>Pattern structure</i> – 3 categories, e.g. Elliptic | <i>Pattern structure</i> – 2 categories, e.g. Simple noun phrase                  |
| <i>Tense</i> – 7 categories, e.g. Pluperfect           | <i>Clause position</i> – 2 categories, e.g. Initial                               |
| <i>Mood</i> – 4 categories, e.g. Subjunctive           | <i>Definiteness</i> – 2 categories, e.g. Definite                                 |
| <i>Voice</i> – 5 categories, e.g. Passive/Deponent     | <i>Gender</i> – 3 categories, e.g. Neuter   |
| <i>Form</i> – 3 categories, e.g. S-form                | <i>Number</i> – 2 categories, e.g. Plural   |
|  | <i>Attribute</i> – 4 categories, e.g. Relative clause                             |
|  | <i>Article</i> – 5 categories, e.g. Definite affix                                |
|  | <i>Other definite attributes</i> – 5 categories, e.g. Demonstrative pronoun       |

**Table 1**  
Overview of linguistic filters for Noun and Verb patterns (Grammatical profile)

### 3.2. Grammatical profile

The `Grammar profile` [16] features **Verb patterns** and **Noun patterns**. Both are based on structures parsed from the two source corpora, and are grouped under different patterns. The two subprofiles have the same general filtering options, such as CEFR levels and hits in receptive and productive data. However, they also have a number of unique linguistic filters (see Table 1 for details), some of which are based on automatic annotation.

### 3.3. Morphological profile

The `Morphological profile` is based on an updated version of `CoDeRoomor` [11] - a list containing 16 230 sense-based morphologically analyzed lemgrams organized into 4 986 morpheme families, of which 4 429 are word families (i.e. root families). To create `CoDeRoomor`, we used `Sen*Lex` list. A team of annotators analyzed each item on the list for their constituent morphemes (e.g. prefix, root) and word formation mechanism (e.g. compounding, derivation). The original `CoDeRoomor` resource can be freely downloaded as csv. or xlsx. files<sup>3</sup> or can be browsed and downloaded in a slightly updated form in the `SweL2P`.<sup>2</sup>

The `Morphological profile` includes **Morpheme family** and **Word family**. They can both be filtered for a morpheme, word, word class and word formation, and `Morpheme family` can additionally be filtered by `morpheme category`. Searching for a particular morpheme (e.g. root `bröd`), shows all lemgrams in the family (once for each level where they appear). Ticking `Only first occurrence`, shows unique items and their count. Selecting a proficiency level(s) shows how that morpheme/word family is represented there.

Morphemic analysis of the vocabulary facilitates various studies into language learning patterns, some examples presented in Volodina et al. [24, 25]. The resource is useful in a broader context, e.g. in NLP for training models for word segmentation into morphemes; in ICALL for the generation of exercises aimed at word-formation patterns, etc.

<sup>3</sup><https://spraakbanken.gu.se/en/resources/coderoomor>

## 4. Potential for research and teaching

The SweL2P has great potential for both research and teaching. Some initial research possibilities have to been explored in relation to morphology in [25], [24] and [26], for MWEs in [15], for place names in [27], and in relation to some initial grammatical patterns in [16]. The resource can be used by researchers but also by teachers to explore what is common or rare at different levels in coursebooks and/or learner essays. Coursebooks can be taken as a proxy for input and help us explore language acquisition in relation to usage-based theories (e.g. [28]). Similarly, a teacher who finds that students find it challenging to learn how to use e.g. double definiteness (Sw. *den gula bilen* [DEF yellow-DEF car-DEF] 'the yellow car') or a particular subjunction can use the resource to explore the actual empirical usage. Thanks to the fact that statistics from both corpora are presented in parallel there is an excellent possibility to ascertain if a certain challenge is likely to be related to a lack of examples in input. Furthermore, since the resource also contains links to the searches in Korp this can then easily be explored further also in reference corpora with the same annotation.

## 5. Concluding remarks and future prospects

Language learning profile resources exist predominantly for English, e.g. English profile [29], CEFR-J [30] and Pearson's GSE Teacher Toolkit [31]. Most languages have nothing similar, the L2 Estonian Teacher's Tools [32] being one of the first non-English profiles. Even when the existing profiles have been based on empirical corpus data, this data is not openly provided, rendering them rather prescriptive. SweL2P takes a descriptive view of the language and provides access to the empirical evidence, i.e. all corpus hits and statistics of actual usage. It lets users zoom in on actual data and draw their own conclusions based on the empirical data with the help of visualizations. Including both receptive and productive frequencies side-by-side the resource gives a more nuanced picture of language learning. Through that and the special efforts invested in the visualization of the data, the SweL2P tool is more readily appropriate for research on SLA than any predecessor known to us. Furthermore, the inclusion of links to searches in Korp makes it easy to compare with other corpora from different varieties of Swedish. Finally, the open nature of the resource makes it highly useful for future learning apps, for training of automatic tools, and for teaching.

In the future, we envisage efforts invested into the disambiguation of morphemes, adding more patterns to the grammar profile (e.g. word order), adding reference corpus statistics, expanding visualization techniques (e.g. word clouds), and user upload of data for analysis.

## Acknowledgments

Work on the Swedish L2 Profile has been supported by a research grant from the Swedish Riksbankens Jubileumsfond *Development of lexical and grammatical competences in immigrant Swedish*, P17-0716:1 (2018–2021). Work on the article has been supported by Nationella språkbanken and HUMINFRA, both funded by the Swedish Research Council (2018–2024, contract 2017-00626; 2022–2024, contract 2021-00176) and their participating partner institutions.

## References

- [1] M. Islam, S. Jin, An overview of data visualization, in: 2019 International Conference on Information Science and Communications Technologies (ICISCT), IEEE, 2019, pp. 1–7.
- [2] C. Sievert, Interactive web-based data visualization with R, plotly, and shiny, CRC Press, 2020.
- [3] S. Coulange, M.-P. Jouannaud, C. Cervini, M. Masperi, From placement to diagnostic testing: Improving feedback to learners and other stakeholders in SELF (Système d’Evaluation en Langues à visée Formative), *Language Learning in Higher Education* 10 (2020) 195–205.
- [4] M. L. Waskom, Seaborn: statistical data visualization, *Journal of Open Source Software* 6 (2021) 3021.
- [5] E. Volodina, I. Pilán, S. R. Eide, H. Heidarsson, You get what you annotate: a pedagogically annotated corpus of coursebooks for Swedish as a Second Language, in: *Proceedings of the 3rd Workshop on NLP for Computer-Assisted Language Learning*, Linköping University Press, 2014.
- [6] E. Volodina, I. Pilán, I. Enström, L. Llozhi, P. Lundkvist, G. Sundberg, M. Sandell, SweLL on the rise: Swedish learner language corpus for European reference level studies, in: *Proceedings of the 10th Language Resources and Evaluation Conference (LREC)*, Portorož, Slovenia, 2016.
- [7] C. of Europe, Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion Volume with new descriptors, Council of Europe Publishing, 2020.
- [8] T. François, E. Volodina, I. Pilán, A. Tack, 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’16)*, 2016, pp. 213–219.
- [9] E. Volodina, I. Pilán, L. Llozhi, B. Degryse, T. François, SweLLex: second language learners’ productive vocabulary, in: *Proceedings of the joint workshop on NLP for Computer Assisted Language Learning and NLP for Language Acquisition*, 2016, pp. 76–84.
- [10] D. Alfter, Exploring natural language processing for single-word and multi-word lexical complexity from a second language learner perspective, *Data Linguistica* 31, University of Gothenburg, 2021.
- [11] E. Volodina, Y. A. Mohammed, T. Lindström Tiedemann, CoDeRoomor: A new dataset for non-inflectional morphology studies of Swedish, in: *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, 2021, pp. 178–189.
- [12] L. Borin, M. Forsberg, M. Hammarstedt, D. Rosén, R. Schäfer, A. Schumacher, Sparv: Språkbanken’s corpus annotation pipeline infrastructure, in: *The Sixth Swedish Language Technology Conference (SLTC)*, Umeå University, 2016, pp. 17–18.
- [13] L. Nieto Piña, Splitting rocks: Learning word sense representations from corpora and lexica, *Data Linguistica* 30, University of Gothenburg, 2019.
- [14] E. Volodina, D. Alfter, T. Lindström Tiedemann, M. S. Lauriala, D. H. Piipponen, Reliability of automatic linguistic annotation: native vs non-native texts, in: *Selected papers from the CLARIN Annual Conference 2021*, Linköping University Electronic Press (LiU E-Press), 2022.
- [15] T. Lindström Tiedemann, D. Alfter, Y. A. Mohammed, D. Piipponen, B. Silén, E. Volodina,



Multiword expressions in Swedish as second language: taxonomy, annotation and initial results, in: Multiword expressions in language resources. Linguistic, lexicographic and computational considerations, acc.

- [16] T. Lindström Tiedemann, Y. A. Mohammed, E. Volodina, Swedish Grammar profiling for empirical L2 research and teaching (subm.).
- [17] D. Alfter, T. Lindström Tiedemann, E. Volodina, LEGATO: A flexible lexicographic annotation tool, in: Proceedings of the 22nd Nordic Conference on Computational Linguistics, 2019, pp. 382–388.
- [18] L. Borin, M. Forsberg, L.-J. Olsson, J. Uppström, The open lexical infrastructure of Språkbanken., in: LREC, 2012, pp. 3598–3602.
- [19] L. Borin, M. Forsberg, L. Lönngrén, SALDO: a touch of yin to WordNet’s yang, Language resources and evaluation 47 (2013) 1191–1211.
- [20] L. Borin, M. Forsberg, Swesaurus; or, The Frankenstein approach to Wordnet construction, in: Proceedings of the Seventh Global Wordnet Conference, 2014, pp. 215–223.
- [21] E. Sköldberg, L. Holmer, E. Volodina, I. Pilán, State-of-the-art on monolingual lexicography for Sweden, Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave 7 (2019) 13–24.
- [22] U. Teleman, S. Hellberg, E. Andersson, Svenska akademiens grammatik, Svenska Akademien and Norstedts ordbok, 1999.
- [23] L. Borin, M. Forsberg, J. Roxendal, Korp-the corpus infrastructure of Språkbanken., in: LREC, volume 2012, 2012, pp. 474–478.
- [24] E. Volodina, Y. Ali Mohammed, T. Lindström Tiedemann, Swedish Word Family Resource – Construction, Applicability, Strengths and first Experiments (subm.).
- [25] E. Volodina, Y. Ali Mohammed, T. Lindström Tiedemann, Lyxig språklig födelsedagspresent from the Swedish Word Family, in: In Volodina, Dannélls, Berdicevskis, Forsberg and Virk (editors). Live and Learn: Festschrift in honor of Lars Born, GU-ISS-2022-03, Department for Swedish, Multilingualism, Language Technology, University of Gothenburg, 2022, pp. 153–160.
- [26] A. Ingves, T. Lindström Tiedemann, Prefix i svenskan – en ordinlärningsresurs för inlärare av svenska som andraspråk?, in: Svenskans beskrivning: Förhandlingar vid trettioåttonde sammankomsten, volume 1, in print.
- [27] T. Lindström Tiedemann, Egennamn, morfologi och andraspråksinläring, in: Namn och gränser: rapport från den sjuttonde nordiska namnforskarkongressen den 8–11 juni 2021, NORNA-förlaget, 2023.
- [28] J. Bybee, Usage-based grammar and second language acquisition, in: Handbook of cognitive linguistics and second language acquisition, Routledge, 2008, pp. 226–246.
- [29] A. O’Keeffe, G. Mark, The English grammar profile of learner competence: Methodology and key findings, International Journal of Corpus Linguistics 22 (2017) 457–489.
- [30] Y. Tono, Coming full circle—from CEFR to CEFR-J and back, CEFR Journal 1 (2019) 5–17.
- [31] V. Benigno, J. de Jong, Developing the global scale of English vocabulary for young learners (6 to 11), 2017.
- [32] T. Üksik, J. Kallas, K. Koppel, K. Tsepelina, R. Pool, Estonian as a second language teacher’s tools, in: Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications, 2021, pp. 130–134.

## **A. Online Resources**

Online resources mentioned in the article:

- Swedish L2 Profile (login: demo)
- CoDeRoom resource
- English Profile
- CEFR-J toolkit
- L2 Estonian Teacher's Tools

Guidelines for manual annotation:

- Lexicographic annotation guidelines: multi-word expressions, adjectives and adverbs
- Official L2P morphology annotation guidelines
- Noun pattern descriptions for Swedish
- Verb pattern descriptions for Swedish