

Automatic Recognition of Abbreviations and Abbreviations' Expansions in Multilingual Electronic Texts

Mojca Kompara

Department of Comparative and General Linguistics, Faculty of Arts,
University of Ljubljana, Slovenia

The paper presents universal rules for creating an algorithm for automatic recognition of abbreviations and abbreviations' expansions in texts written in various languages by focusing on how universal the expansion patterns are, and testing them on multiple languages to see how reliable such patterns are (Zahariev 2004).

As a first step, a demonstration algorithm has been used which focuses on the automatic recognition of abbreviations and abbreviations' expansions in just one language (Slovene) and with a restricted number of characters for each abbreviation. Further development expands the number of characters for each abbreviation and takes into consideration all four types of abbreviation-expansion patterns. A random selection of Slovene text is used to verify the performance of the algorithm and to improve recognition. The universality of rules is taken into consideration by the presence of foreign abbreviation-expansion pairs that exist in Slovene texts and is verified on random foreign texts.

The final version of the algorithm filters texts coming from the *Europarl corpus*, covering all EU languages and bearing in mind specialised characteristics of each language. The acquired data can be used for (semi)automatic production of a dictionary of abbreviations. Such a dictionary represents the future of electronic lexicography (Kompara 2009: 109).

1. INTRODUCTION

Abbreviations are difficult to deal with (Gabrovšek 1994: 23) and represent a growing phenomenon present in all languages. The scope of this article is to present automatic recognition of abbreviations and abbreviations' expansions in multilingual electronic texts. The paper presents universal rules (Zahariev 2004: 10) for creating an algorithm for automatic recognition of abbreviations and abbreviations' expansions in texts written in various languages by focusing on how universal the expansion patterns are, and then testing them on multiple languages to see how reliable such patterns are.

2. DICTIONARIES ARE NO LONGER ENOUGH

In this chapter we present how abbreviations are dealt in dictionaries. We present the insufficient and inefficient inclusion of abbreviations in dictionaries by focusing on the inclusion of abbreviations in monolingual, bilingual, specialised and online dictionaries. All bilingual dictionaries are applicable to the Slovene language and special emphasis is placed on the microstructure of specialised abbreviation dictionaries. Abbreviations are usually included among the entries or/and in supplements of monolingual dictionaries of English *Collins COBUILD English Dictionary* (Sinclair 1999), Italian *Vocabolario della lingua italiana* (Zingarelli 2000), German *Deutsches Universal Wörterbuch* (Drosdowski 1989), Spanish *Clave: Diccionario de Uso del Español* (García Marquez 2002) and French *Le Nouveau Petit Robert* (Robert 1996) but they are not included among the entries of the Slovene monolingual dictionary *Slovar slovenskega knjižnega jezika* (Bajec *et al.* 1970–

1991). As far as bilingual dictionaries are concerned, abbreviations are usually not included in older publications of dictionaries for encoding, such as *Slovensko-angleški slovar* (Grad & Leeming 1997), *Slovensko-italijanski slovar* (Kotnik 1992), *Slovensko-španski slovar* (Grad 2000) and *Slovensko-francoski slovar* (Jesenik 2005) but the exceptions are some new publications *Veliki slovensko-italijanski slovar* (Šlenc 2006) and *Veliki slovensko-nemški slovar* (Debenjak 2003). On the other hand, abbreviations are included in dictionaries for decoding, such as *Veliki angleško-slovenski slovar* (Grad 1998), *Veliki angleško-slovenski slovar Oxford-DZS* (Gabrovšek et al. 2005–2006), *Veliki italijansko-slovenski slovar* (Šlenc 1997), *Veliki nemško-slovenski slovar* (Debenjak 2001), *Špansko-slovenski slovar* (Grad 2001) and *Francosko-slovenski slovar* (Grad 2004). The main problem of all the mentioned dictionaries is the fact that they are not published frequently enough to cover all new abbreviations among their entries, and that abbreviations are not covered extensively.

2. 1. English dictionaries of abbreviations

In English several dictionaries of abbreviations (listed below) were observed. The characteristics of microstructure are visible in table 1.

Index: X – no; \sqrt{X} - yes (not consistently); \sqrt{R} - yes (rarely); $\sqrt{\sqrt{}}$ - yes (consistently); \sqrt{E} - yes (extensive); $\sqrt{\sqrt{}}$ - yes; \sqrt{TE} - yes (too extensive)

(Table 1) *English dictionaries of abbreviations*

Name	Entries	Suppl.	Foreign abbr.	Transl.	Qualifier	Lang.	Encyclo. data
(Paxton 1983)	25,000	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	\sqrt{X}	$\sqrt{\sqrt{}}$
(Fergusson 2000)	40,000	\sqrt{E}	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$
(De Sola 1986)	23,000	\sqrt{E}	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	X	$\sqrt{\sqrt{}}$	\sqrt{R}
(Buttress 1976)	60,000	X	$\sqrt{\sqrt{}}$	X	X	X	X
(Dale & Puttick 1999)	20,000	\sqrt{TE}	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$
(Geddes & Grosset 1999)	10,000	X	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	X
(Gale Research Company 2006)	1,000,000	X	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	X	$\sqrt{\sqrt{}}$
(Barnhart 1995)	60,000	X	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$
(Jung 1991)	50,000	X	$\sqrt{\sqrt{}}$	X	\sqrt{X}	X	X
(Benedetto Mattia 1997)	60,000	$\sqrt{\sqrt{}}$	$\sqrt{\sqrt{}}$	X	X	X	X

As seen from the table the analysed and presented dictionaries differ mainly in size but also in microstructure, especially in terms of translations which are not always present in entries or not present at all. Translation of abbreviations is an essential item that should always be consistently included in specialised dictionaries. A dictionary without translations does not solve problems but causes new ones. As seen from the table above, some dictionaries do not have supplements while others have overly extensive ones. There is also a lot of inconsistency in terms of qualifiers, language and encyclopaedic data. Reverse dictionaries or supplements such as *Reverse Acronyms Initialisms & Abbreviations Dictionary* (Gale Research Company 2007) or (Barnhart 1995) are extremely useful for users but still missing in most dictionaries. The main problem of all dictionaries is the fact that they are too old and not published frequently enough to include new abbreviations.

2.2. German, Italian, Spanish and French dictionaries of abbreviations

Some Italian, German, Spanish and French dictionaries of abbreviations (listed below) were observed. The characteristics of microstructure are visible in table 2.

Index: X – no; √X - yes (not consistently); √R - yes (rarely); √√ - yes (consistently); √E - yes (extensive); √ - yes; √TE - yes (too extensive)

(Table 2) *German, Italian, Spanish and French dictionaries of abbreviations*

Name	Entries	Suppl.	Foreign abbr.	Translation	Qualifier	Lang.	Encyclop. data
(Steinhauer 2005)	50,000	√E	√	X	√	√X	X
(Koblischke 1983)	50,000	√E	√	√	√	√	√
(Righini 2001)	10,000	X	√	√	√	√	√
(Malossini 1999)	8,000	X	√	X	√X	X	X
(Murith & Bocabeille 1992)	20,000	X	√	X	X	X	X
(Galende 1997, 2001)	20,000	X	X	X	X	X	X

There are even more differences in terms of microstructure in German, Italian, Spanish and French dictionaries of abbreviations presented above compared to English dictionaries. This concerns especially translations that should be essential data in such dictionaries too. We notice also that all observed dictionaries are old and do not include new abbreviations and include fewer entries compared to the English dictionaries. As far as Slovene is concerned there is still no updated abbreviation dictionary. The first one *Kratice, Mala izdaja* (Župančič) was written in 1948. Fortunately we have an online dictionary of abbreviations *Slovarček krajšav*,¹ containing over 5,000 entries and providing translations of all foreign abbreviations. Abbreviations represent a growing phenomenon present in all languages. New ones arise suddenly and paper dictionaries do not dedicate enough space and attention to them, and such dictionaries are not published frequently enough to cover new abbreviations. Online dictionaries such as *Slovarček krajšav*,² *Evroterm*,³ *Acronym Finder*,⁴ *The Free Dictionary*,⁵ and others represent an alternative to paper ones and are easily updated. But unfortunately abbreviations in online dictionaries are acquired mainly manually which is still time consuming and not precise enough. The alternative is automatic recognition of abbreviations and abbreviations' expansions from electronic texts.

3. ABBREVIATIONS RECOGNISED AUTOMATICALLY

The pioneer in automatic recognition of abbreviations and abbreviations' expansion is Taghva (1998). Automatic recognition was also dealt with by Yeast (1999), Larkey *et al.* (2000),

¹ <http://bos.zrc-sazu.si/kratice.html>.

² <http://bos.zrc-sazu.si/kratice.html>.

³ <http://www.sigov.si/evroterm/>.

⁴ <http://www.acronymfinder.com>.

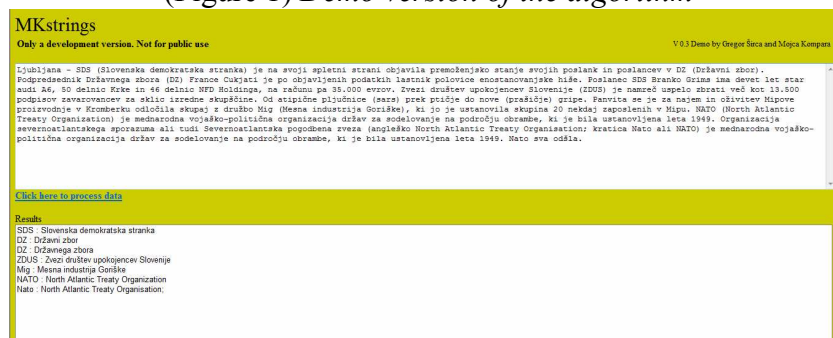
⁵ <http://acronyms.thefreedictionary.com/>.

Pustejovsky *et al.* (2001), Schwartz and Hearst (2003), Park and Byrd (2001), Chang *et al.* (2002) and Zahariev (2004), but his approach is considered special due to the fact that he was not limited to just one language's recognition. In the early stage of the research, words of up to 5 capital letters written in brackets were used as abbreviations' candidates, e.g. (NATO). The reference was Slovene online newspaper *Delo* from 2007. It had 25,588 such words and some occurred more than once. In order to obtain an appropriate amount of abbreviations for further research, all abbreviations occurring more than once and all abbreviations that are not acronyms and/or abbreviations written in capital letters were excluded. Words that are not abbreviations, such as proper names, names of places etc. were excluded, using the *Slovene monolingual dictionary*; after the exclusions the database had over 4,000 abbreviations. After the exclusion of all abbreviations not written in capital letters and considering also the context that should not be written in capital letters, the final list had 2,500 abbreviations. In order to obtain the abbreviations' expansions from the newspaper *Delo*, left context was observed, because expansions are usually placed before the abbreviations, e.g. *European central bank (ECB)*, but still not excluding the possibility of right context. To recognize expansions, 4 types of abbreviations were used. The first type are the so called *covered abbreviations* where letters match the words in left context, e.g. *FF* with the expansion *Filozofska fakulteta*. The second type are *abbreviations with expansions containing prepositions and conjunctions*, e.g. *FDV Fakulteta za družbene vede*. The algorithm takes into consideration also expansions with one or two additional words. The third type concerns *abbreviations composed of the first two letters*, e.g. *NAMA Narodni magazin*. The fourth type covers *abbreviations with prepositions*, e.g. *DZU Družba za upravljanje* where prepositions appear in the abbreviation and also in the expansion. Considering the above-mentioned criteria, 1,800 expansions matched the abbreviations. In the final list containing abbreviation-expansion pairs several problems were observed, such as the occurrence of cases, the multiple occurrences of the same expansion and foreign abbreviations with Slovene expansions.

4. FURTHER DEVELOPMENT AND DEMO PRESENTATION

In the second stage of the development the number of letters in the abbreviations was extended to 10 and right context was observed. All four types of pattern: (*abbreviation*) *expansion*, (*expansion*) *abbreviation*, *abbreviation* (*expansion*), and *expansion* (*abbreviation*) were used. Abbreviations with more than 10 letters were not included and lexicalized abbreviations were also excluded but such abbreviations are well-known and are included in the algorithm via the dictionary of abbreviations *Slovarček krajšav*. After the newly established rules for recognition a demo version of the algorithm was produced. The system called *MKstrings* is composed of two windows. In the first one we add text rich in abbreviations, after clicking *Click here to process data*. In the second window abbreviations and expansions occur as seen in figure 1.

(Figure 1) Demo version of the algorithm



As seen from figure 1, the algorithm is not taking into consideration abbreviations such as *sars*, *Mipu*, *NFD*, *A6*, which is expected according to the rules stated above. Although at first sight the obtained results look really good, the algorithm needs further improvement. To improve the algorithm randomly selected texts rich in abbreviations (from the website *24ur.com*) were used in order to observe how the algorithm behaves. Problems occurred mainly in examples containing the abbreviation e.g. *RS* in the expansion and non-capitalised abbreviation e.g. *DARS* in *Družba za avtoceste v RS (Dars)*. But after taking this into consideration, the problem was still not solved. Prepositions *za* and *v* represented a problem too, because at the present stage the algorithm was able to consider just one preposition in the expansion. Problems occurred also in some copy-pasted examples e.g. *Urada za varstvo konkurence (UVK)*, recognised when retyped. The algorithm does not recognise abbreviations with no expansions. Such examples are automatically excluded. An interesting issue also concerns patterns composed of a foreign abbreviation and a Slovene expansion, e.g. *Združenje evropskih avtomobilskih proizvajalcev (ACEA)*. Such patterns were not observed in the present article and will be recognised in the future. After applying modifications and improvements the software was enlarged in order to be able to filter larger amounts of data. A larger corpus composed of 60 million words (newspaper *Delo* from 2005 to 2009) was used. The algorithm filtered the corpus in 30 minutes and gave 5,820 abbreviation-expansion pairs. The obtained pairs were manually revised and verified using *Google*. Among the revised and verified pairs 4% of false pairs occurred, e.g. *PO predstavljenih podatkih o, NA na vse argumente, IN in novincev*. The precision of the algorithm is 96%. Among the revised and verified pairs were also genuine abbreviations not matching with the right expansions, e.g. *HIV virusom i hepatitisom in virusom* because the expansions were missing. Among the good expansions many occurred more than once and/or with tiny modifications, e.g. usage of different cases or spelling as seen in table 3. In table 3 just 3 expansions out of 6 are genuine.

(Table 3) *Expansions for MNZ*

MNZ	MNZ
1 ministrstva za notranje zadeve	1 ministrstva za notranje zadeve
2 medobčinskih nogometnih zvez	2 medobčinskih nogometnih zvez
3 ministrstvom za notranje zadeve	3 Muzej novejšje zgodovine
4 Medobčinske nogometne zveze	
5 Muzeja novejšje zgodovine	
6 Muzej novejšje zgodovine	

After the exclusion of false pairs, verification and revision of good pairs 2,665 genuine abbreviations-expansion pairs occurred. Among the good pairs there were also some foreign pairs although the recognition focused only on Slovene texts. Among the foreign pairs, some problems occurred in misrecognition of parts of expansions, e.g. *FEE for Environmental Education*, where *Foundation* is missing.

5. RECOGNITION IN FOREIGN LANGUAGES

Among the Slovene analysed texts the algorithm recognised also some foreign abbreviations. To discover whether the algorithm is universal⁶ some English and Italian texts randomly selected and available online were used. In English texts the algorithm recognised patterns such as *Severe acute respiratory syndrome (SARS)*, *World Health Organization (WHO)*, *Center for Group Learning (CGL)*, and *Scottish Licensed Trade News (SLTN)* but did not recognise patterns having abbreviations in the expansion, e.g. *USB Implementers Forum (USB-IF)* and abbreviations written in a special way and/or expansions with punctuation, e.g.

⁶ Such an approach was also used by Zahariev (2004).

Human-interface devices (HIDs). In Italian the situation is similar, the algorithm recognised the following patterns; *Confederazione Generale del Lavoro (CGdL)*, *Popolo della Libertà (PdL)*, *Centro Interdipartimentale di Studi Europei e Mediterranei (CISEM)* but it did not recognise patterns such as *Azienda Nazionale Autonoma delle Strade (ANAS)* because in the text *ANAS* is preceded by the definite article *L'ANAS*. The algorithm does not recognise patterns composed of articles combined with prepositions (*preposizione articolata*), e.g. *della*, *per la*, as seen in the following examples; *Dipartimento di Scienze del Linguaggio, dell'Interpretazione e della Traduzione (DSLIT)*, *Centro Interdipartimentale per la Ricerca Didattica (CIRD)*. The final version of the algorithm filters texts coming from the *Europarl corpus*, covering all EU languages and bearing in mind specialised characteristics of each language. This approach aims to show how universal the expansion patterns are and how reliable such patterns are. For the purpose of this article, texts in Italian, Spanish, French and Slovak were filtered. In all languages the algorithm recognised also foreign patterns e.g. *ADR (Australian Design Rules)* in the Italian text etc.

Italian

In recognising abbreviation-expansion pairs in foreign languages, typology of the foreign language is of extreme importance. In Italian abbreviation-expansion pairs, typological characteristics of the Italian language were taken into consideration. Several stop lists were used in order to come across genuine pairs. A stop list of prepositions e.g. *da*, *a*, *su* etc., conjunctions, definite and indefinite articles and combinations of prepositions and articles called *preposizioni articolate* e.g. *della*, *nella* etc. were used in recognition. After recognition false pairs were manually eliminated but the main problem occurred in pairs where it was difficult to establish an official abbreviation (fixed) and an abbreviation used just once for the purpose of shortening text (unfixed). Abbreviations with more than one meaning were frequent and without experts in the field the right expansion cannot be detached as seen in table 4 where the algorithm gives several possibilities for the abbreviation *CE*.

(Table 4) *Abbreviations with more than one meaning*

It	CE	Comunità europea
It	CE	certificato di esame
It	CE	conformità alle esigenze
It	CE	componenti elettronici

French

Typological differences can be seen also in French. As in the Italian text a stop list of preposition, definite and indefinite articles and conjunction were used in order to provide appropriate recognition. But as seen in table 5 linguistic and non-linguistic expertise is still essential in discerning genuine pairs.

(Table 5) *Abbreviations with more than one meaning*

Fr	SPG	Système de préférences généralisées
Fr	SPG	système des préférences généralisés
Fr	SPG	schéma de préférences généralisées
Fr	SPG	système des préférences généralisées

Spanish

A similar approach was also needed for Spanish texts where in order to obtain genuine pairs stop lists of prepositions, conjunctions and articles were used. The need of experts in establishing genuine pairs remains, as seen from table 6.

(Table 6) *Abbreviations with more than one meaning*

es	FEOGA	Fondo Europeo de Orientación y Garantía Agrícola
es	FEOGA	Fondo Europeo de Orientación y Garantía Agraria

Slovak

Slovak, like Slovene, is a Slavic language and has a completely different typology compared to Romance languages. In Slovak, as in all the Romance languages, a stop list of prepositions and conjunctions was built, but due to the absence of definite or indefinite articles such a list was not needed for articles. On the other hand the occurrence of cases represented a general problem common to all Slavic languages. Multiple occurrences of the same expansion with different cases are seen in the table below. For such cases an automatic approach should be used also for the recognition of just nominative cases. Such an approach is in preparation for Slovene but not yet for Slovak. In Slovene we use the analyser module from *Presis* machine translation software to translate the texts to *Presis Interlingua*. For this task, the analyser module was changed in order to accept just noun phrases in various cases instead of both sentences and noun phrases. The result in *Interlingua* is checked to find out if the noun phrase is nominative. If it is not, the *Interlingua* result is changed to nominative and sent to the generator module of *Presis* to translate *Interlingua* back to Slovene. Furthermore the number of the noun phrase is also checked since nominative dual or plural forms can be the same as non-nominative singular forms (*Alzheimerjeve bolezni*). But unfortunately such an approach works only for the Slovene language.

(Table 7) *Abbreviations with more than one meaning*

sk	AKT	afričkih karibských a tichomorských
sk	AKT	alebo zámorských krajín a teritórií
sk	AKT	afrikké karibské a tichomorské
sk	AKT	africkej karibskej a tichomorskej
sk	AKT	Afriky Karibiku a Tichomoria
sk	EHS	Európskym hospodárskym spoločenstvom
sk	EHS	Európskeho hospodárskeho spoločenstva Dohoda
sk	EHS	Európskeho hospodárskeho spoločenstva
sk	EHS	Európske hospodárske spoločenstvo
sk	EHS	Európskemu hospodárskemu spoločenstvu

As seen from the filtered foreign language texts the main problem concerns the typology of the foreign language and the need for manual cleaning of good examples, fixed and non-fixed pairs. If we want to come across genuine pairs the need of experts in the field is also essential. In the future development data from a multilingual corpus can be used for automatic production of the dictionary of abbreviations. At present the automatic approach is used just

for the production of a Slovene dictionary of abbreviations in *Termania* software.⁷ The production of simple Slovene entries e.g. FDV (*Fakulteta za družbene vede*) where the abbreviation-expansion pair and nominative structure of the expansion are needed is already possible. The production of complex entries (covering qualifiers, language qualifiers and encyclopaedic data) is still under development.

6. CONCLUSION

Abbreviations are not something new or a fashionable way of communication; they were used even by Cicero (Kompara 2005: 80). Abbreviations became part of our everyday life and are produced on a daily basis. They are rapidly growing in number, but as seen in the analysis, neither paper nor online dictionaries are published or updated frequently enough to cover new ones among their entries. Manual acquisition of abbreviation data and manual inclusion of abbreviations into online dictionaries is time-consuming; the solutions are algorithms for automatic recognition of abbreviations and abbreviations' expansions in electronic texts. In the present paper such a system is described and the demo version of the algorithm is presented. The paper describes the development of rules, the algorithm and the software. At a present stage the algorithm is applicable just for Slovene but the occurrence of foreign pairs introduced the idea of multilingual recognition. The *Europarl corpus* was filtered and results for Italian, French, Spanish and Slovak were described. Such an algorithm raises the possibility of creating a semi-automatic dictionary of abbreviations and such a dictionary represents the future of electronic lexicography. Algorithms for automatic recognition of abbreviations present the link between the text and the semi-automatic production of a dictionary of abbreviations (Kompara 2009: 109). That is why the production and further development of the algorithm is essential and useful.

REFERENCES

- Amebis, 1995. *Termania*. [online] Available at: <http://www.termania.net/> [Accessed 16 February 2011].
- Bajec, A. et al., 1970-1991. *Slovar slovenskega knjižnega jezika*. Ljubljana: DZS.
- Barnhart, Robert K. 1995. *The Barnhart Abbreviations Dictionary*. [online] Available at: <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471571466.html> [Accessed 16 February 2011].
- Benedetto, M. F., 1997. *Elsevier's Dictionary of Acronyms, Initialisms, Abbreviations and Symbols*. Amsterdam: Elsevier Science B.V.
- Buttress, F. A., 1976. *World Guide to Abbreviations of Organizations*. Detroit: Grand River Books.
- Chang, J. T., et al. 2002. Creating an Online Dictionary of Abbreviations from MEDLINE. *Journal of American Medical Informatics Association (JAMIA)*, IX(VI), pp. 612-620.
- Dale, R. and Puttick, S., 1999. *Dictionary of Abbreviations and Acronyms*. Hertfordshire: Wordsworth.
- Debenjak, B., 2001. *Veliki nemško-slovenski slovar*. Ljubljana: DZS.
- Debenjak, B., 2003. *Veliki slovensko-nemški slovar*. Ljubljana: DZS.
- De Sola, R., 1986. *Abbreviations Dictionary*, 7th ed. New York, Amsterdam, Oxford: Elsevier.
- Drosdowski, G., 1989. *Deutsches Universal Wörterbuch*. Mannheim: Duden Verlag.
- Farlex, 2004. *The Free Dictionary*. [online] Available at: <http://acronyms.thefreedictionary.com> [Accessed 16 February 2011].

⁷

The dictionary is not yet available online.

- Fergusson, R., 2000. *The New Penguin Dictionary of Abbreviations; from A to ZZ*. London: Penguin.
- Gabrovšek, D., 1994. Kodifikacija angleškega jezika v specializiranih enojezičnih slovarjih: Too much of everything? *Vestnik*, XXVII/I–II, pp. 150–180.
- Gabrovšek, D. et al. 2005–2006. *Veliki angleško-slovenski slovar Oxford*. Ljubljana: DZS.
- Gale Cengage Learning, 2006. *Acronyms, Initialisms and Abbreviations Dictionary*. [online] Available at: <http://www.gale.cengage.com/pdf/samples/sp628573.pdf> [Accessed 16 February 2011].
- Gale Cengage Learning, 2006. *Reverse Acronyms Initialisms & Abbreviations Dictionary*. [online] Available at: <http://www.gale.cengage.com/pdf/samples/sp628662.pdf> [Accessed 16 February 2011].
- Galende, J. C., 1997, 2001. *Diccionario General de Abreviaturas Españolas*. Madrid: Editorial Verbum.
- Garcia Marquez, G., 2002. *Clave: Diccionario de Uso del Español*. Madrid: Ediciones Sm.
- Geddes and Grosset, 1999. *Abbreviations and Acronyms*. New Lanark: Geddes and Grosset.
- Google. [online] Available at: <http://www.google.si/> [Accessed 16 February 2011].
- Grad, A. and Leeming, H., 1997. *Slovensko-angleški slovar*. Ljubljana: DZS.
- Grad, A., 1998. *Veliki angleško-slovenski slovar*. Ljubljana: DZS.
- Grad, A., 2000. *Slovensko-španski slovar*. Ljubljana: DZS.
- Grad, A., 2001. *Špansko-slovenski slovar*. Ljubljana: DZS.
- Grad, A., 2004. *Francosko-slovenski slovar*. Ljubljana: DZS.
- Jesenik, V., 2005. *Slovensko-francoski slovar*. Ljubljana: DZS.
- Jung, H., 1991. *The Dictionary of Acronyms and Abbreviations in Applied Linguistics and Language Learning*. New York: Peter Lang Publishing Group.
- Koblischke, H., 1983. *Großes Abkürzungsbuch*. Leipzig: VEB Bibliographisches Institut Leipzig.
- Kompara, M., 2005. *Slovensko-italijanski glosar krajšav*. B.A. Ljubljana: University of Ljubljana, Faculty of Arts, Department of Translation.
- Kompara, M. 2006. *Slovarček krajšav*. [online] Available at: <http://bos.zrc-sazu.si/kratice.html> [Accessed 16 February 2011].
- Kompara, M., 2009. Prepoznavanje krajšav v slovenskih elektronskih besedilih. *Jezikoslovni zapiski* 15. 1-2, pp. 95-112.
- Kotnik, J., 1992. *Slovensko-italijanski slovar*. Ljubljana: DZS.
- Larkey, L. S., et al. 2000. Acrophile: An Automated Acronym Extractor and Server. *Proceedings of the fifth ACM conference on Digital libraries*, pp. 205-214.
- Malossini, A., 1999. *Dizionario delle sigle e degli acronimi*. Milano: Avallardi.
- Molloy M., 1996. *Acronym Finder*. [online] Available at: <http://www.acronymfinder.com> [Accessed 16 February 2011].
- Murith, J. and Bocabeille. J. M., 1992. *Dictionnaire des Abreviations et Acronymes*. Paris: TEC&DOC – Lavoisier.
- Park, Y. and Byrd, R. J., 2001. Hybrid Text Mining for Finding Abbreviations and Their Definitions. *IMB Thomas J. Watson Research Center*, pp. 167-170.
- Paxton, J., 1983. *Everyman's Dictionary of Abbreviations*. London: J. M. Dent & Sons.
- Pustejovsky, J., et al. 2001. Automation Extraction of Acronym-Meaning Pairs from Medline Databases. In *Proceedings of Medinfo*, pp. 371-375.
- Righini, E., 2001. *Dizionario di Sigle Abbreviazioni e Simboli*. Bologna: Zanichelli.
- Robert, P., 1996. *Le Nouveau Petit Robert*. Paris: Dictionnaires le Robert.
- Schwartz, A. S. and Hearst. M. A., 2003. A simple algorithm for identifying abbreviation definitions in biomedical texts. *Proceedings of the Pacific Symposium on Biocomputing*, pp. 451-462.
- Sinclair, J., 1999. *Collins COBUILD English dictionary*. London: HarperCollins.

- SVEZ, 2000. *Evroterm, Večjezična terminološka zbirka*. [online] Available at: <http://evroterm.gov.si/index.php?jezik=angl> [Accessed 16 February 2011].
- Steinhauer, A., 2005. *Das Wörterbuch der Abkürzungen*. Mannheim: Duden verlag.
- Šlenc, S., 1997. *Veliki italijansko-slovenski slovar*. Ljubljana: DZS.
- Šlenc, S., 2006. *Veliki slovensko-italijanski slovar*. Ljubljana: DZS.
- Taghva, K. and Gilbreth, J., 1998. Recognizing acronyms and their definitions. *IJDAR I(IV)*, pp. 191-198.
- Zahariev, M., 2004. *A (Acronyms)*. Ph. D. Ottawa: School of Computing Science, Simon Fraser University.
- Zingarelli, N., 2000. *Vocabolario della lingua italiana*. Bologna: Zanichelli.
- Župančič, J., 1948. *Kratice, Mala izdaja*. Ljubljana: DZS.
- Yeates, S., 1999. Automatic extraction of acronyms from text. *Proceedings of the Third New Zealand Computer Science Research Students' Conference*, pp. 117-124.
- 24ur.com. [online] Available at: <http://24ur.com/> [Accessed 16 February 2011].

Mojca Kompara

Pot k izviru 18
6320 Portoroz
Slovenia

mokopt@yahoo.com
<http://bos.zrc-sazu.si/kratice.html>