

Tagging Named Entities in 19th Century and Modern Finnish Newspaper Material with a Finnish Semantic Tagger

Kimmo Kettunen

The National Library of Finland
Saimaankatu 6, Mikkeli FI-50100
Kimmo.kettunen@helsinki.fi

Laura Löfberg

Department of Linguistics and English
Language, Lancaster University, UK
l.lofberg@lancaster.ac.uk

Abstract

Named Entity Recognition (NER), search, classification and tagging of names and name like informational elements in texts, has become a standard information extraction procedure for textual data during the last two decades. NER has been applied to many types of texts and different types of entities: newspapers, fiction, historical records, persons, locations, chemical compounds, protein families, animals etc. In general a NER system's performance is genre and domain dependent. Also used entity categories vary a lot (Nadeau and Sekine, 2007). The most general set of named entities is usually some version of three part categorization of locations, persons and corporations.

In this paper we report evaluation results of NER with two different data: digitized Finnish historical newspaper collection Digi and modern Finnish technology news, *Digitoday*. Historical newspaper collection Digi contains 1,960,921 pages of newspaper material from years 1771–1910 both in Finnish and Swedish. We use only material of Finnish documents in our evaluation. The OCR'd newspaper collection has lots of OCR errors; its estimated word level correctness is about 70–75%, and its NER evaluation collection consists of 75 931 words (Kettunen and Pääkkönen, 2016; Kettunen et al., 2016). *Digitoday*'s annotated collection consists of 240

articles in six different sections of the newspaper.

Our new evaluated tool for NER tagging is non-conventional: it is a rule-based Finnish Semantic Tagger, the FST (Löfberg et al., 2005), and its results are compared to those of a standard rule-based NE tagger, FiNER.

1 Introduction

Digital newspapers and journals, either OCR'd or born digital, form a growing global network of data that is available 24/7, and as such they are an important source of information. As the amount of digitized journalistic data grows, also tools for harvesting the data are needed to gather information. Named Entity Recognition has become one of the basic techniques for information extraction of texts since the mid-1990s (Nadeau and Sekine, 2007). In its initial form NER was used to find and mark semantic entities like person, location and organization in texts to enable information extraction related to this kind of material. Later on other types of extractable entities, like time, artefact, event and measure/numerical, have been added to the repertoires of NER software (Nadeau and Sekine, 2007). In this paper we report evaluation results of NER for both historical 19th century Finnish and modern Finnish. Our historical data consists of an evaluation collection out of an OCR'd Finnish historical newspaper collection 1771–1910 (Kettunen et al., 2016). Our present day

Finnish evaluation collection is from a Finnish technology newspaper Digitoday¹.

Kettunen et al. (2016) have reported NER evaluation results of the historical Finnish data with two tools, FiNER and ARPA (Mäkelä, 2014). Both tools achieved maximal F-scores of about 60 at best, but with many categories the results were much weaker. Word level accuracy of the evaluation collection was about 73%, and thus the data can be considered very noisy. Results for modern Finnish NER have not been reported extensively so far. Silfverberg (2015) mentions a few results in his description of transferring an older version of FiNER to a new version. With modern Finnish data F-scores round 90 are achieved. We use an older version of FiNER in this evaluation as a baseline NE tagger. FiNER is described more in Kettunen et al. (2016). Shortly described it is a rule-based NER tagger that uses morphological recognition, morphological disambiguation, gazetteers (name lists), as well as pattern and context rules for name tagging.

Along with FiNER we use a non-standard NER tool, a semantic tagger for Finnish, the FST (Löfberg et al., 2005). The FST is not a NER tool as such; it has first and foremost been developed for semantic analysis of full text. The FST assigns a semantic category to each word in text employing a comprehensive semantic category scheme (USAS Semantic Tagset, available in English² and also in Finnish³). The Finnish Semantic Tagger (the FST) has its origins in Benedict, the EU-funded language technology project from the early 2000s, the aim of which was to discover an optimal way of catering for the needs of dictionary users in modern electronic dictionaries by utilizing state-of-the-art language technology of the early 2000s.

The FST was developed using the English Semantic Tagger as a model. This semantic tagger was developed at the University Centre for Corpus Research on Language (UCREL) at

¹https://github.com/mpsilfve/finer-data/tree/master/digitoday/ner_test_data_annotated

<http://ucrel.lancs.ac.uk/usas/USASSemanticTagset.pdf>

³ <https://github.com/UCREL/Multilingual-USAS/raw/master/Finnish/USASSemanticTags-et-Finnish.pdf>

Lancaster University as part of the UCREL Semantic Analysis System (USAS⁴) framework, and both these equivalent semantic taggers were utilized in the Benedict project in the creation of a context-sensitive search tool for a new intelligent dictionary. The overall architecture of the FST is described in Löfberg et al. (2005) and the intelligent dictionary application in Löfberg et al. (2004).

In different evaluations the FST has been shown to be capable of dealing with most general domains which appear in a modern standard Finnish text. Furthermore, although the semantic lexical resources of the tagger were originally developed for the analysis of general modern standard Finnish, evaluation results have shown that the lexical resources are also applicable to the analysis of both older Finnish text and the more informal type of writing found on the Web. In addition, the semantic lexical resources can be tailored for various domain-specific tasks thanks to the flexible USAS category system.

Lexical resources used by the FST consist of two separate lexicons: the semantically categorized single word lexicon contains 45,871 entries and the multiword expression lexicon contains 6,113 entries, representing all parts of speech.

Our aim in the paper is twofold: first we want to evaluate whether a general computational semantic tool like the FST is able to perform a limited semantic task like NER as well as dedicated NER taggers. Secondly, we try to establish the gap on NER performance of a modern Finnish tool with 19th century low quality OCRed text and good quality modern newspaper text. These two tasks will inform us about the adaptability of the FST to NER in general and also its adaptability to tagging of 19th century Finnish that has lots of errors.

2 Results for the Historical Data

Our historical Finnish evaluation data consists of 75 931 lines of manually annotated newspaper text. Most of the data is from the last decades of 19th century. Our earlier NER evaluations with this data have achieved at best F-scores of 50–60 in some name categories (Kettunen et al., 2016).

⁴ <http://ucrel.lancs.ac.uk/usas/>

We evaluated performance of the FST and FiNER using the *conllev*⁵ script used in Conference on Computational Natural Language Learning (CONLL). *Conllev* uses standard measures of precision, recall and F-score, the last one defined as $2PR/(R+P)$, where P is precision and R recall (Manning and Schütze, 1999). Its evaluation is based on “exact-match evaluation” (Nadeau and Sekine, 2007). In this type of evaluation NER system is evaluated based on the micro-averaged F-measure (MAF) where precision is the percentage of correct named entities found by the NER software; recall is the percentage of correct named entities present in the tagged evaluation corpus that are found by the NER system. In the strict version of evaluation named entity is considered correct only if it is an exact match of the corresponding entity in the tagged evaluation corpus: a result is considered correct only if the boundaries and classification are exactly as annotated (Poibeau and Kosseim, 2001). As the FST does not distinguish multipart names with their boundaries only loose evaluation without entity boundary detection was performed with the FST.

The FST tags three different types of names: personal names, geographical names and other proper names. These are tagged with tags Z1, Z2, and Z3, respectively (Löfberg et al., 2005). Their top level semantic category in the UCREL scheme is *Names & Grammatical Words* (Z), which are considered as closed class words (Hirst, 2009, Rayson et al., 2004). Z3 is a slightly vague category with mostly names of corporations, categories Z1 and Z2 are clearly cut.

Table 1 shows results of the FST’s tagging of locations and persons in our evaluation data compared to those of FiNER. We performed two evaluations with the FST: one with the words as they are, and the other with $w \rightarrow v$ substitution. Variation of w and v is one of the most salient features of 19th century Finnish. Modern Finnish uses w mainly in foreign names like *Wagner*, but in 19th century Finnish w was used frequently instead of v in all words. In many other respects the Finnish of late 19th century does not differ

too much from modern Finnish, and it can be analyzed reasonably well with computational tools that have been developed for modern Finnish (Kettunen and Pääkkönen, 2016).

Tag	F-score FST	F- score FiNER	Found tags FST	Found tags FiNER
Pers	51.1	58.1	1496	2681
Locs	56.7	57.5	1253	1541
Pers w/v	52.2	N/A	1566	N/A
Locs w/v	61.5	N/A	1446	N/A

Table 1. Evaluation of the FST and FiNER with loose criteria and two categories in the historical newspaper collection. W/v stands for w to v substitution in words.

Substitution of w with v decreased number of unknown words to FST with about 2% units and it has a noticeable effect on detection of locations and a small effect on persons. Overall FST recognizes locations better; their recognition with w/v substitution is almost 5 per cent points better than without substitution. FST’s performance with locations outperforms that of FiNER’s slightly, but FST’s performance with person names is 7% points below that of FiNER. Performance of either tagger is not very good, which is expected as the data is very noisy.

It is evident that the main reason for low NER performance is the quality of the OCRed texts. If we analyze the tagged words with a morphological analyzer (Omorfi v. 0.3⁶), we can see that wrongly tagged words are recognized clearly worse by Omorfi than those that are tagged right. Figures are shown in Table 2.

	Locs	Pers
The FST: right tag, word unrec. rate	5.6	0.06
The FST: wrong tag, word unrec. rate	44.0	33.3

Table 2. Percentages of non-recognized words with correctly and wrongly tagged locations and persons – Omorfi 0.3

Another indication of the effect of textual quality to tagging is comparison of amount of

⁵

<http://www.cnts.ua.ac.be/conll2002/ner/bin/conllev.txt>, author Erik Tjong Kim Sang, version 2004-01-26

⁶ <https://github.com/flammie/omorfi>. This release is from year 2016.

tags with equal texts of different quality. We made tests with three versions of a 100,000 word text material that is different from our historical NER evaluation material but derives from the 19th century newspaper collection as well. One text version was old OCR, another manually corrected OCR version and third a new OCRed version. Besides character level errors also word order errors have been corrected in the two new versions. For these texts we did not have a gold standard NE tagged version, and thus we could only count number of NER tags in different texts. Results are shown in Table 3.

	Locs	Gain	Pers	Gain
Old OCR	1866		2562	
Manually corrected OCR	1986	+6.4%	2694	+5.2%
New OCR	2011	+7.8%	2879	+12.4%

Table 3. Number of the FST tags in different quality OCR texts

As the figures show, there is a 5–12% unit increase in the number of tags, when the quality of the texts is better. Although all of the tags are obviously not right, the increase is still noticeable and suggests that improvement in text quality will also improve finding of NEs. Same kind of results were achieved in Kettunen et al. (2016) with FiNER and ARPA.

NER experiments with OCRed data in other languages show usually improvement of NER when the quality of the OCRed data has been improved from very poor to somehow better (Lopresti, 2009). Results of Alex and Burns (2014) imply that with lower level OCR quality (below 70% word level correctness) name recognition is harmed clearly. Packer et al. (2010) report partial correlation of Word Error Rate of the text and achieved NER result; their experiments imply that word order errors are more significant than character level errors. Miller et al. (2000) show that rate of achieved NER performance of a statistical trainable tagger degraded linearly as a function of word error rates. On the other hand, results of Rodriguez et al. (2012) show that manual correction of OCRed material that has 88–92% word accuracy does not increase performance of four different NER tools significantly.

As the word accuracy of the historical newspaper material is low, it would be

expectable, that somehow better recognition results would be achieved, if the word accuracy was round 80–90% instead of 70–75%. Our informal tests with different quality texts suggest this, too, as do the distinctly different unrecognition rates with rightly and wrongly tagged words.

Ehrmann et al. (2016) suggest that application of NE tools on historical texts faces three challenges: i) noisy input texts, ii) lack of coverage in linguistic resources, and iii) dynamics of language. In our case the first obstacle is the most obvious, as was shown. Lack of coverage in linguistic resources e.g. in the form of missing old names in the lexicons of the NE tools is also a considerable source of errors in our case, as our tools are made for modern Finnish. With dynamics of language Ehrmann et al. refer to different rules and conventions for the use of written language in different times. In this respect late 19th century Finnish is not that different from current Finnish, but obviously also this can affect the results and should be studied more thoroughly.

3 Results of the FST and FiNER for the Digitoday Data

Our second evaluation data is modern Finnish, texts of a technology and business oriented web newspaper, *Digitoday*. NE tagged *Digitoday* data⁷ has been classified to eight different content sections according to the web publication (<http://www.digitoday.fi/>). Two sections, *Opinion* and *Entertainment*, have been left out of the tagged data. Each content section has 15–40 tagged files (altogether 240 files) that comprise of articles, one article in each file. The content sections are *yhteiskunta* (Society), *bisnes* (Business), *tiede ja teknologia* (Science and technology), *data, mobiili* (Mobile), *tietoturva* (Data security), *työ ja ura* (Job and career) and *vimpaidet* (Gadgets) We included first 20 articles of each category’s tagged data in the evaluation. *Vimpaimet* had only 15 files, which were all included in the data. Each evaluation data set includes about 2700–4100 lines of text,

⁷ https://github.com/mpsilfve/finer-data/tree/master/digitoday/ner_test_data_annotatated.

Data was collected on the first week of October and November 2016 in two parts.

altogether 31 100 lines with punctuation. About 64% of the evaluation data available in the Github repository was utilized in our evaluation – 155 files out of 240. Structure of the tagged files was simple; they contained one word per line and possibly a NER tag. Punctuation marks were tokenized on lines of their own. The resulting individual files had a few tags like

<paragraph> and <headline>, which were removed. Also dates of publishing were removed.

Table 4 shows evaluation results of the eight sections of the Digitoday data with the FST and FiNER section-by-section. Table 5 shows combined results of the eight sections and Figure 1 shows combined results graphically.

	F-score FST	F-score FiNER	Diff. FiNER vs. the FST's F-scores	Found tags FST	Found Tags FiNER
Business					
Persons	50.00	58.93	+8,93	39	67
Locations	67.14	76.47	+9,33	82	78
Corporations	31.27	65.89	+34,62	69	214
Society					
Persons	50.00	59.86	+9,86	40	79
Locations	78.26	81.63	+3,37	146	138
Corporations	34.65	62.56	+27,91	57	166
Scitech					
Persons	56.67	43.96	-12,71	32	63
Locations	57.14	57.41	+0,27	41	51
Corporations	42.49	61.93	+19,44	88	167
Data					
Persons	56.00	32.26	-23,74	29	72
Locations	64.58	67.42	+2,84	58	51
Corporations	23.85	59.21	+35,36	60	119
Mobile					
Persons	40.00	52.78	+12,78	20	52
Locations	70.77	68.49	-2,28	38	46
Corporations	50.73	62.07	+11,34	139	231
Work and career					
Persons	71.00	72.65	+1,65	79	113
Locations	68.09	76.09	+8,00	51	49
Corporations	48.67	74.64	+25,97	71	151
Data security					
Persons	40.65	46.71	+6,06	47	91
Locations	83.05	83.93	+0,88	66	60
Corporations	29.70	39.08	+9,38	60	119
Gadgets					
Persons	84.62	50.00	-34,62	13	35
Locations	48.28	66.67	+18,39	19	20
Corporations	57.87	68.29	+10,42	90	142

Table 4. Results of FiNER and the FST with Digitoday's data section-by-section

Sections combined	F-score FST	F-score FiNER	Diff. FiNER vs. FST's F-scores	Found tags FST	Found tags FiNER
Persons	56.15	55.39	-0.76	299	572
Locations	70.77	74.58	+3.81	501	493
Corporations	40.11	62.56	+22.45	634	1309
Total for the eight sections				1434	2374

Table 5. Combined results of all Digitoday's sections

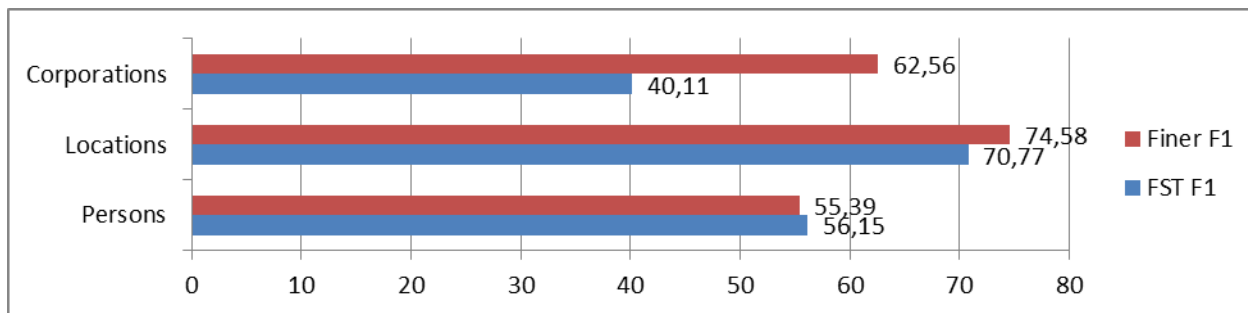


Figure 1. Combined results of all Digitoday’s sections

FiNER achieves best F-scores in most of Digitoday’s sections. Out of all the 24 cases, FiNER performs better in 20 cases and the FST in four. The FST performs worst with corporations, but differences with locations compared to FiNER are small. Performance differences with persons between FiNER and the FST are also not that great, and the FST performs better than FiNER in three of the sections.

Both taggers find locations best and quite evenly in all the Digitoday’s sections. Persons are found varyingly by both taggers, and section wise performance is uneven. Especially bad they are found in the Data and Data security sections. One reason for FiNER’s bad performance in this section is that many products are confused to persons. In Business and Society sections, persons are found more reliably. One reason for the FST’s varying performance with persons is variance of section-by-section usage of Finnish and non-Finnish person names. In some sections mainly Finnish persons are discussed and in some sections mainly foreign persons. The FST recognizes Finnish names relatively well, but it does not cover foreign names as well. The morphological analyzer’s components in the FST are also lexically quite old, which shows in some lacking name analyses, such as *Google*, *Facebook*, *Obama*, *Twitter*, if the words are in inflected forms.

4 Discussion

We have shown in this paper results of NE tagging of both historical OCRed Finnish and modern digital born Finnish with two tools, FiNER and a Finnish Semantic Tagger, the FST. FiNER is a dedicated rule-based NER tool for Finnish, but the FST is a general lexicon-based semantic tagger.

We set a twofold task for our evaluation. Firstly we wanted to compare a general

computational semantics tool the FST to a dedicated NE tagger in named entity search. Secondly we wanted to see, what is the approximate decrease in NER performance of modern Finnish taggers, when they are used with noisy historical Finnish data. Answer to the first question is clear: the FST performs mostly as well as FiNER with persons and locations in modern data. With historical data FiNER outperforms the FST with persons; with locations both taggers perform equally. Corporations were not evaluated in the historical data. In Digitoday’s data FiNER was clearly better than FST with corporations.

Answer to our second question is more ambiguous. With historical data both taggers achieve F-scores round 57 with locations, the FST 61.5 with w/v substitution. With Digitoday’s data F-scores of 70–74.5 are achieved, and thus there is 9–16% point difference in the performance. With persons FiNER’s score with both data are quite even in average. It would be expectable, that FiNER performed better with modern data. Some sections of Digitoday data (*Scitech*, *Data*, *Data Security*) are performing clearly worse than others, and FiNER’s performance only in *Work and Career* is on expectable level. It is possible that section wise topical content has some effect in the results. The FST’s performance with persons is worst with historical data, but with Digitoday’s data it performs much better.

Technology behind FST is relatively old, but it has a sound basis and a long history. Since the beginning of the 1990s and within the framework of several different projects, the UCREL team has been developing an English semantic tagger, the EST, for the annotation of both spoken and written data with the emphasis on general language. The EST has also been redesigned to create a historical semantic tagger for English.

To show versatility of the UCREL semantic tagger approach, we list a few other computational analyses where the EST has been applied to:

- stylistic analysis of written and spoken English
- analysis and standardization of SMS spelling variation,
- analysis of the semantic content and persuasive composition of extremist media,
- corpus stylistics,
- discourse analysis,
- ontology learning,
- phraseology,
- political science research,
- sentiment analysis, and
- deception detection.

More applications are referenced on UCREL's web pages (<http://ucrel.lancs.ac.uk/usas/>; <http://ucrel.lancs.ac.uk/wmatrix/#apps>).

As can be seen from the list, the approach taken in UCREL that is also behind the FST is robust with regards to linguistic applications. Based on our results with both historical and modern Finnish data, we believe that the EST based FST is also a relevant tool for named entity recognition. It is not optimal for the task in its present form, as it lacks e.g. disambiguation of ambiguous name tags at this stage⁸. On the other hand, the FST's open and well documented semantic lexicons are adaptable to different tasks as they can be updated relatively easily. The FST would also benefit from an updated open source morphological analyzer. Omorfi⁹, for example, would be suitable for use, as it has a comprehensive lexicon of over 400 000 base forms. With an up-to-date Finnish morphological analyzer and disambiguation tool the FST would yield better NER results and in the same time it would be a versatile multipurpose semantical analyzer of Finnish.

Overall our results show that a general semantic tool like the FST is able to perform in a restricted semantic task of name recognition almost as well as a dedicated NE tagger. As NER

is a popular task in information extraction and retrieval, our results show that NE tagging does not need to be only a task of dedicated NE taggers, but it can be performed equally well with more general multipurpose semantic tools.

Acknowledgments

Work of the first author is funded by the EU Commission through its European Regional Development Fund, and the program Leverage from the EU 2014–2020.

References

- Maud Ehrmann, Giovanni Colavizza, Yannick Rochat, and Frédéric Kaplan. 2016. Diachronic Evaluation of NER Systems on Old Newspapers. In Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016), 97–107. https://www.linguistics.rub.de/konvens16/pub/13_konvensproc.pdf
- Graeme Hirst. 2009. Ontology and the Lexicon. <ftp://ftp.cs.toronto.edu/pub/gh/Hirst-Ontol-2009.pdf>
- Kimmo Kettunen and Tuula Pääkkönen. 2016. Measuring Lexical Quality of a Historical Finnish Newspaper Collection – Analysis of Garbled OCR Data with Basic Language Technology Tools and Means. *LREC 2016, Tenth International Conference on Language Resources and Evaluation*. http://www.lrec-conf.org/proceedings/lrec2016/pdf/17_Paper.pdf.
- Kimmo Kettunen, Eetu Mäkelä, Juha Kuokkala, Teemu Ruokolainen and Jyrki Niemi. 2016. Modern Tools for Old Content - in Search of Named Entities in a Finnish OCRed Historical Newspaper Collection 1771-1910. Krestel, R., Mottin, D. and Müller, E. (eds.), *Proceedings of Conference "Lernen, Wissen, Daten, Analysen"*, LWDA 2016, <http://ceur-ws.org/Vol-1670/>
- Laura Löfberg, Jukka-Pekka Juntunen, Asko Nykänen, Krista Varantola, Paul Rayson and Dawn Archer. 2004. Using a semantic tagger as dictionary search tool. In 11th EURALEX (European Association for Lexicography) International Congress Euralex 2004: 127–134.

⁸ Many lexicon entries contain more senses than one, and these are arranged in perceived frequency order. For example, it is common in Finnish that a name of location is also a family name.

⁹ <https://github.com/flammie/omorfi>

- Laura Löfberg, Scott Piao, Paul Rayson, Jukka-Pekka Juntunen, Asko Nykänen and Krista Varantola. 2005. A semantic tagger for the Finnish language. http://eprints.lancs.ac.uk/12685/1/cl2005_fst.pdf
- Daniel Lopresti. 2009. Optical character recognition errors and their effects on natural language processing. *International Journal on Document Analysis and Recognition*, 12(3): 141–151.
- Eetu Mäkelä. 2014. Combining a REST Lexical Analysis Web Service with SPARQL for Mashup Semantic Annotation from Text. Presutti, V. et al. (Eds.), *The Semantic Web: ESWC 2014 Satellite Events*. Lecture Notes in Computer Science, vol. 8798, Springer: 424–428.
- Christopher D. Manning and Hinrich Schütze. 1999. *Foundations of Statistical Language Processing*. The MIT Press, Cambridge, Massachusetts.
- Mónica Marrero, Julián Urbano, Sonia Sánchez-Cuadrado, Jorge Morato and Juan Miguel Gómez-Berbís. 2013. Named Entity Recognition: Fallacies, challenges and opportunities. *Computer Standards & Interfaces* 35(5): 482–489.
- David Miller, Sean Boisen, Richard Schwartz, Rebecca Stone and Ralph Weischedel. 2000. Named entity extraction from noisy input: Speech and OCR. *Proceedings of the 6th Applied Natural Language Processing Conference*: 316–324, Seattle, WA. <http://www.anthology.aclweb.org/A/A00/A00-1044.pdf>
- David Nadeau and Satoshi Sekine. 2007. A Survey of Named Entity Recognition and Classification. *Linguisticae Investigationes* 30(1):3–26.
- Thomas L. Packer, Joshua F. Lutes, Aaron P. Stewart, David W. Embley, Eric K. Ringger, Kevin D. Seppi and Lee S. Jensen. 2010. Extracting Person Names from Diverse and Noisy OCR Text. *Proceedings of the fourth workshop on Analytics for noisy unstructured text data*. Toronto, ON, Canada: ACM. <http://dl.acm.org/citation.cfm?id=1871845>.
- Scott Piao, Paul Rayson, Dawn Archer, Francesca Bianchi, Carmen Dayrell, Mahmoud El-Haj, Ricardo-María Jiménez, Dawn Knight, Michal Kren, Laura Löfberg, Rao Muhammad Adeel Nawab, Jawad Shafi, Phoey Lee Teh and Olga Mudraya. 2016. Lexical Coverage Evaluation of Large-scale Multilingual Semantic Lexicons for Twelve Languages. *Proceedings of LREC*. http://www.lrec-conf.org/proceedings/lrec2016/pdf/257_Paper.pdf.
- Thierry Poibeau and Leila Kosseim. 2001. Proper Name Extraction from Non-Journalistic Texts. *Language and Computers*, 37: 144–157.
- Paul Rayson, Dawn Archer, Scott Piao and Tony McEnery. 2004. The UCREL semantic analysis system. *Proceedings of the workshop on Beyond Named Entity Recognition Semantic labelling for NLP tasks in association with 4th International Conference on Language Resources and Evaluation (LREC 2004)*: 7–12. http://www.lancaster.ac.uk/staff/rayson/publications/usas_lrec04ws.pdf
- Kepa Joseba Rodriquez, Mike Bryant, Tobias Blanke and Magdalena Luszczynska. 2012. Comparison of Named Entity Recognition Tools for raw OCR text. *Proceedings of KONVENS 2012 (LThist 2012 wordshop)*, Vienna September 21: 410–414.
- Miikka Silfverberg. 2015. Reverse Engineering a Rule-Based Finnish Named Entity Recognizer. https://kitwiki.csc.fi/twiki/pub/FinCLARIN/KielipankkiEventNERWorkshop2015/Silfverberg_presentation.pdf.