Retrieving Occurrences of Grammatical Constructions

Anna Ehrlemark and **Richard Johansson** and **Benjamin Lyngfelt**
University of Gothenburg
ehrlemark@gmail.com, richard.johansson@gu.se
benjamin.lyngfelt@svenska.gu.se

Abstract

Finding authentic examples of *grammatical constructions* is central in constructionist approaches to linguistics, language processing, and second language learning. In this paper, we address this problem as an information retrieval (IR) task. To facilitate research in this area, we built a benchmark collection by annotating the occurrences of six constructions in a Swedish corpus. Furthermore, we implemented a simple and flexible retrieval system for finding construction occurrences, in which the user specifies a ranking function using lexical-semantic similarities (lexicon-based or distributional). The system was evaluated using standard IR metrics on the new benchmark, and we saw that lexical-semantical rerankers improve significantly over a purely surface-oriented system, but must be carefully tailored for each individual construction.

1 Introduction

Linguistic theories based on the notion of *constructions* are a recent development in linguistics, which has revitalized the discussion in fields such as syntax, lexicography, and argument structure theory. In particular, it is useful for describing *partially schematic* constructions: templatic patterns that exhibit lexical as well as syntactic properties, which are too specific to be referred to general grammar rules, but too general to be attributed to specific lexical units, which is a reason why they historically have been overlooked (Fillmore et al., 1988). These constructions are very frequent, and they are not only theoretically interesting but also important in language teaching (De Knoop and Gilquin, 2016), since they are challenging for second language learners (Prentice and Sköldberg, 2011).

So far, despite the importance of construction-based theory in linguistics and language pedagogy, it has seen limited adoption in natural language processing. A major impediment to their acceptance is probably the lack of resources of a nontrivial size. Efforts to build inventories of this kind – *constructicons* – are now underway for a number of languages. The emerging practice of constructicon development, or *constructicography*, can be seen as a combination of construction grammar and lexicography (Fillmore et al., 2012). A crucial requirement for constructicon building is access to corpus search tools allowing constructicographers to *search* for occurrences of constructions: either for finding prototypical examples (Gries, 2003), or for computing statistics such as word–construction cooccurrence (Stefanowitsch and Gries, 2003). Also, for the reasons mentioned above, finding authentic examples of uses of a particular construction is also useful in language teaching situations.

In this paper, we cast the task of searching for construction occurrences as an information retrieval (IR) problem. This is fruitful in construction-based research and constructicography for a number of reasons. First, while a partially schematic construction can have a surface form that is easy to describe as a structural pattern, its exact semantic properties and restrictions can be hard to capture as a clear-cut binary decision, or may not even be known *a priori*. This makes it natural to re-rank corpus hits according to a semantics-based scoring function. Such an approach allows a researcher to explore the spectrum of search hits, from the most prototypical – which a good retrieval system should place near the top – to the more unusual cases. Also, the IR perspective is natural in terms of interaction: a user can pose queries, rank and re-rank repeatedly according to different functions defined on the fly, to get a comprehensive overview of the various uses of a construction in a corpus.

To open up new opportunities for research in construction retrieval, we built a benchmark collection by annotating the occurrences of six different partially schematic constructions defined in the Swedish Constructicon (Sköldberg et al., 2013). This allows us to explore different retrieval systems and evaluate them according to standard IR evaluation protocols, and we developed a construction retrieval system that finds occurrences of partially schematic constructions in Swedish corpora, based on flexible user-defined ranking functions using lexicon-based and distributional similarities to describe the construction's slot fillers. These ranking functions are learned from a small number of seed examples. Our results show that ranking functions based on lexical-semantic properties are effective, outperforming a purely surface-based baseline for all six constructions. However, because the constructions are so structurally and semantically diverse, we cannot expect a single ranking function to be the best in all cases, and in practice we observed considerable difference among constructions as to which similarities are most useful. This shows that it is crucial for the ranking functions to be flexible and user-defined, which also makes most sense from a usability perspective.

2 Construction Grammar and the Swedish Constructicon

In terms of linguistic theory, the approach is couched in Construction Grammar, for which the aforementioned abundance of semi-general and partially schematic patterns is a key motivation, and where intermingling of linguistic levels is seen as the norm, not the exception. Rejecting the sharp distinction between lexicon and grammar, constructionists regard grammatical rules, lexical idiosyncrasies and "mixed" patterns alike as *constructions*: "conventional, learned form–function pairings at varying levels of complexity and abstraction" (Goldberg 2013, p17). A methodological benefit of assuming the same kind of unit across the board is that the machinery required to account for patterns with both grammatical and lexical properties can also handle those that are purely grammatical or purely lexical, another that the absence of distinct levels eliminates the problem of borderline cases (Fillmore et al., 1988).

Applying this view to descriptive practice, a constructicon is a collection of construction descriptions, a "dictionary of constructions." First introduced for English (Fillmore et al., 2012), there are now constructicon resources under development for a number of languages, including Swedish (Sköldberg et al., 2013; Lyngfelt et al., forthcoming), Brazilian Portuguese (Torrent et al., 2014), and Japanese (Ohara, 2013). The Swedish Constructicon currently covers around 400 constructions, many of which are partially schematic patterns of the kind that is hard to account for from a grammatical or lexical perspective alone. Hence, offering a wide and relevant selection of constructions, it provides a suitable testing ground for the study at hand.

To exemplify the organization of the Swedish Constructicon, Figure 1 shows the most important parts of its entry for V_REFL.RÖRELSE (REFLEXIVE_MOTION), a frequent construction in which motion is expressed using a verb with a reflexive pronoun. The entry contains a *definition* of the construction, that presents its structure and semantics textually, a *structure sketch* describing its surface structure in a semi-formal way, and a small number of annotated *examples* from corpora that show typical uses of the construction.

Name	V_REFL.RÖRELSE
Definition	[An actor]$_{\text{ACTOR}}$, expressed with a [reflexive]$_{\text{REFL}}$, [moves]$_{\text{V}}$ [in a direction, traverses a path, from a place or to a place]$_{\text{LOCATIVE}}$. The verb usually describes the manner of the motion. The construction also encompasses actions that indicate intended motion and non-motion.
Structure sketch	V PN$_{\text{REFL}}$ [PP | ADVP]
Example	[*Vi*]$_{\text{ACTOR}}$ *fick* [*armbåga*]$_{\text{V}}$ [*oss*]$_{\text{REFL}}$ [*fram*]$_{\text{LOCATIVE}}$ *i restaurangen*. [We]$_{\text{ACTOR}}$ had to [elbow]$_{\text{V}}$ [ourselves]$_{\text{REFL}}$ [forward]$_{\text{LOCATIVE}}$ in the restaurant.

Figure 1: Swedish Constructicon entry for V_REFL.RÖRELSE (REFLEXIVE_MOTION).

3 Related work

While we are aware of no previous work approaching the general problem of searching for construction occurrences from a retrieval perspective, quantitative and corpus-based methods have been a crucial part of the construction-linguistic toolbox from its early days (Gries, 2003; Stefanowitsch and Gries, 2003; Hilpert, 2013). In particular, a number of automatic methods have been presented that mine corpora for frequent patterns. For instance, Wible and Tsao (2010) collected generalized n-grams (that is, combinations of words, PoS tags and phrase labels) and applied standard measures of collocational strength to select n-grams that seem to be recurrent patterns. The Swedish constructicon project used similar methods (Forsberg et al., 2014), but also extended the approach by Wible and Tsao (2010) by considering patterns containing phrase labels (e.g. NP-*and*-NP, *as*-Adj-*as*-NP).

The only related work we found that treats the construction detection as a ranking task is by Dubremetz and Nivre (2015), who used a ranking approach to retrieve occurrences of the rare rhetorical *chiasmus* construction. They preferred ranking over classification since the complexity of this construction and the number of borderline cases made a hard classification infeasible, and similarly to our position, they argued that the existence of borderline cases should be embraced instead of ignored. However, since their work is limited to searching for just *chiasmi*, their system was completely tailored for this case.

4 A benchmark collection for evaluating construction retrieval systems

We built a new benchmark for evaluating construction retrieval systems based on six different constructions defined in the Swedish Constructicon. The selected constructions are nontrivial in the sense that their formal description cannot be translated into a standard corpus search query (based on surface features such as words or part-of-speech tags) that captures exactly their true instances – for instance, occurrences of the well-studied *let alone* construction (Fillmore et al., 1988) are easy to spot in corpora.

The initial pass of creating the benchmark was to extract a collection of potential instances for each of the six constructions. We extracted these instances by querying the *Korp* corpus search engine, hosted by the Swedish Language Bank (Borin et al., 2012). Korp stores a large collection of corpora, currently around 10 billion tokens, and allows structural queries based on surface strings but also several types of linguistic annotation; for this experiment, we used a corpus of contemporary fiction. The web service of Korp allows users to pose queries using the CQP language (Christ, 1994), and we selected the instances by translating the structure sketch (as in the example in Figure 1) of each construction into a corresponding CQP query. The corpus search engine then returns a (possibly very large) number of hits, and depending on the formal properties of the construction, fixed lexical content, variable tokens, or syntactic restrictions of particular slots, this list is 'contaminated' to some extent by unrelated hits.

Each collection of hits was then annotated manually as true or false instances of the construction in question. Table 1 shows the number of hits for each query, and the number of true hits among them.

Construction	Total hits	True hits
V_*av*_NP	501	169
PROPORTION_*i/om*	1703	205
REFLEXIVE_MOTION	1300	338
QUANTIFYING_GENITIVE:TIME	945	97
QUANTIFYING_GENITIVE:SCALE	945	166
BOUNDED_EVENT_*på*	2000	43

Table 1: Statistics for the benchmark.

In the following, we detail the six constructions used to create the benchmark.[1]

4.1 V_*av*_NP

V_*av*_NP is a causal VP construction where the event or state expressed by the verb is caused by the bare noun following the preposition *av* 'of'. It includes both literal and metaphorical causative relations such as *stinka av mögel* 'stink of mold' and *dö av skam* 'die of shame' but the formal surface description

[1]For clarity, we refer to the constructions by translating their Swedish names into English in the rest of this paper, except for construction-evoking keywords included in the names.

V *av* NP_{BARE} also catches phrasal verbs such as *dra av moms* ‘subtract sales tax’ and passive voice constructions like *läses av flickor* ‘is read by girls’. The variable construction slots are not restricted to any obvious semantic classes, although emotions and physical states are clearly salient at first glance.

4.2 PROPORTION_*i/om*

PROPORTION_*i/om* is a rate construction that combines two entities, a numerator and a denominator, joined by the preposition *i* ‘in’ or *om* ‘about’, and is restricted to temporal relations such as frequency and speed, and salary rates also fit into this scheme. Its structure sketch NP_{INDEF} [*i* | *om*] N_{DEF} catches true instances such as *två gånger om dagen* ’two times a day’ and *åttio pesos i månaden* ‘eighty pesos a month’, as well as false positives such as *tio dagar i fängelset* ‘ten days in jail’.

4.3 REFLEXIVE_MOTION

As mentioned in Section 2, REFLEXIVE_MOTION is a self-motion construction where an actor expressed with a reflexive traverses a path in a direction from a place or towards a goal. The verb typically describes the means or manner of the motion, while the prepositional/adverbial phrase contributes the direction. The formal description V PN_{REFL} [PP | ADVP] captures prototypical occurrences like *sätta sig ned* ‘sit down’ and *ta sig fram* ‘make one’s way’ as well as instances with verbs that do not usually indicate motion, like *svetta sig igenom* ‘sweat one’s way through’ and *läsa sig bakåt* ‘read one’s way backwards’. False hits in the search result include common reflexive verbs that do not express motion, like *känna sig glad* ‘feel happy’ and *anförtro sig åt* ‘confide in’ as well as certain lexicalized multiword expressions like *tränga sig på* ‘intrude’ and *sätta sig på tvären* ‘be obstinate’.

4.4 QUANTIFYING_GENITIVE:TIME

The QUANTIFYING_GENITIVE:TIME construction is defined as a genitive modifier that specifies the duration of an activity. The formal description DET N_{GEN} N_{INDEF} captures true instances of the construction such as *en stunds tystnad* ‘a moment’s silence’, but also the related scale construction *fem meters djup* ‘five meter’s depth’ (described below), as well as other genitives like *en människas skugga* ‘the shadow of a human’.

4.5 QUANTIFYING_GENITIVE:SCALE

The scale construction QUANTIFYING_GENITIVE:SCALE is the sibling of the time construction described above. Here, the genitive modifier specifies the value on a scale expressed by the noun phrase. It shares the formal description DET N_{GEN} N_{INDEF} with QUANTIFYING_GENITIVE:TIME but is semantically restricted to scalable measures like height, depth, age, size and other things that can be quantified. Consequently, the query captures true instances of the scale construction like *tusen meters höjd* ‘a thousand meter’s height’ and *elva månaders hyra* ‘eleven month’s rent’, but also instances of the aforementioned time construction *tre timmars seglats* ‘three hour’s sailing trip’. Again, other genitive phrases like *en kvinnas fot* ‘the foot of a woman’ also end up in the search batch.

4.6 BOUNDED_EVENT_*på*

The BOUNDED_EVENT_*på* construction is a time expression that modifies the duration in time of a completed action, expressed using the preposition *på*, corresponding to the similar English construction using *in*. It is a specific and rather restricted instance of a more general pattern for prepositional time adverbials. The construction can only be used with events of bounded aspect, and thereby specifies the time required to complete the event. A related but separate construction is used for negated events, where time adverbials with *på* can be used to describe the duration that has passed since an event took place; this construction is called SPECIFIED_TIME:POLARITY. The structure sketch *på* NP naturally translates to a search query that captures all prepositional phrases with the preposition *på*, and even though the noun slot is strictly restricted to time expressions it is impossible to delimit it from related time constructions without taking the wider context into account. The search hits include true instances of the construction such as *rummet tömdes på några sekunder* ‘the room was emptied in a few seconds’ as well as false instances like the negative construction *ingen hade samlat ved på ett år* 'nobody had been collecting firewood for a year'. However, most of the answer set consists of typical PPs like *på en stol* 'on a chair'.

5 Construction retrieval with lexical-semantic reranking

The new benchmark allows us to investigate and compare different systems for retrieving occurrences of constructions in corpora. We now describe the implementation of a retrieval system for finding occurrences of constructions: in particular, we discuss how the ranking function is defined, including the various lexical-semantic similarities the user can choose from, and how the ranking function is trained from a few user-selected seed examples.

The system is built on top of the corpus search service *Korp*, described in Section 4. This is the most comprehensive service of this kind for Swedish, but any similar search tool could be used if building a similar retrieval system for other languages. As when creating the benchmark, the first step of searching for occurrences is to call the underlying corpus search system with a query that describes the surface form of the construction (corresponding to the structure sketch in Figure 1). For instance, if we are looking for occurrences of the QUANTIFYING_GENITIVE:TIME construction (e.g. *an hour's rest*), we use a query corresponding to its structure sketch DET N_{GEN} N_{INDEF}. A number of hits are then returned, out of which some are instances of the construction we are looking for, while others are unrelated. For instance, the structural pattern mentioned above will match any genitive, such as *a dog's life*.

To address this problem, we apply a *reranking* function. The user is asked to provide the system with two additional types of information: (1) a number of positive seed examples of sentences containing true instances of the construction she is looking for and (2) what linguistic properties to consider for particular slots in the search string. For instance, a user could say that the ranking function should consider the distributional similarity function based on the second word in the hit, e.g. the time word in the example above, and then select a number of occurrences such as *an hour's rest, three years' study*. With a carefully designed ranking function and representative seed examples, the system can rank time/activity expressions above other expressions matching that surface pattern.

5.1 Training the reranking function

The reranker is a scoring function $R(x)$ applied to each hit x returned by the corpus search. Learning rankers is widely studied in IR (Liu, 2009); in this work, we assume that each hit x can be analyzed using m different similarity functions σ_j, selected by the user on the fly. Assuming there are labeled examples $x_1, \ldots, x_n$, we write the ranker as

$$R(x) = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} \sigma_j(x, x_i)$$

where α_{ij} is a weight representing the contribution of similarity σ_j applied to the labeled example x_i. If the σ_j are valid kernels, this is the dual form of a linear scoring function, and the α_{ij} can be determined by training a kernelized learner such as the ranking SVM (Joachims, 2002). We prefer to use a simpler learning method that is efficient and that works with just a few examples. For these reasons, we set the weights by computing the *centroid* of the positively labeled instances: that is, by setting all α_{ij} to $1/n$. This method has trivial computation time and allows the use of any similarity function, not just kernels. More complex learners for this scenario, e.g. the one-class SVM (Manevitz and Yousef, 2001), may be considered in future work, and we could also imagine the weights being set manually by the users.

5.2 Similarity functions used in the reranker

The corpus search system takes care of basic surface-oriented features (word forms, morphology, grammatical functions and categories), so the central task of the reranker is to use representations of word meaning to go beyond the simple structural information. We investigate different measures: similarities based on hand-crafted lexicons, and distributional similarity computed from corpora.

5.2.1 Network-based similarity

SALDO (Borin et al., 2013) is a large lexical resource that connects senses of Swedish words into a hierarchical semantic network. To measure similarity between two SALDO entries, we use the measure by Wu and Palmer (1994), based on proximity in the tree and the depth of the lowest common ancestor. This measure is a number between 1 and 0, where 1 is the score for two identical entries. A complication is that our corpora lacks sense annotation; however, since the first sense dominates overwhelmingly in corpora for most lemmas (Johansson et al., 2016), we use the first sense to compute the similarities.

5.2.2 Frame-based similarity

An alternative lexicon-based similarity function is based on the Swedish FrameNet (Friberg Heppin and Toporowska Gronostaj, 2012). This resource, similar to its English counterpart (Fillmore and Baker, 2009), maps lemmas to one or more *frames*, which for the current purposes can be seen as semantic classes. Intuitively, two words have a similar meaning if they belong to the same frame; for instance, *timme* 'hour' and *minut* 'minute' are related because they both belong to the frame CALENDRIC_UNIT. Again, we have to deal with the lack of word sense annotation in the corpora, so we define the similarity to be 1 if the two words share at least one frame, and 0 otherwise.

5.2.3 Distributional similarity

In a distributional model (Turney and Pantel, 2010), the meaning representation of a word is computed by observing the contexts in which it appears in a corpus. This is represented as a vector, which makes it possible to apply geometric operations – most importantly, to compute a word similarity by using a function such as the cosine. We trained `word2vec` (Mikolov et al., 2013) on a 1-billion mixed corpus, preprocessed by lemmatization and compound splitting. We used the default settings, except the dimensionality which was set to 512. To compute the similarity between two words, we applied the cosine to their lemma vectors.[2] Again, this similarity is at most 1, which happens if the vectors are identical.

6 Experiments

We first investigated the effect of the choice of lexical similarity. Table 2 shows the average precision scores for three different similarities: Wu–Palmer in SALDO, frame-based, and distributional. As a baseline we include a lemma-based similarity corresponding to a simple search with a number of specified lemmas in the variable construction slots (that is, we get exactly what we asked for and nothing else). As seed examples, the rerankers were trained on the first 15 positively labeled instances in the collection.

Construction	lemma	SALDO	frame	distributional
V_*av*_NP	0.69	0.73	0.63	**0.86**
PROPORTION_*i/om*	0.64	0.68	**0.95**	0.74
REFLEXIVE_MOTION	0.59	0.53	**0.61**	0.56
QUANTIFYING_GENITIVE:TIME	0.40	0.48	**0.60**	0.49
QUANTIFYING_GENITIVE:SCALE	0.64	0.63	0.52	**0.68**
BOUNDED_EVENT_*på*	0.43	0.51	0.36	**0.60**

Table 2: Effect of the choice of lexical similarity function.

The result clearly shows that reranking based on a lexical-semantic model can give very strong improvement over the lemma-based baseline. However, it should be noted that there is considerable variation in the result. For instance, for PROPORTION_*i/om* and QUANTIFYING_GENITIVE:TIME, the frame-based reranker outperforms the others significantly. As we will see in the detailed analysis, a likely explanation is that the slot fillers in these constructions have well-defined semantic restrictions that correspond cleanly to FrameNet frames, in particular time-related words (frames such as Calendric_unit and Measure_duration). In the case of V_*av*_NP, QUANTIFYING_GENITIVE:SCALE and BOUNDED_EVENT_på it is instead the distributional model that works best. The distributional model seems to work best when the slot fillers are not restricted to a narrowly defined semantic class (corresponding to a FrameNet frame), but instead belong to a broader semantic domain, like emotional states for V_*av*_NP.

The network-based lexicon similarity does better than the baseline in most cases, but never outperforms the other similarities. It is difficult to speculate about why, but we can at least conclude that FrameNet frames are better at capturing narrowly defined semantic classes and distributional models do better at generalizing beyond taxonomic similarity scores.

[2]We pick the first lemma if lemmatization is ambiguous.

6.1 Qualitative evaluation

We conducted a qualitative evaluation of the system by analyzing each construction search separately. The reranked instances were inspected with a particular focus on false positives, which are telling indicators of the shortcomings of our approach. We will also inspect the lower regions of the reranked list and say something about false negatives – true occurrences that end up near the bottom of the reranked search list. This qualitative evaluation is telling, since it may also exemplify how a constructicographer would describe delimiting characteristics of specific constructions.

6.1.1 V_*av*_NP

We expected the V_*av*_NP construction (e.g. *rodna av ilska* 'blush with anger') to be a hard nut to crack because of its great productivity and high lexical variation, but the distributional similarity performs beyond expectations. In particular, this model succeeds in generalizing beyond seen instances and gives high ranking scores to a wide array of new and creative occurrences of the V_*av*_NP construction.

Although there are no distinct semantic restrictions on the variable slots of the V_*av*_NP constructions, the state or event caused by the noun is typically physical or emotional. Representative verbs include *darra* 'shiver', *lida* 'suffer', *dö* 'die', *gråta* 'cry', *rodna* 'blush', *skälva* 'quiver', *kvida* 'whimper', *flåsa* 'pant' and *skaka* 'shake'. The noun slot is typically occupied by event nouns such as *raseri* 'rage', *smärta* 'pain', *ansträngning* 'exertion', *ängslan* 'anxiety', *migrän* 'migraine', *förälskelse* 'infatuation' and *upphetsning* 'excitement'. The distributional model excels in finding commonalities between these words, while the frame-based similarity is too restricted. The slot fillers display so much lexical variation that the frame-semantic lexicon manages to capture just a fraction of them.

6.1.2 PROPORTION_*i/om*

As mentioned above, it is hardly surprising that the search system is good at detecting the PROPORTION_*i/om* construction or that the best performing similarity feature in this case is frame-based. The rate construction (e.g. *tre gånger om dagen* 'three times a day') is strictly restricted to temporal relations, so the denominator will always be a time-related word belonging to a few well-defined frames.

6.1.3 REFLEXIVE_MOTION

The evaluation results for REFLEXIVE_MOTION (e.g. *pressa sig ut* 'press oneself out') are unimpressive. None of the rankers are particularly good at detecting this construction: the frame-based reranker just barely beats the baseline. When inspecting the results, it seems that there is too much diversity in this set, with creative metaphors frequently used, for the lexical models to be very effective.

On a lighter note, while this construction is diverse, we observed that some subsets of the occurrences can be handled nicely: if we are particularly interested in occurrences of the REFLEXIVE_MOTION construction where the verb has a more specific manner meaning, we can handpick seed examples of that kind. By doing so, sentences like (1) rise to the top of the answer set. This means that even if the overall score is low, the user can still use the search system productively to find a certain important subclass.

(1)											
	Jag	tror	att	han	skulle	ha	[sovit	sig	igenom]	hela	eländet.
	I	think.PRS	that	he	would.AUX	have.AUX	sleep.PRF	REFL	through	hole	misery.DEF

'I think that he would have slept through the whole ordeal.'

6.1.4 QUANTIFYING_GENITIVE:TIME

For the time construction QUANTIFYING_GENITIVE:TIME (e.g. *två timmars vila* 'two hours' rest'), the frame-based reranker clearly outperforms the other similarity functions in this case since the construction is restricted to time expressions, which we have already seen fit neatly into a few particular frames. Since time expressions are such a strong feature for detecting this construction, it comes as no surprise that false positives near the top of the ranked list contain time words as well. Example (2) is a false hit that is in fact an instance of the superficially similar QUANTIFYING_GENITIVE:SCALE construction.

(2)							
	Det	var	[femton	år-s	åldersskillnad]	mellan	oss.
	It	is.PST	fifteen	year.PL-GEN	age difference.INDEF	between	us.

'It was fifteen years of age difference between us.'

6.1.5 QUANTIFYING_GENITIVE:SCALE

The lexical similarities give a smaller improvement for the QUANTIFYING_GENITIVE:SCALE than for the related time construction. In particular, while the frame-based feature worked best for the time construction, it receives the lowest average precision in this case. The scalable measures that turn up here are diverse and do not seem to correspond neatly with FrameNet frames.

The distributional similarity does best at generalizing beyond seen examples, and quite impressively hands out high ranking scores to instances as diverse as *50 procents chans* '50 percent chance', *två veckors skäggstubb* 'two weeks' stubble' and *tre spalters bredd* 'three columns' width'. The precision starts to drop when false hits from the QUANTIFYING_GENITIVE:TIME construction and other constructions start to appear.

6.1.6 BOUNDED_EVENT_*på*

Even though this construction (e.g. *på två timmar* 'for two hours') is strictly restricted to time related words, the frame-based reranker performs worst in the evaluation. The explanation is quite straightforward; just spotting the time word is not enough to disambiguate BOUNDED_EVENT_*på* from related time constructions: most of the information that could be used to discriminate can be found outside of the hit. Recall that BOUNDED_EVENT_*på* only occurs with events of bounded lexical aspect.

Since we had little hope that the reranker would be effective for such a contextually dependent construction as BOUNDED_EVENT_*på*, it is a pleasant surprise to see that the distributional model is doing significantly better than the baseline. A possible explanation is that there are lexical preferences at play, beyond the more general time restriction. Short time spans like *på ett ögonblick* 'in an instant' seem more likely to be instances of this construction than cases such as *på en söndag* 'on a Sunday'. Among the top-ranked false positives, we find quite a few hits with the lexicalized phrase *på en gång* 'at once'.

Introducing more contextual information seems necessary for dealing successfully with a construction like this one. However, in this particular case it is not entirely clear how to determine the lexical aspect of the event in an automated fashion. A more straightforward feature to introduce would be negations that can be relatively easily spotted by using a list of negative polarity items.

7 Discussion

We considered the problem of searching for occurrences of a grammatical construction as a retrieval problem, and we created a new benchmark collection with annotated examples for six different constructions defined by the Swedish constructicon. This new resource allows us to investigate the effectiveness of different retrieval models. As a proof of concept, we presented a simple interactive architecture for searching for constructions, where a user provides a number of positive examples (occurrences of the construction) and tailors a ranking function based on a user-defined combination of features, and our benchmark enabled us to carry out detailed quantitative and qualitative investigations of the effect of different models of lexical representation on the retrieval performance of this system. All our experiments were carried out using Swedish, because of the availability of the Swedish Constructicon used to select the constructions in the benchmark, but our approach is general and could be ported to other languages, including English: similar corpus search tools and lexical resources are readily available.

As expected, grammatical constructions are diverse and the ranking function must be tailored for each construction. The most consistent result is that lexical-semantic models based on a constructions' slot fillers improve the reranker, but exactly which of them – lexicon-based or distributional – is most effective depends on the construction. It is important to note that the accuracy of the reranker depends on the construction definition and to which extent such semantic restrictions are in fact at play. The system should therefore be useful in the work of characterizing and defining constructions.

As we have already pointed out in the text, there are several ways to extend this work: more complex learning algorithms for the rankers could be considered, or we could make use of information beyond the slot fillers of the construction. Also, we have now studied the retrieval problem in isolation, but since our main motivation is that the system should be used in the practical work of linguists working with construction grammar, it would be interesting to investigate the usability and interaction aspect as well.

Acknowledgements

We are grateful for the availability of open lexicons, corpora, and technical infrastructures provided by *Språkbanken*, the Swedish Language Bank. The constructicon resource used in this research was developed during the project *A Swedish Constructicon*, which was funded by the Bank of Sweden Tercentenary Foundation (grant P12–0076:1). RJ was supported by the Swedish Research Council under grant 2013–4944, *Distributional methods to represent the meaning of frames and constructions*.

References

Lars Borin, Markus Forsberg, and Johan Roxendal. 2012. Korp – the corpus infrastructure of Språkbanken. In *Proceedings of the Eighth conference on International Language Resources and Evaluation (LREC-2012)*, pages 474–478, Istanbul, Turkey.

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

Oliver Christ. 1994. A modular and flexible architecture for an integrated corpus query system. In *Proceedings of the 3rd Conference on Computational Lexicography and Text Research*, pages 23–32, Budapest, Hungary.

Sabine De Knoop and Gaëtanelle Gilquin, editors. 2016. *Applied Construction Grammar*. Walter de Gruyter.

Marie Dubremetz and Joakim Nivre. 2015. Rhetorical figure detection: the case of chiasmus. In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, pages 23–31, Denver, Colorado, USA, June. Association for Computational Linguistics.

Charles J. Fillmore and Collin Baker. 2009. A frames approach to semantic analysis. In B. Heine and H. Narrog, editors, *The Oxford Handbook of Linguistic Analysis*, pages 313–340. OUP, Oxford.

Charles J. Fillmore, Paul Kay, and Mary C. O'Connor. 1988. Regularity and idiomaticity in grammatical constructions. the case of *let alone*. *Language*, 64:501–538.

Charles J. Fillmore, Russell Lee-Goldman, and Russell Rhomieux. 2012. The FrameNet Constructicon. In H. C. Boas and I. A. Sag, editors, *Sign-Based Construction Grammar*, pages 309–372. CSLI Publications, Stanford.

Markus Forsberg, Richard Johansson, Linnéa Bäckström, Lars Borin, Benjamin Lyngfelt, Joel Olofsson, and Julia Prentice. 2014. From construction candidates to constructicon entries. an experiment using semi-automatic methods for identifying constructions in corpora. *Constructions and Frames*, 6(1):114–135.

Karin Friberg Heppin and Maria Toporowska Gronostaj. 2012. The rocky road towards a Swedish FrameNet – creating SweFN. In *Proceedings of the Eighth conference on International Language Resources and Evaluation (LREC-2012)*, pages 256–261, Istanbul, Turkey.

Adele Goldberg. 2013. Constructionist approaches. In Th. Hoffmann and G. Trousdale, editors, *The Oxford Handbook of Construction Grammar*, pages 15–31. OUP, Oxford.

Stefan Th. Gries. 2003. Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics*, 1:1–27.

Martin Hilpert. 2013. Corpus-based approaches to constructional change. In Th. Hoffmann and G. Trousdale, editors, *The Oxford Handbook of Construction Grammar*, pages 458–477. OUP, Oxford.

Thorsten Joachims. 2002. Optimizing search engines using clickthrough data. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pages 133–142, Edmonton, Canada.

Richard Johansson, Yvonne Adesam, Gerlof Bouma, and Karin Hedberg. 2016. A multi-domain corpus of Swedish word sense annotation. In *Proceedings of the Language Resources and Evaluation Conference (LREC)*, Portorož, Slovenia.

Tie-Yan Liu. 2009. *Learning to Rank for Information Retrieval*. Springer Berlin Heidelberg.

Benjamin Lyngfelt, Linnéa Bäckström, Lars Borin, Anna Ehrlemark, and Rudolf Rydstedt. forthcoming. Constructicography at work: Theory meets practice in the Swedish Constructicon. In B. Lyngfelt, T. T. Torrent, L. Borin, and K. H. Ohara, editors, *Constructicography: constructicon development across languages*. John Benjamins, Amsterdam.

Larry M. Manevitz and Malik Yousef. 2001. One-class SVMs for document classification. *Journal of Machine Learning Research*, 2:139–154.

Tomáš Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositiona lity. In *Advances in Neural Information Processing Systems*, pages 3111–3119.

Kyoko Hirose Ohara. 2013. Toward constructicon building for Japanese in Japanese FrameNet. *Veredas*, 17(1):11–27.

Julia Prentice and Emma Sköldberg. 2011. Figurative word combinations in texts written by adolescents in multilingual school environments. In *Young urban Swedish. Variation and change in multilingual settings*.

Emma Sköldberg, Linnéa Bäckström, Lars Borin, Markus Forsberg, Benjamin Lyngfelt, Leif-Jöran Olsson, Julia Prentice, Rudolf Rydstedt, Sofia Tingsell, and Jonatan Uppström. 2013. Between grammars and dictionaries: a Swedish Constructicon. In *Proceedings of eLex*, pages 310–327, Tallinn, Estonia.

Anatol Stefanowitsch and Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2):209–243.

Tiago Timponi Torrent, Ludmila Meireles Lage, Thais Fernandes Sampaio, Tatiane da Silva Tavares, and Ely Edison da Silva Matos. 2014. Revisiting border conflicts between FrameNet and construction grammar: Annotation policies for the Brazilian Portuguese constructicon. *Constructions and Frames*, 6(1):33–50.

Peter D. Turney and Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37:141–188.

David Wible and Nai-Lung Tsao. 2010. StringNet as a computational resource for discovering and investigating linguistic constructions. In *Proceedings of the NAACL HLT Workshop on Extracting and Using Constructions in Computational Linguistics*, pages 25–31, Los Angeles, United States.

Zhibiao Wu and Martha Palmer. 1994. Verbs semantics and lexical selection. In *Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics*, pages 133–138, Las Cruces, United States.