Improving Verb Phrase Extraction from Historical Text by use of Verb Valency Frames

Eva Pettersson and Joakim Nivre
Department of Linguistics and Philology
Uppsala University
firstname.lastname@lingfil.uu.se

Abstract

In this paper we explore the idea of using verb valency information to improve verb phrase extraction from historical text. As a case study, we perform experiments on Early Modern Swedish data, but the approach could easily be transferred to other languages and/or time periods as well. We show that by using verb valency information in a post-processing step to the verb phrase extraction system, it is possible to remove improbable complements extracted by the parser and insert probable complements not extracted by the parser, leading to an increase in both precision and recall for the extracted complements.

1 Introduction

Information extraction from historical text is a challenging field of research that is of interest not only to language technology researchers but also to historians and other researchers within the humanities, where information extraction is still to a large extent performed more or less manually due to a lack of NLP tools adapted to historical text and insufficient amounts of annotated data for training such tools.

In the *Gender and Work* project (GaW), historians are building a database with information on what men and women did for a living in the Early Modern Swedish society, i.e. approximately 1550–1800 (Ågren et al., 2011). This information is currently extracted by researchers manually going through large volumes of text from this time period, searching for relevant text passages describing working activities. In this process, it has been noticed that working activities often are described in the form of verb phrases, such as *hugga ved* ("chop wood"), *sälja fisk* ("sell fish") or *tjäna som piga* ("serve as a maid"). Based on this observation, Pettersson et al. (2012) developed a method for automatically extracting verb phrases from historical documents by use of spelling normalisation succeeded by tagging and parsing. Using this approach, it is possible to correctly identify a large proportion of the verbs in Early Modern Swedish text. Due to issues such as differences in word order and significantly longer sentences than in present-day Swedish texts (combined with sentence segmentation problems due to inconsistent use of punctuation), it is however still hard for the parser to extract the correct complements associated with each verb.

In this work we propose a method for improving verb phrase extraction results by providing verb valency information to the extraction process. We describe the effect of removing improbable complements from the extracted verb phrases, as well as adding probable complements based on verb valency frames combined with words and phrases occurring in close context to the head verb.

2 Related Work

Syntactic analysis of historical text is a tricky task, due to differences in vocabulary, spelling, word order, and grammar. Sánchez-Marco (2011) trained a tagger for Old Spanish, based on a 20 million token corpus of texts from the 12th to the 16th century, by expanding the dictionary and modifying tokenisation and affixation rules. An accuracy of 94.5% was reported for finding the right part-of-speech, and an accuracy of 89.9% for finding the complete morphological tag.

In many cases, there is a lack of large corpora for training such tools, and alternative methods are called for. Schneider (2012) presented a method

for adapting the Pro3Gres dependency parser to analyse historical English text. In this approach, spelling normalisation is a key factor, transforming the historical spelling to a modern spelling by use of the VARD2 tool (Baron and Rayson, 2008) before parsing. In addition to spelling normalisation, a set of handwritten grammar rules were added to capture unseen interpretations of specific words, for relaxing word order constraints, and for ignoring commas in a sentence. Schneider concluded that spelling normalisation had a large impact on parsing accuracy, whereas the grammar adaptations where easy to implement but lead to small improvements only.

Pettersson et al. (2013) also presented an approach to automatic annotation of historical text based on spelling normalisation. In this approach, the historical spelling is translated to a modern spelling employing character-based statistical machine translation (SMT) techniques, before tagging and parsing is performed by use of standard natural language processing tools developed for present-day language. The method was evaluated on the basis of verb phrase extraction results from Early Modern Swedish text, where the amount of correctly identified verb complements (including partial matches) increased from 32.9% for the text in its original spelling to 46.2% for the text in its automatically modernised spelling. Earlier work by the same authors showed that using contemporary valency dictionaries to remove extracted complements not adhering to the valency frame of a specific verb had a positive effect on verb phrase extraction precision (Pettersson et al., 2012).

In the context of valency-based parsing of modern language, Jakubíček and Kovář (2013) introduced a verb valency-based method for improving Czech parsing. Their experiments were based on the Synt parser, which is a head-driven chart parser with a hand-crafted meta-grammar for Czech, producing a list of ranked phrase-structure trees as output. They used two different dictionaries with valency information to rerank the suggested parses in accordance with the valency frames suggested for the verb in the dictionaries. Evaluation was performed on the Brno Phrasal Treebank using the leaf-ancestor assessment metric, and an improvement from 86.4% to 87.7% was reported for the highest-ranked tree when comparing the Synt parser in its original setting to the inclusion of valency frames for reranking of the output parses.

For modern Swedish, Øvrelid and Nivre (2007) experimented on ways to improve parsing accuracy for core grammatical functions including for example object, subject predicative, and prepositional argument. They found that by providing the parser with linguistically motivated features such as animacy, definiteness, pronoun type and case, a 50% error reduction could be achieved for the syntactic functions targeted in the study.

3 Approach

In this work, we adopt the verb phrase extraction method presented in Pettersson et al. (2013), where verbs and complements are extracted from historical text based on output from NLP tools developed for present-day Swedish. In addition to their approach, we also include a post-processing step, removing and/or inserting verbal complements based on the valency frame of the head verb.

The full process is illustrated in Figure 1, where the first step is tokenisation of the source text by use of standard tools. The tokenised text is then to be linguistically annotated in the form of tagging and parsing. To the best of our knowledge, there is however no tagger nor parser available trained on Early Modern Swedish text. Since these tools are sensitive to spelling, the tokenised text is therefore normalised to a more modern spelling by use of character-based SMT methods, before tagging and parsing is performed using tools trained for modern Swedish. For tagging, we use HunPOS (Halácsy et al., 2007) with a Swedish model based on the Stockholm-Umeå corpus, SUC version 2.0 (Ejerhed and Källgren, 1997). For parsing, we use MaltParser version 1.7.2 (Nivre et al., 2006a) with a pre-trained model based on the Talbanken section of the Swedish Treebank (Nivre et al., 2006b).

After tagging and parsing, the annotations given by the tagger and the parser are projected back to the text in its original spelling, resulting in a tagged and parsed version of the historical text, from which the verbs and their complements are extracted. The complements included for extraction are the following: subject (for passive verbs only, where the subject normally corresponds to the direct object in an active verb construction), direct object, indirect object, prepositional complement, infinitive complement, subject predicative, verb particle, and reflexive pronoun. As mentioned, we also add a post-processing filter as a complementary step, using valency information to

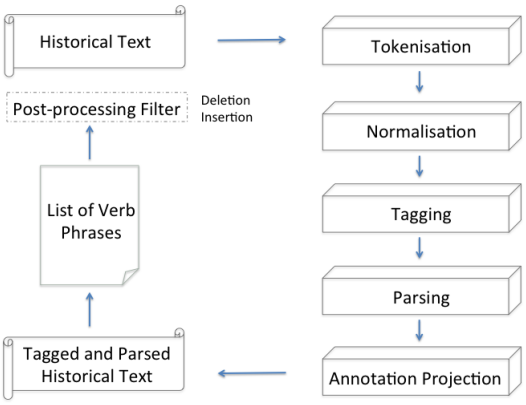

Figure 1: Method overview.

modify the complements suggested by the parser.

3.1 Deletion of Improbable Complements

As discussed in Section 1, certain characteristics of historical text make it difficult for the parser to correctly extract the complements of a verb. Therefore, we add valency information in a post-processing step, filtering away extracted complements that do not conform to the valency frame of the verb. A similar idea was presented in Pettersson et al. (2012), where filtering was based on valency frames given in two contemporary dictionaries, i.e. Lexin[1] and Parole[2]. However, some word forms in historical text are not frequent enough in contemporary language to occur in modern dictionaries. Examples from the GaW training corpus are *absentera* (old word for ”be absent”), *umgälla* (old word for ”suffer for”), and *ärna* (old word for ”intend to”). Moreover, the meaning of verbs tend to change over time, and it is not obvious that verb valency frames for present-day Swedish also holds for historical Swedish. An example from the GaW corpus is the verb *slå* (”hit”) which in both Lexin and Parole is listed as a monotransitive verb (”to hit someone”). In the GaW corpus however, it is repeatedly used as a ditransitive verb, as in *Sedhan hadhe Erich OluffSon slaghit Pelle Pederssonn tre blånader* (”Then Erich OluffSon had hit Pelle Pederssonn three bruises”). A comparison between the valency frames present in the GaW corpus and the frames present in the Lexin dictionary shows that only 16% of the verb forms that are present in both the old and the modern resource (108 out of 675 verb forms) have equal valency frames. In our approach to deletion of improbable complements, we therefore base the valency frames not only on the contemporary valency dictionaries, but also on the verbal complements occurring in the training part of the GaW corpus.

Deletion experiments are performed for all complement types extracted from the parser except for subjects, since a verb is typically expected to have a subject. We present deletion experiments for the following five settings:

1. **Lexin**
   For each extracted complement, if the head verb is present in the Lexin valency dictionary and the valency frame in Lexin does not allow for a complement of the type indicated by the parse label, the complement is removed from the extracted verb phrase.

2. **Parole**
   For each extracted complement, if the head verb is present in the Parole valency dictionary and the valency frame in Parole does not allow for a complement of the type indicated by the parse label, the complement is removed from the extracted verb phrase.

3. **GaW Corpus**
   For each extracted complement, if the head verb is present in the training part of the GaW corpus, and none of the occurrences in the corpus contain a complement of the type indicated by the parse label, the complement is removed from the extracted verb phrase.

4. **All combined**
   For each extracted complement, if the head verb is present in all three resources men-

[1] http://spraakbanken.gu.se/lexin/valens_lexikon.html
[2] http://spraakbanken.gu.se/swe/resurs/parole

tioned above, and none of these resources allow for a complement of the type indicated by the parse label, the complement is removed from the extracted verb phrase. Likewise, if the head verb is present in only two of these three resources, and none of these two resources allow for a complement of the type indicated by the parse label, the complement is removed from the extracted verb phrase. Finally, if the head verb is present in one resource exclusively, and this resource does not allow for a complement of the type indicated by the parse label, the complement is removed from the extracted verb phrase.

5. **All one-by-one**
   For each extracted complement, if the head verb is present in the best-performing resource, i.e. the resource yielding the highest complement extraction f-score in the first three experiments, and the valency frame in this resource does not allow for a complement of the type indicated by the parse label, the complement is removed from the extracted verb phrase. Otherwise, if the head verb is present in the second best-performing resource, and the valency frame in this resource does not allow for a complement of the type indicated by the parse label, the complement is removed from the extracted verb phrase. Only if the head verb is not present in any of the two best-performing resources, the third resource is consulted.

3.2 Insertion of Probable Complements

Apart from filtering away unlikely complements extracted by the parser, we also aim at inserting probable complements not found by the parser, by searching the parsed sentence for words and phrases that match the valency frame of the head verb, but which have not been extracted by the parser. Since the word order is more varying in Early Modern Swedish than in present-day Swedish, all complements are searched for both to the left and to the right of the head verb.

In the insertion experiments, we focus on phrasal verbs in the broader sense, including particles, reflexives, and prepositional complements. We believe that these complement types are relatively easy to recognise in a sentence. Furthermore, if for example a reflexive pronoun is found close to the head verb in the sentence, and the valency frame suggests a reflexive pronoun, then the probability that this reflexive belongs to the verb is rather high. The same argument holds for prepositional phrases containing the expected preposition to form a prepositional complement, and for prepositions or adverbials identical to a particle expected by the valency frame of the head verb. For direct and indirect objects on the other hand, even if we find a noun phrase close to the verb, it would still be hard to determine whether this noun phrase actually corresponds to a direct or indirect object, since noun phrases may occur with many different functions in a clause, and the word order is not fixed, particularly not for historical text. Therefore we would run a high risk of extracting for example the subject noun phrase instead of the direct or indirect object noun phrase, especially for languages like Swedish, where subject/object distinctions are not manifested morphologically other than for pronouns. Furthermore, direct objects are not always expressed in the form of noun phrases, but are quite often expressed as for instance clauses, as in the following example from the GaW corpus: *fordra at Barnet skal döpas hemma* (”demand that the Child should be christened at home”). Similarly, subject predicatives may also be expressed in varying ways and infinitive complements are often ambiguous to other functions. Thus, these categories are excluded from the insertion experiments.

In accordance with the arguments given above, the following three experiments are performed for insertion of probable complements:

1. **Insertion of prepositional complement**
   If the valency frame of the head verb (in any of the three valency resources) allows for a prepositional complement, and a prepositional phrase containing the expected preposition is found either to the left or to the right of the head verb, this prepositional phrase is added to the extracted verb phrase with a prepositional complement label.

2. **Insertion of particle**
   If the valency frame of the head verb allows for a particle, and a word that is identical to the expected particle and tagged as preposition or adverb is found either to the left or to the right of the head verb, this preposition or adverb is added to the extracted verb phrase with a particle label.

3. **Insertion of reflexive**
   If the valency frame of the head verb allows for a reflexive pronoun, and the word form *sig* (”oneself”), or the alternative historical spelling *sigh*, is found either to the left or to the right of the head verb, this word form is added to the extracted verb phrase with a reflexive label.

4 Data

Verb valency frames are extracted from three sources: the contemporary Lexin valency dictionary, the contemporary Parole valency dictionary, and the training and development parts of the GaW corpus of Early Modern Swedish court records and church documents. Evaluation is performed on the evaluation part of the GaW corpus. All the verbs in the GaW corpus have been manually annotated as such, and all complements adhering to the verbs have been annotated with labels denoting subject (for passive verbs only), direct object, indirect object, prepositional complement, infinitive complement, subject predicative, verb particle, and reflexive pronoun. Furthermore, the training and development parts of the corpus have been annotated with information on the manually modernised spelling for each original word form occurring in the text.

Both in the Lexin dictionary and in the Parole dictionary, verb valency frames are connected to the present tense form of the verb only, without information on other inflectional forms of the verb. In the verb phrase extraction process however, we need to connect whatever inflectional form of the verb that is used in the sentence to the correct valency frame. For broader coverage of the valency dictionaries, the present tense forms were therefore expanded to other inflectional forms based on the Saldo dictionary and the SUC corpus. The Saldo dictionary is a dictionary of present-day Swedish word forms, with morphological and inflectional information (Borin et al., 2008). By comparing the present tense verb form in Lexin or Parole to the Saldo dictionary, it is thus possible to extract a lemma corresponding to the verb form, and from that lemma all the inflectional forms adhering to that lemma. For verb forms not found in the Saldo dictionary, the SUC corpus was consulted. Since this corpus has been manually annotated with lemma information, all inflectional forms of the same lemma occurring in the corpus may thus be extracted. For Lexin and Parole verb forms not found in neither Saldo nor SUC, only the present tense form of the verb is stored with its corresponding valency frame.

For the GaW corpus, we have a similar problem in that only those verb forms that occur in the corpus will be assigned a valency frame, and if several forms of the same verb occur in the corpus, these will be assigned valency frames separate from each other. To deal with this, we use the same method of comparison to Saldo and SUC for retrieving the full set of word forms associated with a verb form, assigning the same valency frame to all verb forms belonging to the same lemma. In this process, we use the manually normalised form of each verb for comparison towards Saldo and SUC, to avoid mismatches due to spelling variation in the historical corpus.

It could be argued that instead of generating all fullforms for a verb, it would be more efficient to perform lemmatisation prior to comparison. This would however potentially impose more ambiguity to the valency frames, since word forms in the SUC corpus are associated with their base form rather than the actual lemma. This means that present tense forms such as *är* (”is”) and *varar* (”lasts”) are both associated with the same base form *vara* (”to be/to last”), even though their inflectional paradigms and valency frames differ significantly. For properly lemmatised sources, these word forms would instead have been associated with different lemmas, e.g. *vara1* and *vara2*.

Table 1 shows the number of verb forms found in Saldo and SUC respectively, during the process of expanding the valency frames to more inflectional forms. The GaW corpus has been divided into training (train), development (dev) and test sets, where the training part is the same data set as was used for training and tuning in Pettersson et al. (2013), and the development set is the same data set as was used for evaluation in the same paper. In total, the training and development parts contain 600 sentences each, whereas the test set contains 300 sentences. Since the test set will only be used for evaluation, no expansion to inflectional forms is needed for this particular data set.

Table 2 lists the total number of entries in the language resources, before and after word form expansion. We will use the training part of our corpus as a basis for valency frames during model selection, where the development part is used for

	Verbs	Saldo	SUC	Not found
Lexin	3,281	3,181	33	67
Parole	4,304	4,263	26	15
GaW Train	1,329	1,168	14	147
GaW Dev	1,410	1,245	15	150
GaW Test	987	n/a	n/a	n/a

Table 1: Verb forms found in Saldo and SUC during the process of expanding the valency frames to more inflectional forms.

repeated testing. In the final evaluation, the training and development sets are merged to a combined valency resource, and evaluation scores are given for the test part of the corpus.

	verb forms	expanded forms
Lexin	3,281	42,545
Parole	4,304	32,640
GaW Train	1,329	10,032
GaW Dev	1,410	10,394
GaW Test	987	n/a

Table 2: Number of verb forms in the language resources, before and after word form expansion.

5 Evaluation

Evaluation is performed in terms of precision, recall and f-score based on the extracted complements, where the baseline case is the original verb phrase extraction system without any of the above specified amendments. We define true positives as correctly extracted complements. Likewise, false positives are complements extracted by the system that are not present in the gold standard, whereas false negatives are complements that are present in the gold standard but not extracted by the system. Since we are specifically aiming at extracting the correct complements, intransitive verbs that were also identified as intransitive by the extraction system will not contribute to the set of true positives. Intransitive verbs for which the system has extracted complements will however contribute to the set of false positives, whereas verbs identified as intransitive by the system though complements are present in the gold standard will add to the set of false negatives.

We also make a distinction between labelled and unlabelled precision and recall, where labelled precision and recall requires that the correct label for the complement has been assigned, i.e. direct object, prepositional complement etc, whereas unlabelled precision and recall only concerns the extracted word sequences, regardless of what label the parser has assigned to the complement.

Since the overall aim of the verb phrase extraction process is to present to historians text passages that may be of interest, partial matches are also regarded as true positives, as these would still point the user to the right text passage. True positives thus include the following cases, with authentic examples from the GaW corpus:

- **Exact match**
  Gold complement: *2 klimpar smör*
  Extracted complement: *2 klimpar smör*
  ”2 lumps of butter”
- **Substring type A**
  Gold complement: *de penningar och* ***medel***
  Extracted complement: ***medel***
  ”(the money and) **resources**”
- **Substring type B**
  Gold complement: ***detta***
  Extracted complement: ***detta*** *efter honom*
  ”**this** (after him)”
- **Overlap**
  Gold complement: ***förswagat*** *ock förtrygt*
  Extracted complement: *nogh* ***förswagat***
  ”(probably) **weakened** (and oppressed)”

6 Model Selection

In the model selection phase, we try different strategies for deletion and insertion of complements, using the training part of the corpus as a basis for valency frames, and the development part of the corpus for testing.

6.1 Deletion of Improbable Complements

For deletion of improbable complements, we first need to decide which of the five settings listed in Section 3.1 that should be chosen. We therefore ran experiments where deletion is performed for all complement types (except subject), evaluating the results for each setting separately. The results for unlabelled complement extraction are summarized in Table 3.

	Precision	Recall	F-score
Baseline	53.30	51.22	52.24
Lexin	59.76	35.44	44.49
Parole	55.22	38.49	45.36
GaW corpus	57.51	46.77	51.59
All combined	56.62	47.76	51.81
All one-by-one	57.64	46.01	51.17

Table 3: Unlabelled results for deletion of improbable complements with different settings.

As seen from the results, all settings improve pre-

cision as compared to the original system. However, recall varies to a great extent. For the largest resource, i.e. the Lexin dictionary, precision is the highest, but recall is very low. This indicates that a great amount of verb forms are found in the Lexin dictionary, but with valency frames that do not correspond to the way the verbs are used in historical texts, meaning that complements are erroneously deleted. This confirms our initial hypothesis that due to language change, contemporary dictionaries are not sufficient for guiding a parser with valency information. Further arguments for this hypothesis is the fact that even though the GaW training corpus is by far the smallest valency resource, using only this resource for defining verb valency frames results in a substantially higher f-score value than using Lexin or Parole. In fact, the f-score results for using the GaW corpus only are almost as high as for using all resources combined.

Since all methods improve precision as compared to the baseline, we choose the combined method for further experiments, since this method has the highest recall and also the highest f-score.

In the next round of experiments, we want to find out which complement types should be candidates for deletion. The hypothesis is that some complement types may be more thoroughly covered in the valency resources than others. If so, deletion of complements may only be a successful method for some complement types, whereas others should be left unmodified in the deletion process. To test this hypothesis, we tried deletion for each complement type separately, keeping only those that improve f-score as compared to the baseline system. These experiments were run with the combined setting, in accordance with the arguments given above. The results are presented in Table 4, where it can be noticed that only deletion of direct objects and subject predicatives are successful in improving the f-score value as compared to the baseline. Keeping these two categories as candidates for deletion, a precision of 54.96% is achieved, with a recall of 50.25%, as compared to the baseline precision of 53.30 and recall of 51.22.

	Precision	**Recall**	**F-score**
Baseline	53.30	51.22	52.24
A) direct object	54.29	50.62	52.39
B) indirect object	53.37	50.92	52.12
C) prep compl	56.06	47.51	51.43
D) inf compl	53.34	51.02	52.15
E) subj predicative	53.93	50.84	52.34
F) particle	53.45	50.84	52.11
G) reflexive	53.19	50.50	51.81
A + E	54.96	50.25	52.50

Table 4: Unlabelled results for deletion of improbable complements for the setting ”all combined”, varying the complements included for deletion.

6.2 Insertion of Probable Complements

As described in Section 3.2, the insertion experiments are targeted at particles, reflexives, and prepositional objects. Whenever the valency frame of the head verb in the extracted phrase allows for a complement of the specified type, the parsed sentence is searched for words and phrases matching the complement at hand. In the insertion experiments, we tried the following enhancements of the original insertion strategy:

1. Inclusion of stopwords, for which no complements are to be added. The set of stopwords were empirically defined as word forms belonging to any of the lemmas *vara* (”be”), *bli* (”become”), *ha* (”have”) and *finnas* (”exist”).

2. Prohibiting punctuation to occur between the head verb and the candidate complement.

3. Inclusion of a distance threshold, defining how many tokens that may come in between the head verb and the candidate complement. We tried a number of different thresholds, out of which a threshold of 5 tokens turned out to yield the best results.

The insertion results are presented in Table 5, showing that without any restrictions in the insertion process, recall can be increased from 51.22% to 53.63%. This is however at the expense of a substantial drop in precision from 53.30% to 37.47% as compared to the baseline system. Restrictions in the form of A) stopwords for which no complements are inserted, B) prohibition of punctuation between the head verb and the candidate complement, and C) defining a threshold for how many tokens are allowed to occur between the head verb and the candidate complement, all had a positive effect on precision and f-score. Thus, in the best setting, i.e. where all three restrictions are implemented, a precision of 52.57% is achieved, with a recall of 52.45%.

To find out which complements should be included for insertion, we also tried insertion for each complement type separately. As seen from Table 6, the best results are achieved when all

	Precision	Recall	F-score
Baseline	53.30	51.22	52.24
Original	37.47	53.63	44.12
A) Stopwords	45.69	53.41	49.25
B) Punctuation	47.81	53.04	50.29
C) Threshold	51.42	52.54	51.97
A + B + C	52.57	52.45	52.51

Table 5: Unlabelled results for insertion of probable complements.

three complement types are included for insertion.

	Precision	Recall	F-score
Baseline	53.30	51.22	52.24
A) prep compl	52.70	51.86	52.28
B) particle	53.23	51.28	52.24
C) reflexive	53.20	51.75	52.46
A + C	52.60	52.39	52.49
A + B + C	52.57	52.45	52.51

Table 6: Unlabelled results for insertion of probable complements, varying the complements included for insertion.

7 Results

Table 7 presents the complement extraction results on the test corpus, with the training and development part of the GaW corpus merged into a single historical valency resource. Results are presented for the baseline system (without additional deletion or insertion), for the best deletion setting as argued in Section 6.1, for the best insertion setting as argued in Section 6.2, and for both deletion and insertion combined.

	Unlabelled		
	Precision	Recall	F-score
Baseline	61.82	48.68	54.47
Deletion	63.02	47.61	54.24
Insertion	61.88	50.80	55.80
Delete + Insert	63.04	49.74	55.61
	Labelled		
	Precision	**Recall**	**F-score**
Baseline	53.25	38.34	44.58
Deletion	54.75	37.97	44.84
Insertion	53.67	40.45	46.13
Delete + Insert	55.12	40.08	46.41

Table 7: Complement extraction results.

As expected, performing only deletion of complements leads to an increase in precision at the expense of a decrease in recall. Performing only insertion on the other hand leads to an increase in recall without decreasing precision, demonstrating that inserting complements introduces true positives to a higher extent than false positives, which is satisfactory. In fact, insertion of complements results in a slightly higher f-score value than the combination of deletion and insertion. However, the best precision is achieved when both deletion and insertion are performed, yielding a precision of 63.04%, as compared to 61.82% for the baseline system. This setting also improves both precision and recall as compared to the baseline.

8 Conclusion

We have presented a method for improving verb phrase extraction from historical text, by automatically deleting improbable verbal complements extracted by the parser, while at the same time inserting probable complements not extracted by the parser. Our approach is based on verb valency frames rendered from historical corpora and from contemporary valency dictionaries, where the historical corpus had the largest positive effect even though the contemporary dictionaries covered more verb forms. This supports our hypothesis that since language changes over time, valency frames for present-day language may not be enough to cover the syntax in historical text. By automatically deleting and inserting complements based on a combination of the historical corpus and the contemporary dictionaries, an increase in both precision and recall is achieved, as compared to the baseline system.

For historians working with old texts, there is a need for NLP tools to effectively search large volumes of text automatically for text passages of special interest. We believe our method for verb phrase extraction from historical text to be a useful tool for this purpose. Still, there is room for improvement, since the best precision achieved for complement extraction is 63.04%, with a recall of 49.74%. In the current approach, verb valencies are exploited in a post-processing phase, with the original extracted verb phrases as input. Future work includes to explore the possibility of providing valency information already in the parser training phase, enriching the part-of-speech tags with information on whether a certain verb is likely to occur with e.g. a particle or prepositional complement. The hypothesis is that a parser trained on this kind of data will be keen to search harder for the expected complements. It would also be interesting to explore the use of lexical semantics for identifying specific types of complements.

References

Maria Ågren, Rosemarie Fiebranz, Erik Lindberg, and Jonas Lindström. 2011. Making verbs count. The research project 'Gender and Work' and its methodology. *Scandinavian Economic History Review*, 59(3):271–291. Forthcoming.

Alistair Baron and Paul Rayson. 2008. Vard2: A tool for dealing with spelling variation in historical corpora. In *Postgraduate Conference in Corpus Linguistics*, Aston University, Birmingham.

Lars Borin, Markus Forsberg, and Lennart Lönngren. 2008. Saldo 1.0 (svenskt associationslexikon version 2). Språkbanken, University of Gothenburg.

Eva Ejerhed and Gunnel Källgren. 1997. Stockholm Umeå Corpus. Version 1.0. Produced by Department of Linguistics, Umeå University and Department of Linguistics, Stockholm University. ISBN 91-7191-348-3.

Péter Halácsy, András Kornai, and Csaba Oravecz. 2007. HunPos - an open source trigram tagger. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pages 209–212, Prague, Czech Republic.

Miloš Jakubíček and Vojtěch Kovář. 2013. Enhancing czech parsing with verb valency frames. In *CICLing 2013*, pages 282–293, Greece. Springer Verlag.

Joakim Nivre, Johan Hall, and Jens Nilsson. 2006a. MaltParser: A data-driven parser-generator for dependency parsing. In *Proceedings of the 5th international conference on Language Resources and Evaluation (LREC)*, pages 2216–2219, Genoa, Italy, May.

Joakim Nivre, Jens Nilsson, and Johan Hall. 2006b. Talbanken05: A Swedish treebank with phrase structure and dependency annotation. In *Proceedings of the 5th international conference on Language Resources and Evaluation (LREC)*, pages 24–26, Genoa, Italy, May.

Lilja Øvrelid and Joakim Nivre. 2007. When word order and part-of-speech tags are not enough – Swedish dependency parsing with rich linguistic features. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP)*, pages 447–451.

Eva Pettersson, Beáta Megyesi, and Joakim Nivre. 2012. Parsing the Past - Identification of Verb Constructions in Historical Text. In *Proceedings of the 6th EACL Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 65–74, Avignon, France, April. Association for Computational Linguistics.

Eva Pettersson, Beáta Megyesi, and Joakim Nivre. 2013. An SMT approach to automatic annotation of historical text. In *Proceedings of the Workshop on Computational Historical Linguistics at NODALIDA. NEALT Proceedings Series 18; Linkping Electronic Conference Proceedings.*, volume 87, pages 54–69.

Cristina Sánchez-Marco, Gemma Boleda, and Lluís Padró. 2011. Extending the tool, or how to annotate historical language varieties. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 1–9, Portland, OR, USA, June. Association for Computational Linguistics.

Gerold Schneider. 2012. Adapting a parser to Historical English. In *Outpost of Historical Corpus Linguistics: From the Helsinki Corpus to a Proliferation of Resources*.