

Chapter 9

Multiword expressions in Swedish as a second language: Taxonomy, annotation, and initial results

Therese Lindström Tiedemann^a, David Alfter^b, Yousuf Ali Mohammed^b, Daniela Piipponen^a, Beatrice Silén^a & Elena Volodina^b


^aUniversity of Helsinki, Finland ^bUniversity of Gothenburg, Sweden

This chapter introduces part of the Swedish L2 profiles, a new resource for Swedish as a second language. Multiword expressions (MWEs) in this resource are based on knowledge-based automatic annotation of MWEs, which we show works quite well for Swedish. In contrast, manual annotation of the compositionality of each MWE proved difficult, probably due to different interpretations of “compositionality” by the two annotators. We show that experts and non-experts can rank MWEs very similarly according to relative receptive difficulty, with particularly high agreement for the easiest items. A qualitative comparison of the proficiency levels associated with the MWEs based on coursebook occurrences and the results from crowdsourcing and direct ranking indicate that MWEs which appear in few books of the same level are more likely to be difficult to associate with an appropriate level based on coursebook corpus data. Furthermore, results show that compositionality and/or transparency might influence the relative ranking. Finally, there is a clear increase in MWE lemmas at higher proficiency levels at the group level, and at the highest level receptive and productive data include the same percentage of MWEs.

1 Introduction

Previous research has clearly shown that multiword expressions (MWEs) are an important part of idiomatic language use (e.g. Paquot 2019), but also that they



Therese Lindström Tiedemann, David Alfter, Yousuf Ali Mohammed, Daniela Piipponen, Beatrice Silén & Elena Volodina. 2024. Multiword expressions in Swedish as a second language: Taxonomy, annotation, and initial results. In Voula Giouli & Verginica Barbu Mititelu (eds.), *Multiword expressions in lexical resources: Linguistic, lexicographic, and computational perspectives*, 309–348. Berlin: Language Science Press. DOI: 10.5281/zenodo.10998647 

are a challenge even to advanced second language (L2) learners (Pawley & Syder 1983, Wray 2002) and show a clear correlation to one's level of proficiency (Forsberg 2010). MWEs can be seen as:

...sequences of words that are in some regard *not entirely predictable*, whether on account of a meaning that is wildly or subtly different from the words they contain, a function that is only achieved with the whole expression, or features of structure such as morphology or word order that are non-canonical. (Wray 2013: 317, italics added)

This clearly entails that MWEs are an additional complication in second language acquisition (SLA). It also means that similarly to single word lexemes, MWEs have to be learnt as lexemes.

Based on previous research showing the challenges of MWEs in SLA and in relation to current advances in automatic evaluation, it is important to consider whether MWEs can be seen as particularly criterial of certain levels, but also how well MWEs can be automatically annotated in learner texts even though these texts tend to contain issues which do not follow the norm in the target language. Furthermore, ways of linking MWEs to proficiency levels need to be explored.

We are primarily interested in MWEs in relation to the acquisition of Swedish as a second language (L2 Swedish), however most of our results should be of interest also in relation to other languages and for SLA in general. Second language acquisition of Swedish MWEs has been studied through experiments, questionnaires and occasionally also in learner texts (e.g. Prentice & Sköldberg 2013; Enström 1990; Abrahamsson & Hyltenstam 2009).

In this chapter we present studies of Swedish MWEs based on authentic L2 data, both receptive and productive, linked to proficiency levels according to the Common European Framework of Reference for Languages (CEFR, Council of Europe 2001). We argue for the usefulness of automatic annotation of MWEs combined with additional manual annotation, and discuss the possibilities of linking MWEs to CEFR levels based on authentic data, crowdsourcing and expert annotation. Since many languages tend to have less resources than English and because we know that there will always be new expressions which will need to be linked to proficiency levels, we want to find a cheap and reliable way of linking MWEs to levels. We therefore explore crowdsourcing with relative judgement. Since it is likely to be cheaper to use non-experts and because it is interesting as a research question, we want to see if experts (L2 Swedish teachers, assessors, researchers) and non-experts (L2 Swedish learners) agree on their ratings. In addition, we also explore the possibility of explicit level ranking by experts. In this

chapter, we summarise previously published results (see Alfter et al. 2021, Lindström Tiedemann et al. 2022) and present further qualitative analyses of some of our results from these experiments.

Our study is based on MWEs found through automatic annotation of coursebooks for L2 Swedish aimed at adult learners (the Coctail corpus, Volodina et al. 2014) and L2 Swedish learner essays (the SweLL-pilot corpus, Volodina et al. 2016a). We summarise the results of our annotation check as previously published (Volodina et al. 2022b), showing that MWEs in these different materials can be well annotated automatically.

The identified MWEs were manually categorised according to our Swedish taxonomy for MWEs, see Section 3. By using our taxonomy to compare receptive and productive usage of MWEs we present how our data, including our manual annotation, can be accessed through an open lexical resource online (Swedish L2 Profiles)¹ that can be used for further research, as well as for teaching.

We aim to answer the following research questions:

1. How well can MWEs be automatically annotated in L2 coursebooks and L2 learner texts?
2. How well does the occurrence of MWEs in authentic materials coincide with (a) ranking results from an expert or a non-expert crowd; (b) direct annotation by experts?
3. How do different MWE types appear over CEFR levels in receptive and productive data for L2 Swedish? Are certain MWE types more challenging to L2 Swedish learners based on a comparison of their occurrence in receptive and productive data?

First we present some previous research on MWEs in relation to SLA and L2 Swedish in particular (Section 2). We then present our Swedish MWE taxonomy (Section 3) after which our materials and method are introduced, including the annotation tools that we use (Section 4). In Section 5 we present our results, and in Section 6 we summarise our conclusions and look ahead.

2 Previous research

MWEs are a broad and vaguely defined phenomenon. Research articles and books contain a multitude of different terms with similar meanings: collocations (Bhalla

¹<https://spraakbanken.gu.se/larkalabb/svlp> (login: demo)

& Klimcikova 2019), phraseological units (Paquot 2019), lexicalised phrases (Sag et al. 2002), fixed expressions (Villada Moirón 2005), formulaic language (Durrant 2018), lexical bundles (Granger 2018), words-with-spaces (Sag et al. 2002), formulaic sequences (Wray & Perkins 2000, Wray 2002). Wray (2002) lists c. 60 terms for similar concepts and notes the problem of the varying terminology and that terms tend to have strong connections to certain theories or methods. This also means that even when the same term is used we cannot be certain that it means the same. When working on MWEs in a language other than English this causes additional challenges even with a language as closely related to English as Swedish, since this plethora of terms needs to be compared to terminology which has been used in descriptions in that language.

The multitude of terms is partly a result of the many different approaches to MWEs. Bhalla & Klimcikova (2019) name three main approaches to studying and classifying collocations, and by extension MWEs, namely:

1. Psychological – lexical associations in the mental lexicon;
2. Phraseological – dealing predominantly with separating MWEs from free word combinations based on semantic principles; and
3. Distributional – focusing on the manifestations of MWEs in corpora based on frequency, distribution, and degree of co-occurrence.

This captures the complexity of the phenomenon and the variety of ways in which it can be perceived. It also underscores the practical needs to identify MWEs and to categorise them into subcategories. There is a need to explain their (typical and atypical) behaviour in relation to various fields, e.g. lexicography, language learning, clinical linguistics and Natural Language Processing (NLP). For lexicography, it is important to have an approach to listing MWEs, to grouping them into specialised lexicons, as well as an approach for the identification of new MWEs (Agirre et al. 2006). Identification of MWEs relies on NLP approaches (Baldwin & Bond 2002, Sag et al. 2002, Piao et al. 2005, Attia et al. 2010, de Caseli et al. 2010, Watrin & François 2011, Shigeto et al. 2013) and thus requires formalisation of the definition of MWEs, something we explore further in relation to our study below.

2.1 MWEs in SLA research

Several studies have shown that MWEs are a major part of our lexical competence. Jackendoff (1997) claims that c. 50% of our mental lexicon consists of

MWEs, while Erman (2007: 28) argues that they may in fact form an even larger part of our language since Mel'čuk (1998: 24) claims that MWEs (or phrasemes) "outnumber words roughly ten to one" (where, by "words", presumably, Mel'čuk means single lexical items). Interestingly, experiments have shown that L2 and L1 speakers process language differently. L2 speakers apparently rely primarily on frequency, whereas L1 speakers rely more on Mutual Information (MI) in processing MWEs (Ellis 2012: 24).

Some researchers have claimed that MWEs are frequent also in learner language, sometimes assuming that they are more common at lower proficiency levels (Wray 2002: 173 citing the work of others). Ellis (2012: 18) claims that "Zipf's (1935) law and the 'phrasal teddy bear' explain the paradox whereby formulas seed language acquisition and yet learners typically do not achieve native-like formulaicity". Formulaic language has been claimed "the biggest stumbling block to sounding nativelike" (Wray 2002: ix). Similarly, CEFR documentation claims that "idiomatic expressions and colloquialisms" are not likely to be fully acquired before C2 (Council of Europe 2009: 185, 187), although many "idiomatic expressions and colloquialisms" should be *understood* at C1 (Council of Europe 2009: 124, 143).

Research has shown both an increased use of MWEs and more native-like usage in terms of distribution, as the proficiency increases (e.g. Forsberg 2006 as cited in Ringbom 2012 for prefabs in L2 French (L1 Swedish); Forsberg & Bartning 2010 with regards to lexical formulaic sequences). Still, MWEs remain difficult even for advanced learners (Nesselhauf 2003: 237, Ringbom 2012: 496; Ekberg 2013) as do specific MWEs such as idioms and proverbs (Abrahamsson & Hyltenstam 2009, Prentice 2010) and since it is "clearly impossible to teach all (or even most) of the collocations in a language, criteria have to be set up to determine which collocations should be included in a given syllabus" (Nesselhauf 2003: 238). Furthermore, Forsberg & Bartning (2010: 150) showed that the increase was not always statistically significant, something they believed might be due to the low number of essays per level and also the fact that the texts are often fairly short.

Unidiomaticity in learner languages has often (and for a long time) been linked to MWEs (cf. Pawley & Syder 1983) probably due to the fact that there is such a multitude of complicating factors in relation to MWEs. De Cock et al. (2014: 78) claim that the problems with MWEs concern: (1) the extent to which they are used, (2) the MWEs that are used, and (3) how they are used.

How MWEs are used by L1 and L2 speakers have been important issues in recent research within SLA and within learner corpus research, but often from a fairly open perspective on collocations which focuses on how words are used

together with other words based on statistical measures such as MI and log likelihood. However, as stressed by Forsberg & Bartning (2010: 148) these measures require quite large datasets, which is a prerequisite not met by our dataset. We have therefore opted to focus on a knowledge-based approach.

2.2 MWEs and language teaching

Nesselhauf (2003: 223) concluded that MWEs should be seen as “an important part” of L2 teaching, in particular at advanced levels, and that the difficulties which learners experience with collocations require more research. In connection with language learning, lists of MWEs are very useful. There are lexical resources of this kind, but for languages such as Swedish it is hard to find materials with indications of proficiency levels. Furthermore, materials where levels have explicit and transparent information about their grounding in empirical data are quite rare, open access to receptive and productive data being even less common.

One possibility is to use corpora, and several studies have shown that corpora can be useful both to introduce MWEs and to work with *noticing* (Schmidt 2012) strategies (cf. Meunier 2012). Nevertheless, teachers and learners rarely have access to information about how the MWEs occur in learner language or even in data aimed at learners. This is a shame since access to MWEs in data opens possibilities for working with noticing as well as contextualising the usage (see e.g. Boers et al. (2006) who studied how MWEs can be taught with the help of noticing).

Online lexicographic reference sites with information about the MWEs which can be expected at particular CEFR levels are available for English. The English Profile² (Hawkins & Filipović 2012, Green 2012, Kurtes & Saville 2008) is explicitly based on learner data, but does not provide access to frequencies of use or more than the odd example of use in the entry itself. The English Vocabulary Profile (Capel 2012, 2015) makes it possible to select phrases, phrasal verbs or idioms, all of which contain some MWEs. However, there is currently no possibility to select MWEs as a superordinate category. The English Grammar Profile (O’Keeffe & Mark 2017) enables the selection of e.g. phrases/exclamations, expressions with *be*, or items which have been subcategorised as “phrasal”. However, there is no category for MWEs in general. Similarly, Pearson’s Global Scale of English (GSE)³ provides lists of MWEs with proficiency level information, but no frequencies and no access to empirical data which the reference has been based on. The user can choose phrasal verbs and/or phrases. The category ‘phrases’ in

²<https://www.englishprofile.org/>

³<https://www.pearson.com/languages/why-pearson/the-global-scale-of-english.html>

Pearson's GSE includes phatic communication, asking about prices, introducing yourself, and idiomatic expressions and is therefore broader than MWEs, since it deals more with communicative phrases. The Swedish L2 lexical profile, which we are introducing here, provides more information about MWEs. Furthermore, it not only includes frequencies from both receptive and productive empirical data, it also provides access to the data.

2.3 Assigning proficiency levels to lexical items

Even though CEFR focuses on communicative competence, the CEFR documentation (Council of Europe 2001, 2020) still indicates that we should try to associate lexicon items to CEFR proficiency levels. Previous work on assigning levels to words and MWEs based on corpora have mainly focused on coursebook corpora (Gala et al. 2013, 2014, François et al. 2014, 2016, Dürlich & François 2018, Tack et al. 2018), although some works also used learner corpora (Volodina et al. 2016b, Alfter et al. 2016). It has generally been found that a simple method of assigning levels, i.e., using the first level at which an expression occurs, performs better, or at least equally well to more sophisticated methods (Gala et al. 2014, Alfter 2021). However, as the majority of works have focused on coursebooks, it should be noted that other methods of assigning levels, such as threshold approaches, may be more suitable for learner language (Alfter 2021, Yamaguchi et al. 2022). Furthermore, frequency based approaches such as the above may not be well suited to assigning levels to MWEs, as these expressions tend to be less frequent.

2.4 MWEs, compositionality and transparency

Some MWEs such as idioms have often been discussed as examples which go against the compositionality principle in language. However, compositionality is easily confused with transparency, and there is a need to investigate its relation to other features of the MWEs and their constituents (Schulte im Walde 2024 [this volume]). Research on idioms have included debates regarding how semantically analysable idioms are (Cieślicka 2015). These discussions have compared idioms like *spill the beans* and *kick the bucket*, the first has been seen as semantically compositional, since we can imagine each word in the expression being a metaphorical rendering of something: *spill* 'tell something' and *beans* 'secrets' (Nunberg et al. 1994, Cieślicka 2015). However, the second expression is not compositional in this way. This means that an idiom can be seen as decomposable or compositional even when it is figurative and non-transparent. Hence, according to Cieślicka (2015) and Nunberg et al. (1994: 495) decomposability (cf. also compositionality) is not the same as transparency.

2.5 Swedish MWEs

As in international research, there is also a multitude of Swedish terms which have been used in relation to MWEs. Sometimes the Swedish terms are very similar to the English terms, but they do not necessarily mean exactly the same. We aim to make it possible to relate our work to both international and Swedish terminology, which is why we sometimes give both an English and a Swedish term for the sake of clarity. Swedish terms will be preceded by (sv) just as Swedish examples.

Lexicalised phrases, (sv) *lexikaliserade fraser* (lit. ‘lexicalised phrases’), are discussed by Anward & Linell (1976) in a more restrictive sense than the one presented in Sag et al. (2002) and which we have adopted in our study. They focused on a type of lexicalised phrases which have connective prosody where the main stress is on the right-hand part of the expression. They exemplify this with (sv) *en varm korv* (lit. ‘a hot sausage’) ‘a hot sausage’ as opposed to the compound (sv) *en varmkorv* (lit. ‘a hot-sausage’) ‘a hot dog’. Furthermore, according to them, lexicalised phrases can be inflected and syntactically modified internally through separate lexemes. This is sometimes possible in the non-contiguous lexicalised phrases in the Saldo lexicon (Borin et al. 2013) and in our taxonomy, but not always since this is also related to the compositionality and transparency of the expression.

Anward & Linell (1976: 80–81) further divided lexicalised phrases into rather specific subcategories such as premodified noun phrases (NP), e.g. (sv) *Vita huset* (lit. ‘the white house’) ‘The White House’; definite NP with a preposed epithet, e.g. (sv) *profeten Jesaja* (lit. ‘the prophet Jesaja’) ‘Isaiah the Prophet’; definite NP with a postmodifier, e.g. (sv) *Gustav III* (lit. ‘Gustav the third’) ‘Gustav the third’ or (sv) *mannen på gatan* (lit. ‘the man on the street’) ‘common man’; adjectival phrases with prepositional phrasal modifiers, e.g. (sv) *ont i halsen* (lit. ‘sore in the throat’) ‘a sore throat’ etc.

In Swedish linguistics, MWEs ((sv) *flerordsenheter* (lit. ‘multiword-units’) see e.g. Prentice & Sköldbberg 2013) are primarily treated as specific subcategories: e.g. particle verbs ((sv) *partikelverb*);⁴ reflexive verbs ((sv) *reflexiva verb*); idioms ((sv) *idiom*); proverbs ((sv) *ordspråk*) and lexicalised compounds ((sv) *lexikaliserade sammansättningar*). Support verb constructions have also been studied and are referred to as (sv) *funktionsverbförbindelse* (lit. ‘function verb relation’) in the Swedish Academy Grammar (SAG, Teleman et al. 1999); e.g. (sv) *falla i glömska* (lit. ‘to fall into oblivion’) ‘to be forgotten’. Even though the English term *support verb* is very different from the Swedish term, we believe that the meaning is sufficiently close to allow us to use this term.

⁴We call these “particle verb” in English to reflect the Swedish terminology even though they are similar to phrasal verbs.

The term *collocation* ((sv) *kollokation*) is also used in Swedish. Prentice & Sköldb-berg (2013) define collocations as words with a strong association between them and among the examples one can find e.g. (sv) *fatta beslut* (lit. ‘to grab (a) decision’) ‘to come to a decision’ which we would classify as a support verb construction. Exactly what makes something a “strong association” is not clear, but as international literature has often seen “collocations” as a statistical association we believe *MWE*, and the Swedish term *flerordsenhet*, are better terms to use in the context of our work.

An additional complication both in learning and automatically finding and annotating Swedish MWEs is that there are some MWEs which show variation regarding whether they are written as a MWE, or as a single word. For instance, there are adverbs, which are highly lexicalised but which are often written as separate words, e.g. (sv) *i dag* (lit. ‘in day’) ‘today’, (sv) *i går*⁵ (lit. ‘in yesterday’) ‘yesterday’, *över huvud taget* (lit. ‘over head taken’) ‘at all’. The official recommendation for many of these has been to write them apart, but there has been a fair amount of variation, and lately the recommendations have become more relaxed and primarily emphasise consistency (cf. Karlsson 2017, Svenska Akademien 2015). In our manual annotation we currently only annotate the multiword instances of these words and it is only those that appear in the listings in the MWE part of the Swedish L2 Profiles. However, any analysis of these types of MWEs should also consider the single-word variants and it would be good if future work could also include them in the profile next to the multiword instances.

3 Swedish MWE taxonomy

It is important to explore (1) whether some MWEs appear to be easier to learn and (2) which MWEs or MWE types tend to be learnt only at more advanced proficiency levels. Individual MWEs are likely to be highly linked to certain topics. However, since there are many different kinds of MWEs it is interesting to see if learning patterns can be found if we look at how the MWE types occur in both coursebooks for L2 learners and in texts which learners produce, rather than looking at individual MWEs. If so, types could be taken into account more both in teaching and in assessment. In this section we present how we have designed our MWE taxonomy based on previous international research on MWEs, research on Swedish MWEs and in relation to the automatic annotation pipeline we use.

⁵The word *går* is only used in this expression and in the noun *gårdagen* ‘yesterday’ in present day Swedish.

Erman & Warren (2000) distinguished MWEs (formulaic sequences) into lexical, grammatical and discursive ones (prefabs) and applying this taxonomy Forsberg (2008) and Lewis (2008) both found that lexical MWEs were most problematic to L2 learners (L2 French and L2 English respectively) (Forsberg & Bartning 2010). Our taxonomy has a similar division but it is more detailed (see Figure 1). We have two ambitions with our taxonomy:

1. A taxonomy that supports L2 Swedish research, teaching, and learning. It should be connected to what learners might find easy or difficult in learning L2 Swedish. For instance, particle verbs and reflexive verbs can be challenging for learners (cf. Enström 1990, Ekberg 1999).
2. A taxonomy that is computationally useful. While MWEs in this work were automatically identified, our taxonomy could further enhance automatic MWE recognition, which in turn could impact downstream tasks such as parsing efficiency positively.

We want to be able to start from the output of the annotation pipeline. The Sparv-pipeline (Borin et al. 2016) which we use (see Section 4.2 for more details on Sparv) is knowledge-based and depends upon entries currently in the Saldo lexicon (Borin et al. 2013). As part of the lemmatisation, MWEs in texts are identified through Saldo. This means that if the MWE does not have an entry in Saldo it will not be recognised, and if something is not seen as part of a MWE in Saldo it will not be part of the MWE in the list of MWEs which we work with. The latter is the case with certain prepositions since they can either be seen as part of the MWE or as part of the valency of the MWE, cf. (sv) *ha ont (i)* (lit. ‘have ache (in X)’) ‘have a (X) ache’, or (sv) *ta reda (på något)* (lit. ‘take control/organisation (on something)’) ‘find out (something)’.

There are several potential problems with identifying MWEs based on Saldo lookup for work on L2 Swedish:

1. The MWE annotation might not be reliable. There was no previous evaluation of how reliable the Sparv-pipeline is at identifying MWEs, that is, whether it produces too many false positives (overgenerating) or false negatives (undergenerating). We have therefore performed an annotation check as presented in Volodina et al. (2022b) and will summarise and discuss this in Section 5.1 with regards to MWEs.
2. The annotation pipeline may not be reliable on L2 production. L2 production does not necessarily conform to the standard variety of the target

language. This means that the recognition of MWEs is likely to be more complicated, since the pipeline has been trained on fairly normlike texts written (primarily) by L1 speakers. This is also something which we studied in Volodina et al. (2022b) and which is summarised in Section 5.1.

3. The lexicon might not contain the MWEs which are used. Sparv lemmatisation is based on the Saldo lexicon which is far from exhaustive when it comes to MWEs. Borin (2021) claims that MWEs make up 6% of the Saldo lexicon. For comparison, Sag et al. (2002) cite that 41% of WordNet consists of MWE entries (but WordNet entries only include senses for lexical word classes, whereas Saldo also includes senses for grammatical word classes and this is likely to affect the percentage of MWEs). As seen in Section 2, there have been claims that the number of MWEs are equal to the number of single-item entries (Jackendoff 1997), or that there may be ten MWEs to every single item (Mel'čuk 1998). Thus, it is relatively safe to assume that a fair share of Swedish MWEs are not listed in Saldo and would therefore be missed during the automatic linguistic annotation.
4. Saldo does not include 'strong collocates' (i.e. institutionalised phrases). These are also an important "near-phraseological" knowledge for L2 learners and have frequently been looked at in studies of formulaic language in SLA (cf. Section 2). Hence future research needs to find ways to add less lexicalised MWEs to a lexicographic resource aimed at language learners.

In Section 5.1 we show that (1) and (2) are not really an issue and in fact the same checks indicate that (3) also is not a large problem for our data since most MWEs were annotated. We will however have to leave (4) to future research.

In our taxonomy we have tried to take into account previous research on MWEs both regarding second language acquisition and the Swedish language. Our aim is that the categories should be easily relatable to both, and possible to justify formally in such a way that they could facilitate later computational use of the taxonomy.

We focus on conventionalised collocations but we have decided against using the term *collocation*. This is partly because it is often associated with the statistical method of identifying MWEs based e.g. on n-grams or less lexicalised phrases as discussed above. However, *lexicalised phrases* in our sense is sometimes the same as a *collocation* according to others; for instance, Cowie (1994)'s and Nesselhauf (2003)'s use of collocation relies on there being an "arbitrary restriction on substitutability" (Nesselhauf 2003: 225) which is similar to our idea of lexicalised phrases.

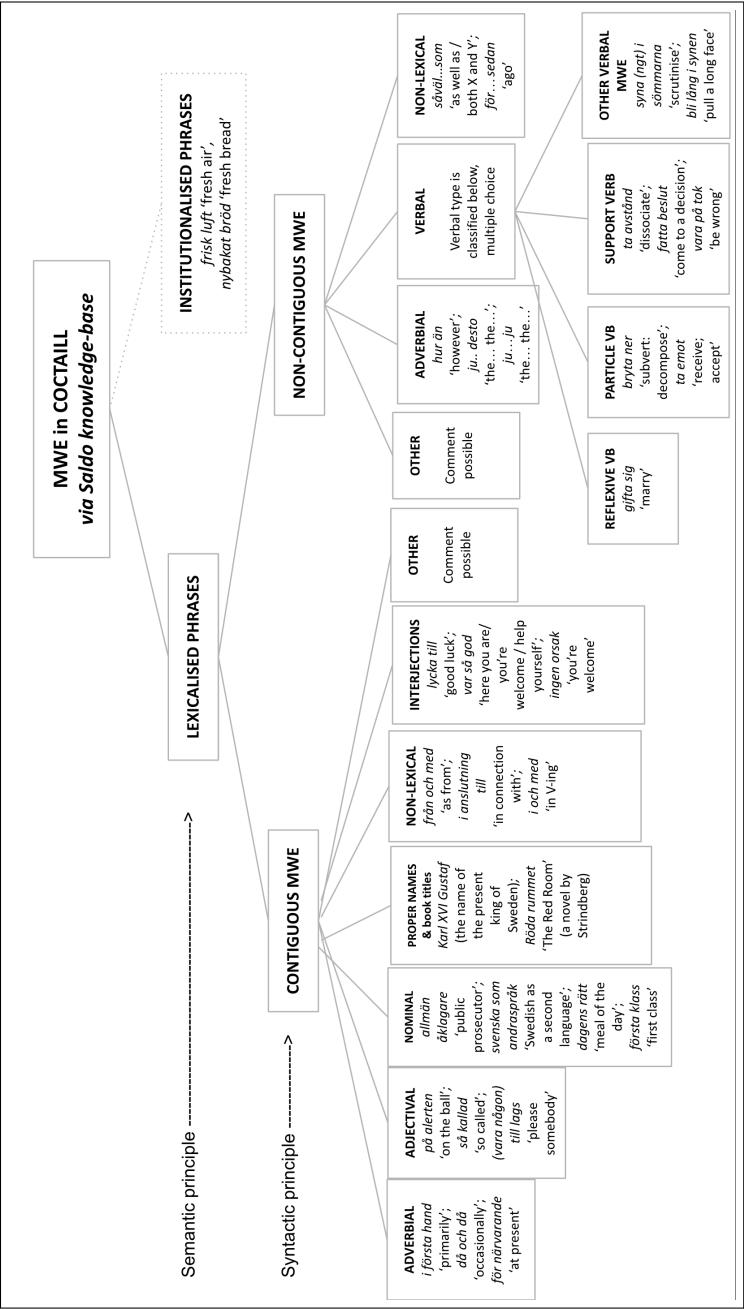


Figure 1: Our MWE taxonomy, below each category there are some Swedish examples to help the annotators remember the definition.

In our taxonomy (Figure 1), we divide MWEs into *lexicalised phrases* (cf. Sag et al. (2002)) and *institutionalised phrases* (cf. “strong collocates” or Cowie’s “free combinations”). The latter are not currently included in the Sparv pipeline and are therefore not included in our current work. Lexicalised phrases are further divided into *contiguous* and *non-contiguous MWEs*⁶ according to syntactic principles and later subcategorised according to word class where the syntactic function associated with different word classes is of particular importance (cf. Leseva et al. 2024 [this volume]). The contiguous MWEs are therefore divided into: adverbial MWEs, adjectival MWEs, nominal MWEs, non-lexical MWEs, but also proper names where we include book titles. One might not need to learn book titles, but we classify any which occur since they are a type of MWEs and some of them are such that they can be expected to be part of the common knowledge of Swedish speakers, e.g. *Röda rummet* ‘The Red Room’, a famous novel by the Swedish nineteenth century author August Strindberg. Finally, we also include a contiguous category of interjections since e.g. greetings are common in learner language and often consist of lexicalised MWEs.

Among the non-contiguous MWEs we find adverbial MWEs, verbal MWEs and non-lexical MWEs. The verbal category contains several sub-categories which are often mentioned in both teaching and research. Therefore we annotate if the verbal MWE is: a reflexive verb e.g. (sv) *gifta sig* (lit. ‘to marry oneself’) ‘to marry’, *lära sig* (lit. ‘to learn oneself’) ‘to learn’, a particle verb e.g. *bryta ner* (lit. ‘to break down’) ‘to subvert, to decompose’, *ta emot* (lit. ‘to take against’) ‘to accept’, a support verb construction e.g. *ta avstånd* (lit. ‘to take distance’) ‘to dissociate’, *fatta beslut* (lit. ‘to grab (a) decision’) ‘to come to a decision’, *vara på tok* (lit. ‘to be on mistake’) ‘to be wrong’, or some other type of verbal MWE. This final category includes more idiomatic expressions such as (sv) *syna något i sömmarna* (lit. ‘inspect something in the seams’) ‘check something carefully’ and (sv) *bli lång i synen* (lit. ‘become long in the sight’) ‘pull a long face’. In addition, both contiguous and non-contiguous MWEs can be classified as “other” and a comment can be added since some items could be difficult to annotate into these categories and it is important to be able to come back to any such items at a later point.

During initial annotation we included the categories of *idiom* and *fixed expression* which were removed before the final annotation. They proved to be too problematic to define in such a way that they did not overlap with each other or

⁶Cf. *continuous* and *non-continuous MWEs* in this volume. Since we are presenting our taxonomy here, we need to use the terminology we have chosen there. The terminology we chose for our taxonomy is also what is used in our online resource.

with other categories such as interjections. What one annotator saw as a fixed expression was sometimes seen as an interjection by the other, e.g. for greetings.

There are also some other distinctions which can sometimes be a bit problematic, e.g. (sv) *nybakat bröd* (lit. ‘newly baked bread’) ‘fresh bread’ is categorised as an institutionalised phrase, while (sv) *dagligt bröd* (lit. ‘daily bread’) ‘daily income’ is categorised as a nominal MWE. In a religious context, which is probably the most common context for the expression, this could also be considered an “institutionalised phrase”, but, this is not really the sense of institutionalised phrases in our taxonomy.

We discussed making further divisions according to transparency and compositionality, and experimented with annotating compositionality on a scale from 0–100. However, rating the compositionality (or transparency, see further Section 4.2) proved very difficult to do in a systematic way for the annotators, and their annotations indicated that they might have interpreted the concept of compositionality differently.

4 Materials and methods

In this section we introduce the corpora which we have used for this study (Section 4.1), and how we have automatically identified and manually annotated MWEs in them (Section 4.2). We then briefly describe how we have checked the automatic MWE annotation (Section 4.3). In Section 4.4 we explain how we have linked the MWEs to levels and also summarise a couple of studies where we have studied whether crowdsourcing could be used to link lexical items to proficiency levels. Since this chapter also demonstrates how the annotation can be accessed and used for further analysis of MWEs, we also introduce the graphical user interface of the Swedish L2 Profiles here (Section 4.5).

4.1 Corpora and Sen*Lex

We study the development of L2 Swedish based on two Swedish corpora: Coctail (Corpus of CEFR-based Textbooks as Input for Learner Levels’ modelling, Volodina et al. 2014), a corpus of coursebooks used in teaching L2 Swedish to adults, and the SweLL-pilot corpus (Swedish Learner Language Pilot corpus, Volodina et al. 2016a), a corpus of L2 Swedish essays. Coctail is used as a representation of receptive proficiency, however it can also be used as a proxy for common input at the different CEFR levels. The SweLL pilot is used to study the productive proficiency at different levels.

The corpora have been processed automatically using the Sparv pipeline (Borin et al. 2016), including tokenisation, lemmatisation, part-of-speech (POS)-tagging, dependency parsing, and word sense disambiguation. The pipeline also identifies MWEs during the process of lemmatisation based on a knowledge-based method that makes use of the Saldo lexicon (Borin et al. 2013). We use this automatic MWE annotation as a basis for our further manual annotation, i.e. only MWEs identified in this process are additionally annotated. We have also evaluated the success of the automatic annotation on a variety of texts which we use: coursebook texts and learner texts from different proficiency levels (see Section 4.3 and for results see Section 5.1).

The corpora have previously been used to derive two lexical resources aimed at language learners: (a) a CEFR-graded resource for Swedish as a second language, SVALex “SVenska som Andraspråk Lexikal resurs” (lit. ‘SWedish as a Second language Lexical resource’) (François et al. 2016) which is based on Coctail and which shows the expected receptive knowledge, and, (b) SweLLex (Swedish Learner language Lexicon, Volodina et al. 2016b) based on the SweLL-pilot corpus and which targets productive knowledge. In these lexical resources you can find the lemgrams (i.e. lemma + part of speech) of the words which occur in the corpora, but homographs are not separated if they have the same part of speech and the same inflectional paradigm. Hence, (sv) *val* ‘election’ and (sv) *val* ‘whale’ are different lemgrams since their inflectional paradigms are different: (1) *ett val*, *flera val* ‘one election, several elections’, (2) *en val*, *flera valar* ‘one whale, several whales’. But (sv) *gå* which can mean several different things including ‘walk’, ‘go’ or ‘be possible’ cannot be distinguished based only on lemgrams, therefore, word sense disambiguation is needed.

We recreated the lists with automatic word sense disambiguation, resulting in a list where each item includes lemma + POS + sense. The resulting lists are called SenSVALex and SenSweLLex, but are usually treated as one and referred to as *Sen*Lex*, cf. Alfter (2021: 31–32). *Sen*Lex* includes a total of 16 324 items (excluding some problematic cases but including MWEs). Any of these items which has been annotated as MWE are categorised manually according to MWE types (cf. Section 4.2).

4.2 Automatic and manual MWE annotation

Bearing in mind the limitations of Saldo as a knowledge base for the identification of MWEs (see Section 3), we nonetheless argue that the Saldo-based, i.e., knowledge-based, MWE identification is useful, objective and reliable (cf. the annotation check results in Section 5.1 and Volodina et al. 2022b). We therefore

explore the phraseological dimension of L2 vocabulary starting from the MWEs identified through the Sparv pipeline (Borin et al. 2016) and based on Saldo (Borin et al. 2013). Two annotators⁷ categorised the MWEs further into subcategories (cf. Section 3).

All automatically identified MWEs were added to a database for further manual annotation in Legato (Alfter et al. 2019), where only the types in our taxonomy could be selected. The manual annotation was done according to guidelines.⁸ An initial round of annotation was analysed and resulted in a modified taxonomy as presented in Section 3, as well as clarifications in the guidelines which the annotators made use of during the second and final round of annotation. The first author was available for supervision during annotations.

In the first round of the manual annotation our annotators were asked to indicate compositionality. This was excluded from the second round because it seems that our annotators understood the concept of compositionality differently. Previous research has also shown that compositionality is sometimes confused with transparency and vice versa (Cieślicka 2015, Nunberg et al. 1994). Instead, we decided to ask if contiguous MWEs were (morpho)syntactically modifiable or not in the final annotation.

Compositionality and transparency in relation to MWEs is definitely an area that requires further investigation, and a larger experiment with rankings of transparency and compositionality would be very interesting. However, this can only be done if a better way can be found of either defining or operationalising the concepts. Nevertheless, when there appeared to be a large difference between the ranking from the crowdsourcing experiment and the level of first occurrence in the coursebooks we decided to compare these results to the initial annotations from one of the annotators regarding the “compositionality” of the MWEs. We compared the relative rankings from the crowdsourcing experiment to the compositionality judgements. We did not use both since they seemed to interpret the task or concept differently.

4.3 Annotation check

As part of a more comprehensive annotation check (Volodina et al. 2022b) we also checked the annotation quality of MWEs. In this chapter we will summarise

⁷Both annotators are L1 speakers of Swedish. One has a MA in Scandinavian languages and one has a PhD in the same.

⁸<https://docs.google.com/document/d/1nZOKf-54FEkjIQFnPUmZZRWqib6y7gpCuKQO-XadeqM/edit?usp=sharing>

the parts about MWEs which have previously been published and discuss the results further.

The check was done by letting two annotators⁹ go through the automatic annotation of three texts per level (5 levels A1–C1) from three different sources: (1) the coursebook corpus, Coctail, (2) the original learner corpus, SweLL-pilot, and (3) the same as (2), only normalised which, among other things, standardised the spelling (cf. Rudebeck et al. 2021 for the normalisation guidelines, which we followed to facilitate comparisons).

Each annotator received a spreadsheet with the texts and their annotation in one tab per source (coursebooks, original learner texts, normalised learner essays respectively). Each annotation type was presented in a column of its own next to which a separate column was used for corrections. There was one token per row, which meant that MWEs spread over several rows. Next to the column of lemma, or MWE annotation, there was an extra column for corrections. The columns which should not be changed were locked so that only one or two of the researchers had access to them. The check was done according to a set of guidelines and under the supervision of the first author.

After the check was finished by both annotators a first comparison was run by one of the researchers, comparing the cells in the columns for corrections. Then a more qualitative check was done where the first author checked the changes that had been made to any items in relation to MWEs, i.e. additions or deletions of tokens from MWEs which had been identified, as well as if any MWEs had been noted as having been missed completely (undergeneration) or identified mistakenly (overgeneration) (see Section 5.1). An analysis of the whole check has been published in Volodina et al. (2022b).

4.4 Proficiency level assignment

CEFR levels are assigned based on the appearance of items in L2 Swedish coursebooks, as found in Coctail, and in L2 Swedish learner essays, as found in SweLL-pilot (cf. Section 2.3). Of course there are lexical items which do not occur among the words in the corpora and we therefore wanted to see how those could be linked to CEFR levels, but also how alternative ways of linking levels to MWEs would compare to the levels of lemmagrams which had already been graded based on coursebooks and learner essays. For this reason, we tried to rank MWEs through crowdsourcing by experts and non-experts, as well as by direct labelling by experts (Alfter et al. 2021). These results are not available in the Swedish L2

⁹Both annotators have an MA in Scandinavian languages. One is an L1 speaker of Swedish and the other an advanced L2 speaker.

Profiles but they are partly presented in Section 5.2. The crowdsourcing experiment meant that participants were asked to say which out of four MWEs they judged to be the easiest, and which they judged as the most difficult. We sorted the MWEs into three groups which were presented in separate experiments:

1. Group 1: Interjections, fixed expressions and idioms;
2. Group 2: Verbal MWEs and
3. Group 3: Adverbial, adjectival and non-lexical MWEs.

In parallel with the crowdsourcing experiment, we also asked three L2 Swedish professionals who had good knowledge of CEFR to first go through all of the crowdsourcing tasks, and then perform an explicit ranking assignment in a spreadsheet, where they had to assign CEFR levels from A1 to C2+ to each MWE which was part of the crowdsourcing experiment (Alfter et al. 2021).

Apart from the quantitative analysis in Alfter et al. (2021), we have analysed some of the results qualitatively in a previous publication (Lindström Tiedemann et al. 2022) and some of the main results from the latter are summarised in Section 5.2.1. This analysis was based on the MWEs which had been ranked as easiest and hardest in the different participant groups and included the seven easiest and the seven most difficult items from each type of MWEs and from each participant group. The MWEs were compared qualitatively across groups, and also in relation to the fact that if all groups had picked the same items as the seven easiest there would be 21 in total (7×3 MWE types). The results were also compared to coursebook occurrences, and newspapers and blogs to some extent, as well as to the direct ranking results.

In this chapter we present a further qualitative analysis of some of the results from the crowdsourcing experiment. This will help us gain a better understanding of why some expressions seem to be ranked very differently in the coursebook rankings in comparison to the general implicit rankings by the crowdsourcers. We examine the items in group 1: interjections etc., from the crowdsourcing experiment. This was the group with the best correlations in the crowdsourcing experiment, but still some results were a bit surprising in comparison to the first occurrence in coursebooks, and it would be interesting to see if those have anything in common.

Since we originally picked twelve items per level from the coursebooks, and since the crowdsourcing experiment results in a continuum from 1–60 we make a naive working assumption that 12 steps equal one level, even though this certainly does not have to be the case. We do this just as a way of deciding on a

selection principle for the items to look at more closely. We therefore focus on items that were one level away from the level they were chosen to represent, based on the occurrences in the coursebooks, or more than one level. We check if the same expression occurs in more than one coursebook aimed at the same level, to see if they could be seen as *core items* that several books considered important to include at the same or adjacent levels (Volodina et al. 2022a).

4.5 MWEs in the Swedish L2 Profiles

Based on the manual annotation described in Section 4.2 above, MWEs can now be accessed and filtered in different ways in the lexical profile within the Swedish L2 Profiles. There we provide lists of MWEs which appear in coursebooks for L2 Swedish learners (the Coctail corpus) or in learner essays (the SweLL-pilot corpus) including information about the proficiency levels where they appear in coursebooks (presented as receptive) and learner essays (presented as productive). The information can be filtered according to part-of-speech, MWE type, or CEFR level.

The profile includes absolute and relative frequencies and links to Korp (Borin et al. 2012) at Språkbanken Text where the corpus evidence can be consulted.¹⁰ This is what we use for our analysis of how MWEs occur in the data Section 5.3, and it is openly available to other researchers and teachers who wish to explore the resource.

5 Analysis and results

In this section we first present the results of our annotation check where we wanted to see how well the Sparv-pipeline identifies MWEs in both coursebooks and learner texts (Section 5.1). This includes a discussion of our results in Volodina et al. (2022b). In Section 5.2 we analyse different ways of assessing the proficiency level which should be associated with different MWEs. Finally, we analyse the distribution of MWEs across proficiency levels based on their occurrence in coursebooks and learner essays as presented in the Swedish L2 Profiles (Section 5.3).

5.1 Quality of the automatic annotation of MWEs

We focus on lexicalised MWEs such as verbal MWEs (e.g. particle verbs and reflexive verbs), greetings, multiword prepositions. These might not occur that of-

¹⁰The productive data requires a licence to access the actual texts.

ten in texts and they can be non-contiguous or allow some variation which can complicate their automatic recognition. For this reason we have opted to use a knowledge-based approach to MWE which we presented above (Section 4.2) and which we present an evaluation of here based on Volodina et al. (2022b), and here we also discuss these results further.

The automatic MWE annotation works best on coursebook data which has been written by L1 speakers. It also works better on the normalised L2 data than on the original L2 data, indicating that original L2 data is a bit more problematic as expected. There is no clear correlation between issues in the annotation check and certain proficiency levels in either of the datasets, instead the precision and recall seem to vary idiosyncratically, but we suspect it may be related to genre, topic and task type (Volodina et al. 2022b).

Overall, the MWE annotation works fairly well: 7–8 out of 10 MWEs were correctly annotated as MWEs, 2–3 were missed and some items were labelled as MWEs even though they were not MWEs, or they were only partially recognised (Volodina et al. 2022b: 158). In 45% of the missed cases it turns out that the MWE is also missing in the Saldo lexicon (Volodina et al. 2022b) which, as mentioned above, was one of the weaknesses we expected to see when using a knowledge-based MWE annotation system. Still, since MWEs are mostly well annotated and since the MWEs which are included in Saldo are likely to be the most well-known in Swedish, we believe this is a good result. It is clear that this could be improved by adding more items to Saldo, e.g. based on the items we have found to be missing. Nevertheless, it is clear that some MWEs which are listed in Saldo were also missed in the annotation, and it would be a good idea to look at these instances in more detail in the future to see if the annotation could somehow be improved, and maybe such work could also contribute to our understanding of MWEs.

Checking MWE annotation seems to be cognitively more difficult than checking lemmas, POS etc. (Volodina et al. 2022b). Furthermore, there are several MWEs that include a placeholder and such MWEs are not yet fully part of Saldo even though similar constructions have been studied extensively in relation to the Swedish Constructicon (see e.g. Sköldberg et al. 2013, Lyngfelt et al. 2018). This is related to the difficulty of annotating such instances automatically and requires further research.

Our assistants who checked the annotation agreed quite well in the MWE check. Krippendorff's alpha (Krippendorff 1980) show inter-annotator agreement at 0.85 for Coctail, 0.74 for original learner essays and 0.89 for the normalised learner essays, the highest value (Volodina et al. 2022b). This could possibly be because the normalised version of the learner essays was checked more closely in time by the two assistants and hence the discussions with the supervising

researcher might have been more similar in relation to this set. The annotation of Coctail and SweLL original had been checked quite a long time before then by assistant 1 and the check had resulted in some discussions about annotation and annotation check practices. This also partly meant that assistant 1 was allowed to go back and change her annotations to some extent and the guidelines were clarified for assistant 2 on these accounts.

Differences in the MWE check show that the assistants partly disagreed on what a MWE is, or differed in how good they were at spotting certain types of MWEs: one recognising grammatical MWEs such as (sv) *trots att* ‘even though’ more easily, and the other recognising complex verbs such as (sv) *få barn* (lit. ‘receive (a) child/children’) ‘have a child/children’ more easily (Volodina et al. 2022b). This could also partly have been affected by the fact that after assistant 1 started checking the data we saw that many MWEs were seen as missing a preposition (cf. Section 3) which led to discussions with Saldo staff who explained that prepositions are usually not seen as part of the MWE, but rather as part of the valency of the MWE and therefore are not listed in Saldo. Assistant 2 received this information before the check and hence could bear it in mind from the start.

5.2 MWE and proficiency levels

The results of the crowdsourcing experiment show that non-experts and experts rank MWEs very similarly, whereas direct ranking seems to be difficult and show less agreement between the annotators and also show some disagreement with the crowdsourced relative rankings by the same expert (Alfter et al. 2021). Unfortunately, since we chose to use relative judgement in the crowdsourcing experiment we have not yet found a way of directly linking items to a particular proficiency level. Instead the results are on a continuum with no indication of precise levels.

The fact that explicit level assignments seems to be more difficult and less consistent is something which we have concluded correlates well with previous studies on assessment of proficiency, although most of those studies have been on data consisting of full texts rather than decontextualised expressions which are bound to be even more difficult to link to an explicit level (Alfter et al. 2021).

5.2.1 The easiest and the most difficult items

In a previous qualitative analysis of the items which were ranked as the easiest seven or the most difficult seven in each of the three sets (i.e. interjections, verbal, adverbial) and for all three groups (i.e. learners, teachers, experts) we found

that the crowd participant groups agreed fairly well, but they agreed more on the easiest words (Lindström Tiedemann et al. 2022). This is hardly surprising since the topics are also more clearly defined in relation to the lower CEFR levels than the higher levels which are more associated with professions and special interests.

There were 28 MWEs in total which appeared among the easiest seven – instead of 21 which would have been the case if experts and non-experts had been in complete agreement (7 x 3 groups of MWEs). Thirteen (46.4%) of the easiest MWEs were among the easiest seven for both expert groups and for the non-experts. Five (17.9%) were among the easiest according to the L2 speakers (non-experts) and one of the expert groups. Hence there was partial agreement for eighteen MWEs (64.3%). The most difficult seven expressions were as many as 35 MWEs (compared to 21 which would have meant total agreement on the seven expressions, but not necessarily on their order). Only nine (25.7%) were the same in all three groups, and an additional seven (20%) showed agreement between the L2 speakers and one of the expert groups. That is, there was only 45.7% partial agreement among the most difficult items.

Comparing these results to the rankings based on the coursebooks showed that 18 (85.7%) of the easiest MWEs picked by the L2 speakers appeared at A1 level in coursebooks, somewhat less of the items picked by the experts: 76.2% of the L2 teachers' and 71.4% of the CEFR experts' (Lindström Tiedemann et al. 2022). Hence the L2 speakers' relative judgement was in a sense more in line with coursebook rankings for the easiest expressions. The most difficult expressions were harder to compare to coursebook occurrences. L2 speakers had nine (43%) MWEs from C1 among their most difficult seven, L2 teachers eleven (52%) and CEFR experts seven (33%).

Some of the MWEs were ranked as quite easy in comparison to the coursebook-based levels and sometimes it seemed as though this could be because the expressions were commonplace everyday expressions (Lindström Tiedemann et al. 2022). Prentice & Sköldberg (2013) claim that MWEs can be more common in informal genres. Some of the MWEs which were ranked as easy clearly do appear more in blog corpora than in newspaper corpora (Lindström Tiedemann et al. 2022) which could be a sign of this, but this should be investigated further. Other expressions which were ranked as easy were clearly more international expressions such as (sv) *logga in* 'to log in', which was seen as relatively easy by all three groups, but which only occurred in coursebook texts at C1 level, except for appearance in an exercise at A2 level. In addition, when ranking MWEs explicitly to levels, the three CEFR experts ranked this item as A1 or A2. This shows that it might be important to include all vocabulary from coursebooks, also from

exercises, and not only from readings texts, since some lexemes might only be used in exercises.

5.2.2 Comparing relative ranking to coursebook occurrences

Since the crowdsourced rankings are relative, they cannot easily be compared to CEFR levels. Still, it is interesting to try to do so by focusing on items which appear to have been ranked quite differently from the level which they were picked for. Since we picked 12 items per level we here focus on items which in the relative ranking ended up one level (eleven steps on a continuum from 1–60) or more, from their level based on their first occurrence in the coursebooks. Some of these MWEs do not even occur at adjacent levels in different coursebooks. In many cases where the discrepancy between the ‘levels’ was large we only have evidence from one book at the level of first occurrence. There are 20 cases (33%) where the ranking has a discrepancy of eleven or more steps on the scale from 1–60. Out of these 17 (85%) have a first occurrence based on one single book, and hence cannot be considered core items at that level, and this could be the reason why they have been ranked quite differently in the crowdsourcing experiment. In six (30%) cases we have examples from other levels, of which two (10%) include the level that was estimated based on the average implicit ranking by the crowd if we assume that the first 12 items on the continuum of 60 items equal A1, the second 12 A2, etc.¹¹

5.2.3 Comparing relative ranking to compositionality or transparency

Fourteen of the twenty cases which appeared to be ranked quite differently to the level which they were picked to represent based on coursebook occurrences had non-compositionality scores which were $\geq 50\%$ in the manual annotation. In fact, for as many as 13 it was $\geq 69\%$; and for nine it is as high as 98–100%. Therefore it is possible that these items were ranked as rather difficult in the crowdsourcing experiment due to their non-compositionality or opacity, since this meant that their meaning could not be guessed based on the combination of the constituent words. High non-compositionality meant that the items were more difficult than expected based on the first occurrence in coursebooks (if we estimate that 1–12 on the continuum from 1–60 should be equivalent to A1 and 48–60 to C1). The only two items (out of the 20) which were seen as opaque and which still received a

¹¹This was used as a naive principle to relate the continuum to levels as a means to clarify the differences and facilitate the analysis. However it is of course possible that all items in the relative ranking are seen as equivalent to A1 or C1 items by the participants.

‘lower’ implicit ranking than the coursebook projections were (sv) *spetsa öronen* (lit. ‘to sharpen the ears’) ‘prick up one’s ears’ and (sv) *många järn i elden* (lit. ‘(to have) many irons in the fire’) meaning ‘(to be) busy’. They were both classified as one level lower in the implicit ranking (if we assume that the order should be equivalent to 12 items per level).

Items that were ranked as easier by the expert and non-expert crowds were often classified as quite compositional (30–69, average 44). And conversely items which were classified as more difficult were seen as relatively non-compositional by the same annotator (74–100, average 88). To conclude, it could be that receptive levels of MWEs can often be correlated to compositionality or transparency, however it is unclear which it was that annotator focused on.

5.2.4 Comparing relative ranking to L1 reference corpora

There was one item in group 1 that the non-expert crowd and the expert crowd agreed was most/second most difficult, namely (sv) *på pin kiv* ‘just to tease’. This item was classified as C1 based on the coursebook texts. It appears three times, but only in one of the books on C1 level.

På pin kiv is a rare expression and 100% non-compositional or opaque according to our annotation. It occurs only rarely 0.1/1 million tokens (25 actual occurrences) in the Swedish newspaper corpus *Göteborgsposten (GP) 1994–2013*, and the same, 0.1/1 million (2 actual occurrences) in the Finland-Swedish newspaper corpus *Hufvudstadsbladet*. It does not occur at all in Finland-Swedish blogs (corpus *Bloggtexter*), but it does occur at the same relative frequency as in the newspaper corpora, 0.1 (35 times), in Swedish-Swedish blogs (corpus *Bloggmix 2006–2013*).¹²

The constituent (sv) *pin* is a rare adjective which only appears in this MWE and hence works as a good immediate idiom key. There are homonyms: (sv) *pin* (noun) which is only used in the Swedish proverb *vill man vara fin får man lida pin* (lit. ‘he/she who wants to look nice must suffer pain’) ‘good looks hurt’ where *pin* is interpreted as a form of pain, and (sv) *pin* (adverb) meaning ‘to the highest level’ used in the compound, e.g. (sv) *pinfärs*. *Kiv*, similarly, is quite rare and its only meaning in one of the most authoritative Swedish dictionaries (Svenska Akademien 2021) is “minor disagreement”. Based on both the opaqueness of the constituents of the MWE and the rare occurrences in L1 corpora it is quite understandable that this was ranked as among the most difficult items for L2 learners by experts and non-experts alike.

¹² All corpora were accessed through Korp at <https://spraakbanken.gu.se/korp/>

Interestingly, although not that surprising, the two easiest items in group 1 were also ranked the same in all three crowd participant groups: (sv) *god morgon* ‘good morning’ and (sv) *god natt* ‘good night’. Furthermore, these items were classified as 45–46% non-compositional and they are fairly transparent and also very common greetings. Greetings are among the lemmas that are usually taught explicitly at the beginning of language courses.

5.3 Frequency and MWE types

MWEs make up 4.9–9.4% of the sense-based lemgrams from the receptive corpus and 1.4–9.4% from the productive corpus (cf. Figure 2). The percentage is higher in the receptive data all the way up to C1 level. However, the difference decreases steadily. The percentage is thus much lower than the c. 50% which Jackendoff (1997) estimated that MWEs should make up of our mental lexicon. However, our estimates are higher than Borin (2021)’s for all of the Saldo lexicon, which he claimed consisted of 6% MWEs.

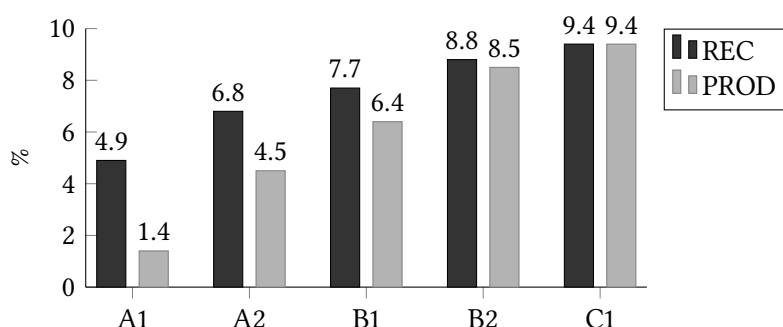


Figure 2: The percentage of MWEs on our sense-based list of lemgrams from Coctail (REC) and SweLL-pilot (PROD)

The fact that data produced by L2 learners reach the same percentage of MWEs in the texts at C1 level as the percentage used in coursebooks at the same level seems very encouraging. It would be good to compare this with other corpora, but to do so in a reasonable way we would need to extract a list of types consisting of lemgram + sense in the same way as here and that will have to be left for future work. In addition, we should try to find a way to include MWEs which are partly non-normlike in the L2 data and which therefore have not been annotated by the pipeline, e.g. due to a mistaken form or non-normlike word order or even a word that is the wrong word in the context but synonymous. To study MWE acquisition we need to find ways to compare such occurrences to the rest; but, it

seems they can only be captured manually unless they are consistent ‘mistakes’ which many learners tend to make.

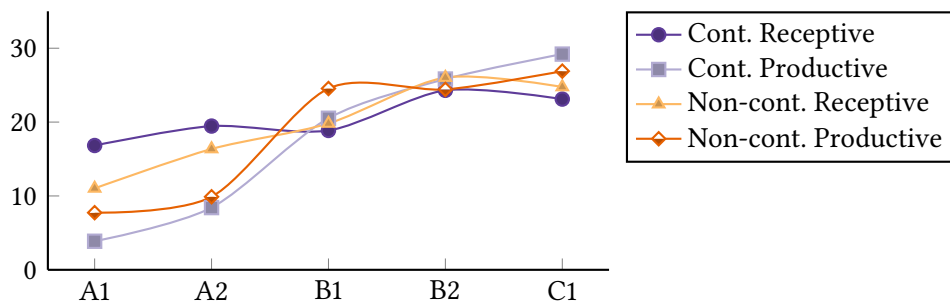


Figure 3: Development of contiguous and non-contiguous MWEs over CEFR proficiency levels in both coursebook data (receptive) and learner data (productive). The frequencies are based on the number of types in each category, not the frequency of occurrence of each type.

Contiguous and non-contiguous MWEs show similar development over the CEFR levels as shown in Figure 3. There is a clear increase in both broad types of MWEs, and from B1 level contiguous MWEs have very similar levels in receptive and productive data. Surprisingly, there are more non-contiguous MWE lemmas in productive learner data from B1 than in the coursebook data. But from B2 the levels are similar in production too. The higher B1 level seems likely to be due to a task effect (cf. Caines & Buttery 2017). Of course we know that in the productive data we are only catching items which have been correctly spelled and might therefore miss some instances which are reasonably normlike and which are attempts to produce MWEs which learners are starting to grasp. It is also possible that some MWEs have been used in the wrong context or in an unidiomatic way. For instance, the preposition might be wrong, since prepositions are not usually included in the MWE in Saldo. This means that a non-standard preposition will not cause the MWE to go unnoticed by the system. This makes it even more interesting that at B1 level there were more non-contiguous MWE lemmas in the learner data than in coursebooks. Including less, in the MWE itself, proves to be a possible advantage when working with learner texts. If MWEs can be identified even when they are not used in a normlike manner, that facilitates a closer analysis of how normlike the usage is.

Zooming in on the subcategories among the contiguous MWEs (Figures 4 & 5), adverbial MWEs stand out at all levels in both the receptive and the productive data. It would be interesting to compare this to other genres. It is also striking that at A1, in the productive data, the only contiguous MWE type is contiguous

adverbial MWEs, but at A2 there are already several others. The only subcategory among the contiguous MWEs which is still missing in the productive A2 data is proper nouns and this could possibly be related to pseudonymisation of the data, although it seems unlikely in this case since the MWE proper nouns that are included in Saldo are generally famous people, companies, organisations etc. This means that it is more likely to be the tasks at A2 that are restricting the use of MWE proper nouns. The category is still missing at B1 and is quite rare at B2 and very rare at C1.

Interjections are quite rare in productive data and there is a decrease in the use of MWE interjections in the receptive data. Still, it is interesting that there are MWE interjections at all levels in the coursebooks, even at C1 level. It would be interesting to have a closer look at the interjections that appear and in which text types they are used in the coursebooks. One would expect that they would only be used in dialogues, but considering the fairly high number of lemmas at more advanced levels when dialogues are more rare this requires further analysis. The fact that they are rare in learner language is most likely due to the tasks.

All non-contiguous MWEs in our data are verbal (Figure 6). The most common subcategory by far is particle verbs in both receptive and productive data. The frequency clearly increases from A1 to B2 in the receptive data and then drops a bit at C1. In the productive data there is a drop at A2. Otherwise there is a steady increase, so it seems likely that the drop at A2 is task-related or a sign that they are starting to really produce MWEs which they know and not whole-phrase constructions. In the productive data there are no instances of the categories *other* or *reflexive verbs* at A1 and interestingly there are no support verb constructions at C1, this may well also be task related.

Reflexive verbs start to appear at A2 in the learner data and show a clear increase at B1 and then remain quite stable. After this the levels are similar to the coursebook data possibly indicating that a ceiling has been reached. The number of lemmas are then slightly higher than in the coursebooks, but this is probably due to the topics covered in the respective texts. Reflexive verbs are quite common in different languages. However, the verbs which are reflexive differ even between closely related languages (Enström 1990: 41) (e.g. (sv) *lära sig* (lit. 'teach oneself') 'to learn'). In fact, they differ even between varieties of the same language, e.g. (sv) *ändra sig* (lit. 'change oneself') 'change one's mind' is not necessarily reflexive in Swedish as spoken in Finland (cf. af Hällström & Reuter 2008). Björklund (2007) similarly mentions differences in the usage of e.g. (sv) *köpa sig* (lit. 'to buy oneself') 'to buy' and (sv) *köpa* 'to buy' in Sweden and Finland, which however could be due to the small size of her data.

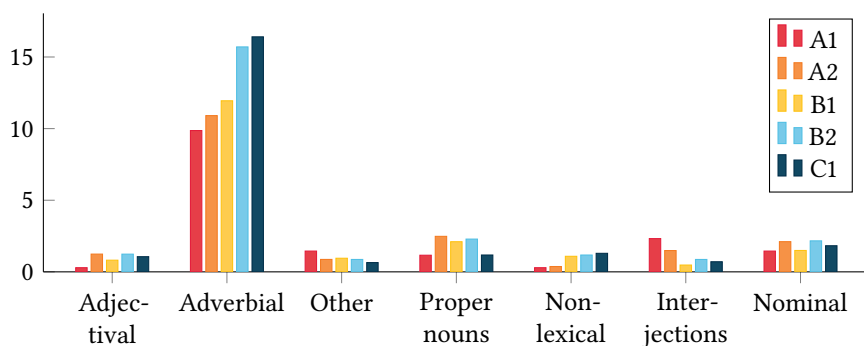


Figure 4: Contiguous MWE types in coursebook data (receptive) over CEFR levels per 10 000 tokens, i.e. the frequency of occurrence (cf. tokens) of the different types (cf. lemmas) is not considered here.

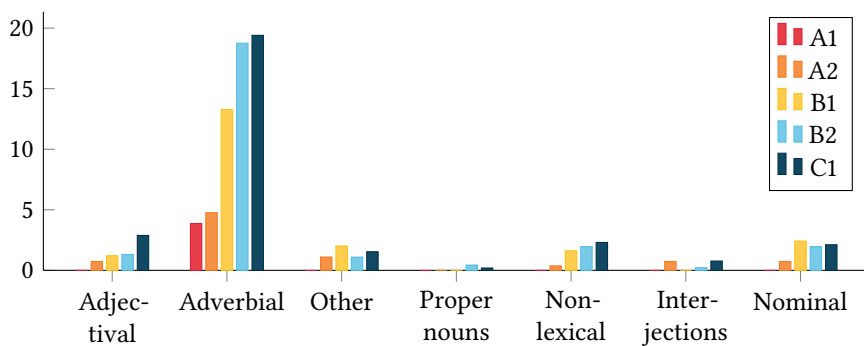


Figure 5: Contiguous MWE types in learner data (productive) over CEFR levels per 10 000 tokens, i.e. the frequency of occurrence (cf. tokens) of the different types (cf. lemmas) is not considered here.

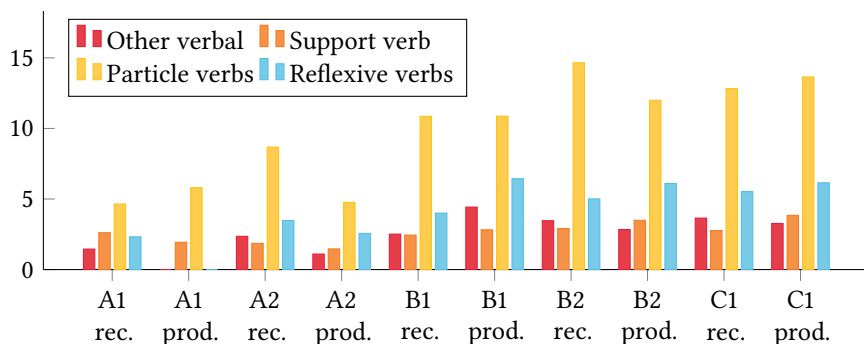


Figure 6: Non-contiguous MWE types in relation to 10 000 tokens, i.e. the frequency of occurrence (tokens) of the different types (cf. lemmas) is not considered here.

It is quite possible that there may be interference from previously known languages which cause difficulty in learning which verbs are reflexive. It is likely that this is why it is quite common that the reflexive pronoun is left out. Furthermore, in relation to usage-based theories of language acquisition how easy it is to learn that a verb is reflexive or not could depend on which variety of the target language you have been in contact with more, if there is indeed variation with regards to whether the verb is reflexive in different varieties. Nevertheless, previous research has shown that there were not that many mistakes in connection with Swedish reflexive verbs, but they seemed to be underused both on the type and the token level in a comparison between L1 and advanced L2 writers (Enström 1990: 93–94).

The fact that Enström (1990) found that there were not many actual mistakes in comparison with the standard regarding how the reflexive verbs were used means that a quantitative comparison based on the reflexive MWEs is likely to give quite a correct image of the use in both coursebooks and learner texts. It is particularly interesting to see that the frequency of reflexive verbs at B1–C1 in our data is higher in the learner essays than in coursebooks. Quite possibly this is related to the essay topics, but it should be investigated further.

6 Conclusions and future work

The Swedish L2 Profiles provides access to Swedish MWEs sorted by CEFR levels and with a possibility of comparing the usage in receptive and productive data by frequency as well as in context. This makes it a versatile resource e.g. for researchers and teachers, and it is clearly different from the English Profile and the Pearson Toolkit. Not only do MWEs occur in one place with a possibility to sort by types, but it also gives open access to frequencies and all of the empirical corpus evidence and presents *productive* and *receptive* overviews side by side.

In this chapter we have summarised how we have ascertained that the knowledge-based automatic MWE annotation works well enough for future studies of L2 Swedish. We have also presented our MWE taxonomy where we have tried to find an appropriate way of building on previous research and making it useful also in connection to automatic annotation. The first round of manual annotation showed that some MWE types can be difficult to keep apart, but after adjusting our taxonomy the two annotators agreed quite well.

Our attempts to find new ways of linking MWEs to receptive proficiency levels proved to work well only as relative measurements, not in relation to discrete levels (Alfter et al. 2021). The best agreement seems to be with regards to ranking

easy MWEs rather than difficult ones. This seems to have support in the CEFR documentation (Council of Europe 2001), since the beginner levels have a very clear focus on certain topics and themes, whereas advanced levels are more varied because they relate to different areas of expertise (cf. Lindström Tiedemann et al. 2022).

In this chapter, we have extended our analysis of the crowdsourcing results through a qualitative analysis of items which were ranked differently than their coursebook first level of occurrence. The results indicate a possible tendency for items which occur in only one book at its first level of occurrence to be ranked less reliably based on coursebooks. This could be because we are investigating MWEs rather than single items, which previous work has focused on. Future research, should continue to investigate different ways of linking items to levels and make sure to investigate both single words and MWEs. If empirical data is used as a basis, rather than e.g. crowdsourcing, both receptive and productive data should be investigated.

Compositionality (or transparency) proved to be problematic since the annotators did not appear to have interpreted the task in the same way. Still, by using the compositionality score from *one* of the annotators, we explore if there might be a tendency that relative rankings are linked to compositionality or transparency. Our results indicate that this might be the case, at least with regards to whether interjections or fixed expressions (group 1 in the crowdsourcing experiment) are interpreted as easier or more difficult than the first level of occurrence in coursebooks. This should be investigated further but it requires better ways of making sure whether it is transparency or compositionality that is estimated.

Some of the crowd rankings were also compared to occurrences in newspapers and blogs where the most difficult idiom was found to be rare in all of the corpora. Some further comparisons of this kind were done in a previous qualitative analysis (Lindström Tiedemann et al. 2022), see also Section 5.2.1. This showed that some MWEs may be more common in informal genres such as blogs, as claimed in Prentice & Sköldberg (2013).

We see tendencies for MWE usage to depend on many different factors including the tasks and genres which we compare. This means that we need resources such as the Swedish L2 Profiles which facilitates the comparison of texts aimed at learners and texts written by learners, and which also provides a possibility of easily comparing data to other corpora. We also have to remember that learner data is often quite small and often consists of fairly short texts (cf. Forsberg & Bartning 2010) which can complicate analysis and in particular comparisons with L1 production which often consists of longer texts and much larger corpora.

Making use of the data from the Swedish L2 Profiles categorised into MWE types, we see a clear development for MWEs in both receptive and productive data. The percentage of MWEs among the lemmagrams+sense types might possibly be an indicator of the productive proficiency level since there is a very clear increase per level and it is only at C1 that this is the same as in coursebooks at the group level, but also at B2 it is very close. Some subcategories show a clearer development (e.g. reflexive verbs, adverbial MWEs) whereas others seem fairly stable (e.g. support verbs, other contiguous MWEs such as idioms), and occasionally there is an indication of possible overuse. Future studies should try to investigate this further while restricting the genre and topic to match as much as possible, but also by investigating MWEs per text in both receptive and productive data in relation to the CEFR levels. However, this is complicated by the fact that learner texts tend to be rather short. This means that it is unlikely that several MWEs will be used in one text.

Acknowledgments

We would like to thank Riksbankens Jubileumsfond who financed most of the work presented in this paper through grant P17-0716:1. Some parts were also financed by the University of Helsinki and by Språkbanken Text, the University of Gothenburg.

We also wish to thank the editors and the anonymous reviewers for their valuable feedback on earlier versions of this chapter.

Abbreviations

CEFR	Common European Framework of Reference for Languages
GSE	Global Scale of English
L1	first language
L2	second language
MI	mutual information
MWE	multiword expression
NLP	natural language processing
NP	noun phrase
POS	part-of-speech
SAG	The Swedish Academy Grammar (Teleman et al. 1999)
SAOL	The Swedish Academy Glossary (Svenska Akademien 2015)

References

- Abrahamsson, Niclas & Kenneth Hyltenstam. 2009. Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning* 59(2). 249–306.
- af Hällström, Charlotta & Mikael Reuter. 2008. *Finlandssvensk ordbok*. Electronic version of the 4th edition. Schildts förlag & Forskningscentralen för de inhemska språken. <https://kaino.kotus.fi/fsob/>.
- Agirre, Eneko, Izaskun Aldezabal & Eli Pociello. 2006. Lexicalization and multiword expressions in the Basque WordNet. In Petr Sojka, Key-Sun Choi, Christine Fellbaum & Piek Vossen (eds.), *Proceedings of third international WordNet conference*, 131–138.
- Alfter, David. 2021. *Exploring natural language processing for single-word and multi-word lexical complexity from a second language learner perspective*. University of Gothenburg. (Doctoral dissertation).
- Alfter, David, Yuri Bizzoni, Anders Agebjörn, Elena Volodina & Ildikó Pilán. 2016. From distributions to labels: A lexical proficiency analysis using learner corpora. In Elena Volodina, Gintarė Grigonytė, Ildikó Pilán, Kristina Nilsson Björkenstam & Lars Borin (eds.), *Proceedings of the joint Workshop on NLP for Computer-Assisted Language Learning and NLP for Language Acquisition (NLP4CALL & NLP4LA)*, 1–7. Umeå: Linköping University Electronic Press. <https://aclanthology.org/W16-6501>.
- Alfter, David, Therese Lindström Tiedemann & Elena Volodina. 2019. LEGATO: A flexible lexicographic annotation tool. In Mareike Hartmann & Barbara Plank (eds.), *22nd Nordic Conference on Computational Linguistics (NoDaLiDa)*, 382–388. Linköping University Electronic Press.
- Alfter, David, Therese Lindström Tiedemann & Elena Volodina. 2021. Crowdsourcing relative rankings of multi-word expressions: Experts versus non-experts. *Northern European Journal of Language Technology* 7.
- Anward, Jan & Per Linell. 1976. Om lexikaliserade fraser i svenskan. *Nysvenska Studier* 55–56. 77–119.
- Attia, Mohammed, Antonio Toral, Lamia Tounsi, Pavel Pecina & Josef Van Genabith. 2010. Automatic extraction of Arabic multiword expressions. In Éric Laporte, Preslav Nakov, Carlos Ramisch & Aline Villavicencio (eds.), *Proceedings of the 2010 workshop on multiword expressions: From theory to applications*, 19–27. Coling 2010 Organizing Committee.
- Baldwin, Timothy & Francis Bond. 2002. Multiword expressions: Some problems for Japanese NLP. In *Proceedings of the 8th annual meeting of the Association for Natural Language Processing*, 379–382. Keihanna, Japan.

- Bhalla, Vishal & Klara Klimcikova. 2019. Evaluation of automatic collocation extraction methods for language learning. In Helen Yannakoudakis, Ekaterina Kochmar, Claudia Leacock, Nitin Madnani, Ildikó Pilán & Torsten Zesch (eds.), *Proceedings of the 14th workshop on innovative use of NLP for building educational applications*, 264–274. Association of Computational Linguistics.
- Björklund, Siv. 2007. Reflexiva verb och infinitivfraser. *Språkbruk*. <https://www.sprakbruk.fi/-/reflexiva-verb-och-infinitivfraser>.
- Boers, Frank, June Eyckmans, Jenny Kappel, Hélène Stengers & Murielle Demecheleer. 2006. Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research* 10(3). 245–261.
- Borin, Lars. 2021. Multiword expressions: A tough typological nut for Swedish *framenet++*. In Dana Dannélls, Lars Borin & Karin Friberg Heppin (eds.), *The Swedish FrameNet++: harmonization, integration, method development and practical language technology applications*, 221–260. Amsterdam/Philadelphia: John Benjamins.
- Borin, Lars, Markus Forsberg, Martin Hammarstedt, Dan Rosén, Roland Schäfer & Anne Schumacher. 2016. Sparv: Språkbanken's corpus annotation pipeline infrastructure. In *The sixth Swedish language technology conference (SLTC)*. Umeå University.
- Borin, Lars, Markus Forsberg & Lennart Lönngren. 2013. SALDO: A touch of yin to WordNet's yang. *Language Resources and Evaluation* 47. 1191–1211.
- Borin, Lars, Markus Forsberg & Johan Roxendal. 2012. Korp: The corpus infrastructure of Språkbanken. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerch, Mehmet Doğan Uğur, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the eighth international conference on Language Resources and Evaluation (LREC)*, 474–478. European Language Resources Association.
- Caines, Andrew & Paula Buttery. 2017. The effect of task and topic on opportunity of use in learner corpora. In Vaclav Brezina & Lynne Flowerdew (eds.), *Learner corpus research: New perspectives and applications*, 5–27. Bloomsbury Publishing Academic London.
- Capel, Annette. 2012. Completing the English vocabulary profile: C1 and C2 vocabulary. *English Profile Journal* 3.
- Capel, Annette. 2015. The English vocabulary profile. *English profile in practice* 5. 9–27.
- Cieślicka, Anna B. 2015. Idiom acquisition and processing by second/foreign language learners. In Roberto R. Heredia & Anna B. Cieślicka (eds.), *Bilingual figurative language processing*, 288–244. Cambridge University Press.

- Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Modern Languages Division (Strasbourg) & Cambridge University Press.
- Council of Europe. 2009. *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR)*. Neus Figueras, Brian North, Sauli Takala, Piet Van Avermaet & Norman Verhelst (eds.). Council of Europe, Language policy division.
- Council of Europe. 2020. *Common European Framework of Reference for Languages: learning, teaching, assessment*. Council of Europe Publishing.
- Cowie, Anthony P. 1994. Phraseology. In R. E. Asher (ed.), *Encyclopedia of language and linguistics*, 3168–3171. Oxford: Pergamon.
- De Cock, Sylvie, Sylviane Granger, Geoffrey Leech & Tony McEnery. 2014. An automated approach to the phrasicon of EFL learners. In Sylviane Granger (ed.), *Learner English on computer*, 67–79. Routledge.
- de Caseli, Helena Medeiros, Carlos Ramisch, Maria das Graças Volpe Nunes & Aline Villavicencio. 2010. Alignment-based extraction of multiword expressions. *Language Resources and Evaluation* 44. 59–77.
- Dürlich, Luise & Thomas François. 2018. EFLLex: A graded lexical resource for learners of English as a foreign language. In Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis & Takenobu Tokunaga (eds.), *Proceedings of the eleventh international conference on Language Resources and Evaluation (LREC)*. European Language Resources Association (ELRA). <https://aclanthology.org/L18-1140>.
- Durrant, Phil. 2018. Formulaic language in English for academic purposes. In Anna Siyanova-Chanturia & Ana Pellicer-Sánchez (eds.), *Understanding formulaic language: A second language acquisition perspective*, 211–227. Routledge.
- Ekberg, Lena. 1999. Användningen av komplexa predikat hos invandrarbarn i Rosengård. In Lars-Gunnar Andersson, Aina Lundqvist, Kerstin Norén & Lena Rogström (eds.), *Svenskans beskrivning 23: förhandlingar vid tjugotredje Sammankomsten för svenskans beskrivning: Göteborg den 15–16 maj 1998*, 86–95. Lund: Lund University Press.
- Ekberg, Lena. 2013. Grammatik och lexikon hos svenska i andraspråk på nästan infödd nivå. In Kenneth Hyltenstam & Inger Lindberg (eds.), *Svenska som andraspråk*, 259–279. Lund: Studentlitteratur.
- Ellis, Nick C. 2012. Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics* 32. 17–44.

- Enström, Ingegerd. 1990. *Feltyper i invandrargymnasisters användning av partikelverb, prefixverb och reflexiva verb* (Rapporter från institutionen för nordiska språk/svenska (NORDRAPP) 4). Göteborg: Göteborgs universitet.
- Erman, Britt. 2007. Cognitive processes as evidence of the idiom principle. *International Journal of Corpus Linguistics* 12(1). 25–53.
- Erman, Britt & Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text & Talk* 20(1). 29–62.
- Forsberg, Fanny. 2006. *Le langage préfabriqué en français parlé L2: Étude acquisitionnelle et comparative*. Stockholm: Stockholm University. (Doctoral dissertation).
- Forsberg, Fanny. 2008. *Le langage préfabriqué: Formes, fonctions et fréquences en français parlé L2 et L1* (Contemporary Studies in Descriptive Linguistics 20). Oxford: Peter Lang.
- Forsberg, Fanny. 2010. Using conventional sequences in L2 French. *International Review of Applied Linguistics in Language Teaching* 48. 25–51.
- Forsberg, Fanny & Inge Bartning. 2010. Can linguistic features discriminate between the communicative CEFR-levels? A pilot study of written L2 French. In Inge Bartning, Maisa Martin & Ineke Vedder (eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (EuroSLA Monographs 1), 133–158. EuroSLA.
- François, Thomas, Núria Gala, Patrick Watrin & Cédric Fairon. 2014. FLELex: A graded lexical resource for French foreign learners. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the ninth international conference on Language Resources and Evaluation (LREC'14)*, 3766–3773. Reykjavik, Iceland: European Language Resources Association (ELRA). <https://aclanthology.org/L14-1>.
- François, Thomas, Elena Volodina, Ildikó Pilán & Anaïs Tack. 2016. SVALex: A CEFR-graded lexical resource for Swedish foreign and second language learners. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the tenth international conference on Language Resources and Evaluation (LREC)*, 213–219. <https://aclanthology.org/L16-1>.
- Gala, Núria, Thomas François, Delphine Bernhard & Cédric Fairon. 2014. Un modèle pour prédire la complexité lexicale et graduer les mots. In Phillipe Blache, Frédéric Béchet & Brigitte Bigi (eds.), *Proceedings of TALN 2014*, 91–102. Association pour le Traitement Automatique des Langues. <https://aclanthology.org/F14-2000>.

- Gala, Núria, Thomas François & Cédric Fairon. 2013. Towards a French lexicon with difficulty measures: NLP helping to bridge the gap between traditional dictionaries and specialized lexicons. In Iztok Kosem, Jelena Kallas, Polona Gantar, Simon Krek, Margit Langemets & Maria Tuulik (eds.), *Electronic lexicography in the 21st century: Thinking outside the paper. proceedings of the eLex 2013 conference*, 132–151. <http://eki.ee/elex2013/proceedings/eLex2013-proceedings.pdf>.
- Granger, Syviane. 2018. Formulaic sequences in learner corpora. In Anna Siyanova-Chanturia & Ana Pellicer-Sanchez (eds.), *Understanding formulaic language: A second language acquisition perspective*, 228–247. Routledge.
- Green, Anthony. 2012. *Language functions revisited: Theoretical and empirical bases for language construct definition across the ability range* (English Profile Studies 2). Cambridge, UK: Cambridge University Press.
- Hawkins, John A & Luna Filipović. 2012. *Criterial features in L2 English: Specifying the reference levels of the Common European Framework*, vol. 1. Cambridge University Press.
- Jackendoff, Ray. 1997. *The architecture of the language faculty* (Linguistic Inquiry Monographs 28). Cambridge, MA: MIT Press.
- Karlsson, Ola. 2017. *Svenska skrivregler*. Liber.
- Krippendorff, Klaus. 1980. *Content Analysis: An introduction to its methodology*. Beverly Hills, CA: Sage.
- Kurtes, Svetlana & Nick Saville. 2008. The English profile programme: An overview. *Research Notes* 33. 2–4.
- Leseva, Svetlozara, Verginica Barbu Mititelu, Ivelina Stoyanova & Mihaela Cristescu. 2024. A uniform multilingual approach to the description of multiword expressions. In Voula Giouli & Verginica Barbu Mititelu (eds.), *Multiword expressions in lexical resources: Linguistic, lexicographic, and computational perspectives*, 73–116. Berlin: Language Science Press. DOI: 10.5281/zenodo.10998635.
- Lewis, Margareta. 2008. *The idiom principle in L2 English: Assessing elusive formulaic sequences as indicators of idiomaticity, fluency, and proficiency*. Stockholm University. (Doctoral dissertation).
- Lindström Tiedemann, Therese, David Alfter & Elena Volodina. 2022. CEFR-nivåer och svenska flerordsuttryck. In Siv Björklund, Bodil Haagensen, Marianne Nordman & Anders Westerlund (eds.), *Svenskan i Finland 19: Föredrag vid den nittonde sammankomsten för beskrivningen av svenskan i Finland, Vasa den 6–7 maj 2021*, 218–233. Svensk-österbottniska samfundet.

- Lyngfelt, Benjamin, Linnéa Bäckström, Lars Borin, Anna Ehrlemark & Rudolf Rydstedt. 2018. Constructicography at work: Theory meets practice in the Swedish constructicon. In *Constructicography: Constructicon development across languages*, 41–106. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/cal.22.03lyn.
- Mel'čuk, Igor. 1998. Collocations and lexical functions. In A. P. Cowie (ed.), *Phraseology: Theory, analysis, and applications*, 23–53. Clarendon Press.
- Meunier, Fanny. 2012. Formulaic language and language teaching. *Annual Review of Applied Linguistics* 32. 111–129.
- Nesselhauf, Nadja. 2003. The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics* 24(2). 223–242.
- Nunberg, Geoffrey, Ivan A. Sag & Thomas Wasow. 1994. Idioms. *Language* 70(3). 491–538.
- O'Keeffe, Anne & Geraldine Mark. 2017. The English grammar profile of learner competence: Methodology and key findings. *International Journal of Corpus Linguistics* 22(4). 457–489.
- Paquot, Magali. 2019. The phraseological dimension in interlanguage complexity research. *Second Language Research* 35(1). 121–145.
- Pawley, Andrew & Frances Hodgetts Syder. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In Jack C. Richards & R. W. Schmidt (eds.), *Language and communication*, 203–239. Routledge.
- Piao, Scott Songlin, Paul Rayson, Dawn Archer & Tony McEnery. 2005. Comparing and combining a semantic tagger and a statistical tool for MWE extraction. *Computer Speech & Language* 19(4). 378–397.
- Prentice, Julia. 2010. *Käppen i hjulen: Behärskning av svenska konventionaliserade uttryck bland gymnasieelever med varierande språklig bakgrund* (ROSA 12). University of Gothenburg. <http://hdl.handle.net/2077/23261>.
- Prentice, Julia & Emma Sköldberg. 2013. Flerordsenheter – ur ett andraspråksperspektiv. In Kenneth Hyltenstam & Inger Lindberg (eds.), *Svenska som andraspråk: I forskning, undervisning och samhälle*, 2nd edn., 197–220. Studentlitteratur.
- Ringbom, Håkan. 2012. A country in focus: Review of recent applied linguistics research in Finland and Sweden, with specific reference to foreign language learning and teaching. *Language Teacher* 45(4). 490–514.
- Rudebeck, Lisa, Gunlög Sundberg & Mats Wirén. 2021. *SweLL normalization guidelines* (GU-ISS Research report series). University of Gothenburg. <http://hdl.handle.net/2077/69432>.

- Sag, Ivan A., Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In Alexander F. Gelbukh (ed.), *Proceedings of the third international conference on Intelligent Text Processing and Computational Linguistics (CICLing 2002)*, 1–15. Springer.
- Schmidt, Richard. 2012. Attention, awareness, and individual differences in language learning. In Wai Meng Chan, Kwee Nyet Chin, Sunil Kumar Bhatt & Izumi Walker (eds.), *Perspectives on individual characteristics and foreign language education* (SSFLE 6), 27–50. De Gruyter Mouton. DOI: 10.1515/9781614510932.27.
- Schulte im Walde, Sabine. 2024. Collecting and investigating features of compositionality ratings. In Voula Giouli & Verginica Barbu Mititelu (eds.), *Multiword expressions in lexical resources: Linguistic, lexicographic, and computational perspectives*, 269–308. Berlin: Language Science Press. DOI: 10.5281/zenodo.10998645.
- Shigeto, Yutaro, Ai Azuma, Sorami Hisamoto, Shuhei Kondo, Tomoya Kose, Keisuke Sakaguchi, Akifumi Yoshimoto, Frances Yung & Yuji Matsumoto. 2013. Construction of English MWE dictionary and its application to POS tagging. In Valia Kordoni, Carlos Ramisch & Aline Villavicencio (eds.), *Proceedings of the 9th workshop on multiword expressions (MWE 2013)*, 139–144. Atlanta, GA: Association for Computational Linguistics.
- Sköldberg, Emma, Linnéa Bäckström, Lars Borin, Markus Forsberg, Benjamin Lyngfelt, Leif-Jöran Olsson, Julia Prentice, Rudolf Rydstedt, Sofia Tingsell & Jonatan Uppström. 2013. Between grammars and dictionaries: A Swedish constructicon. In Iztok Kosem, Jelena Kallas, Polona Gantar, Simon Krek, Margit Langemets & Maria Tuulik (eds.), *Electronic lexicography in the 21st century: Thinking outside the paper. Proceedings of the eLex 2013 conference*, 310–327. http://eki.ee/elex2013/proceedings/eLex2013_21_Skoldberg+etal.pdf.
- Svenska Akademien. 2015. *Svenska Akademiens ordlista*. 14th edn. Svenska Akademien & Norstedts ordbok. <https://www.svenskaakademien.se/svenska-spraket/svenska-akademiens-ordlista-saol>.
- Svenska Akademien. 2021. *Svensk ordbok utgiven av Svenska Akademien*. 2nd edn. Svenska Akademien & Nordstedts ordbok. <https://www.svenskaakademien.se/svenska-spraket/svensk-ordbok-utgiven-av-svenska-akademien-so>.
- Tack, Anaïs, Thomas François, Piet Desmet & Cédric Fairon. 2018. NT2Lex: A CEFR-graded lexical resource for Dutch as a foreign language linked to open Dutch WordNet. In Joel Tetreault, Jill Burstein, Ekaterina Kochmar, Claudia Leacock & Helen Yannakoudakis (eds.), *Proceedings of the 13th workshop on innovative use of NLP for Building Educational Applications (BEA)*, 137–146. New

- Orleans, LA: Association for Computational Linguistics. <https://aclanthology.org/W18-05>.
- Teleman, Ulf, Staffan Hellberg, Erik Andersson & Lisa Christensen. 1999. *Svenska akademiens grammatik*. Norstedts ordbok & Svenska Akademien.
- Villada Moirón, María Begoña. 2005. *Data-driven identification of fixed expressions and their modifiability*. University of Groningen. (Doctoral dissertation).
- Volodina, Elena, David Alfter & Therese Lindström Tiedemann. 2022a. Crowdsourcing ratings for single lexical items: A core vocabulary perspective. *Slovenščina 2.0* 10(2). 5–61. DOI: 10.4312/slo2.0.2022.2.5-61.
- Volodina, Elena, David Alfter, Therese Lindström Tiedemann, Maisa Lauriala & Daniela Piipponen. 2022b. Reliability of automatic linguistic annotation: Native vs non-native texts. In Monica Monachini & Maria Eskevich (eds.), *Selected papers from the CLARIN Annual Conference 2021*, 151–167. Linköping Electronic Press. DOI: 10.3384/9789179294441.
- Volodina, Elena, Ildikó Pilán, Stian Rødven Eide & Hannes Heidarsson. 2014. You get what you annotate: A pedagogically annotated corpus of coursebooks for Swedish as a second language. In Elena Volodina, Lars Borin & Ildikó Pilán (eds.), *Proceedings of the 3rd workshop on NLP for computer-assisted language learning (NLP4CALL)*, 128–144. <https://aclanthology.org/W14-3500>.
- Volodina, Elena, Ildikó Pilán, Ingegerd Enström, Lorena Llozhi, Peter Lundkvist, Gunlög Sundberg & Monica Sandell. 2016a. Swell on the rise: Swedish learner language corpus for European reference level studies. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the tenth international conference on Language Resources and Evaluation (LREC'16)*, 206–212. Portorož, Slovenia: ACL. <https://aclanthology.org/L16-1>.
- Volodina, Elena, Ildikó Pilán, Lorena Llozhi, Baptiste Degryse & Thomas François. 2016b. SweLLex: Second language learners' productive vocabulary. In Elena Volodina, Gintarė Grigonytė, Ildikó Pilán, Kristina Nilsson Björkenstam & Lars Borin (eds.), *Proceedings of the joint Workshop on NLP for Computer-Assisted Language Learning and NLP for Language Acquisition (NLP4CALL & NLP4LA)*, 76–84. Umeå: Linköping Electronic Conference Proceedings. <https://aclanthology.org/W16-6500>.
- Watrín, Patrick & Thomas François. 2011. An n-gram frequency database reference to handle MWE extraction in NLP applications. In Valia Kordoni, Carlos Ramisch & Aline Villavicencio (eds.), *Proceedings of the workshop on multiword expressions: From parsing and generation to the real world*, 83–91. Portland, OR:

- Association for Computational Linguistics. <https://aclanthology.org/W11-0813>.
- Wray, Alison. 2002. *Formulaic language and the lexicon*. Cambridge University Press.
- Wray, Alison. 2013. Formulaic language. *Language teaching* 46(3). 316–334.
- Wray, Alison & Michael R. Perkins. 2000. The functions of formulaic language: An integrated model. *Language & Communication* 20(1). 1–28.
- Yamaguchi, Nami, David Alfter, Kaori Sugiyama & Thomas François. 2022. Towards a verb profile: Distribution of verbal tenses in FFL textbooks and in learner productions. In David Alfter, Elena Volodina, Thomas François, Piet Desmet, Frederik Cornillie, Arne Jönsson & Evelina Rennes (eds.), *Proceedings of the 11th workshop on NLP for computer assisted language learning*, 123–142. Louvain-la-Neuve, Belgium: LiU Electronic Press. <https://aclanthology.org/2022.nlp4call-1.13>.