Proceedings of the Language Technologies for All (LT4All) , pages 388–392
Paris, UNESCO Headquarters, 5-6 December, 2019.

The Case of Polish on its Way to Become a Well-Resourced-Language

Zygmunt Vetulani[1], Grażyna Vetulani[2]
Adam Mickiewicz University in Poznań[1,2]
ul. Wieniawskiego 1, 61-712 Poznań, Poland,
{vetulani, gravet}@amu.edu.pl

Abstract

We intend to illustrate – on our own research – the effort one needs to invest in order to obtain a real size application of deep text understanding for a language being categorized as *less-resourced.* We present some language resources and tools we had to develop for processing text in Polish, a language which was still in this category in the 1990s. The resources were used to implement a full scale prototype of the rule-based system POLINT-112-SMS for improving information workflow for emergency management purposes.

Keywords: less-resourced-language, Language Technology, language resources, dictionaries, lexicon grammars, tools, applications with NL competence.

1. Introduction

Polish, a West Slavic language spoken by about 45 million native speakers, has written literature since XIII century and in XXth century was among the best described languages. Still it was classified for a long time as a language with a very poor technological infrastructure. Using today's terminology, Polish was considered as a "less-resources-language" at least until the late 1990s. Negative effects of this scarcity time may still be observed. The EU initiatives of of awareness actions followed by appropriate funding measures appeared to be an important milestone in the development of human language technologies (HLT) in Central Europe. Successive incorporation to the international (and multilingual) research was effect of new bridges emerging between the technologically advanced research communities and newcomers often contributing with new original ideas. International grants bringing together partners from various language communities and representing different linguistic traditions were often oriented to creation of basic language resources and tools requested by quickly growing multilingual sector of language industries, also in Poland.

The general condition of language technologies for Polish is now good enough to no classify Polish as "less-resourced". Having said that, there are still many gaps to be filled concerning language resources and tools.

The conference-forum LT4ALL 2019 organized under the auspices of UNESCO that has just ended in Paris was today's equivalent of the EU actions taken a quarter of a century ago. It was addressed primarily to the representatives – researchers and administrative staff – of clearly "less-resourced-languages", but also to other LT community members, in order to sensitize all of them to new threats of a "world-at-various technological speeds".

In this paper we want to share our experience in the ambitious task to acquire the capacity to develop LT based technologies in the situation of initial scarcity of digital language resources and tools for Polish.

2. Beginnings

Intensive development of Polish literature since Renaissance and Baroque was followed by high quality linguistic descriptions of the language, so that in 20th century it was among the best described languages. Rapid progress of computer technologies attracted the attention of researchers on possibility of machine translation already in late 1940s. In Poland, among the first attempts to use computers in processing Polish text and speech worthy a note are works conducted in 1970s in Warsaw and Poznań (L. Bolc, S. Szpakowicz for text processing, W. Jassem, M. Steffen-Batóg for speech generation and analysis). In late 1970s and early 1980s the first attempts to implement toy systems understanding Polish were done (independently) by W. Lubaszewski, St. Szpakowicz, Z. Vetulani. The common feature of these first works was inaccessibility of real-size electronic language resources and NLP dedicated tools for Polish. Still, these initial works (see e.g. Z. Vetulani, 1984 and 1988) appeared stimulating a decade later when we started working on real size application prototypes of systems involving text understanding (Vetulani, 1991). These initial works highlighted necessity of real scale language resources.

3. Basic resources for text processing

On the ground of our trials to design non-trivial question-answering systems with deep understanding (rule based) (Vetulani, 1984, 1988) we realized that the most urgent necessity was to dispose of electronic dictionaries, computer processible grammars of Polish and corpora (both application specific, but also corpora for general language). Several projects contributed to fill (partially) the lack of the above mentioned resources and tools.

3.1 Dictionary project POLEX (1994-1996)

Polish is a language with a complex inflection system and has relatively free word order. Therefore simple adaptation of processing algorithms efficient for English or French appeared hard to apply, as basic information concerning the function of a word in the sentence is typically being encoded in the word form, independently of its linear position in the sentence.

Good grammatical description of Polish existed until recently only in form of traditional dictionaries and grammars addressed to the human users and were of low usefulness for automatic processing because of lack of precision. Huge amount of work invested by grammarians

until 1990s did not lead to a standard of description of Polish words that would eliminate the necessity of individual linguistic competence of users to interpret dictionaries. We proposed unambiguous inflectional description apt to eliminate linguistic competence of dictionary users and therefore be appropriate for machine text processors. The solution we proposed is an outcome of the POLEX Polish Lexicon Project (1994-1996)[1].

The POLEX Polish Lexicon is a morphological dictionary which includes the core Polish vocabulary of general interest of the traditional paper dictionary (Szymczak, 1981). It is based on a precise machine-interpretable formalism (coding system), the same for all categories (classes of speech) (Vetulani et al., 1998).

The dictionary entries are of the following form:

BASIC_FORM+LIST_OF_STEMS+PARADIGMATIC_CODE+DISTRIBUTION_OF_STEMS

The paradigmatic inflection code contains full paradigmatical information, i.e. the way how to associate endings to stems to obtain a required form of the word. The distribution associates stems to the paradigmatic positions.

The first public release of the resource contained over: 42,000 nouns, 12,000 verbs, 15,000 adjectives, 25,000 participles, and about 200 pronouns[2].

3.2 First steps towards Lexicon grammars for Polish

In the early 1970s Maurice Gross (LADL, Paris 7) the concept of a grammatical lexicon based on the idea of storing words together with possibly all relevant syntactic and semantic information (Gross, 1975). This idea, inspired by Z. S. Harris' transformation theory, was developed first for French, then for other languages. Consequently, predicative words were studied from the point of view of their aptitude to form elementary sentences. Gross introduced the term lexicon-grammar (fr. lexique-grammaire) to mean the method to describe the meaning of predicative words by providing description of how these words form simple sentences. What distinguishes lexicon grammars from traditional grammatical descriptions of a language is that lexicon grammar entries contain possibly full grammatical description[3] of well distinguished senses of words. This property makes that lexical-grammars are adequate for application in language processing systems.

The EUREKA project GENELEX (1990-1994)[4] was an initiative to define a generic model for lexicons; to design and develop software tools for lexicon management (Marie-Hélène ANTONI-LAY et al., 1994) based on the ideas of lexicon-grammar. Anoni-Lay gives two reasons to build large size lexicons. "The first reason is that Natural Language applications keep on moving from research environments to the real world of practical applications. Since real world applications invariably require larger linguistic coverage, the number of entries in electronic dictionaries inevitably increases. The second reason lies in the tendency to insert an increasing amount of linguistic information into a lexicon. (…) In the eighties, new attempts were made with an emphasis on grammars, but an engineering problem arose: how to manage a huge set of more or less interdependent rules. The recent tendency is to organise the rules independently, to call them syntactic/semantic properties, and to store this information in the lexicon. A great part of the grammatical knowledge is put in the lexicon (…). This leads to systems with fewer rules and more complex lexicons." (ibid.).

Two COPERNICUS projects CEGLEX – COPERNICUS 1032 (1995-1996) and GRAMLEX – COPERNICUS 621 (1995-1998) were executed under the EC funded 4th Framework Program which accepted participation of Central European Countries. One of the goals of these projects was to test the potential of the extension of novel LT solutions to languages that were going to be considered as official in future EU members after 2004.

3.2.1 PECO-COPERNICUS project CEGLEX (1995-1996)

The main goal of the CEGLEX consortium (Vetulani et al. 1994) was to test the GENELEX proposal of a generic model for re-usable lexicons – first implemented for a number of West-European languages, among other French, English, German, Italian – for three more languages spoken in Central Europe: Czech, Hungarian and Polish.

The CEGLEX/GENELEX model claims to be:

- theory-welcoming,
- complete, i.e. to cover all relevant phenomena on three classical layers: morphological, syntactical, and semantic,
- easily transportable.

The three layers of the CEGLEX/GENELEX model were confronted with the data of the considered languages with generally positive results, especially for Czech and Polish. For Polish this confrontation consisted in the adaptation of the model to the Polish data. On this occasion some modifications were proposed, in particular concerning the representation of the inflection phenomena. It is worth noticing that the POLISH CEGLEX module went further than GENELEX as we considered also semantic layer which was only marginally addressed in GENELEX. The outcome of CEGLEX was the first successful attempt to test on representative linguistic data feasibility of machine readable lexicon-grammar covering all three layers.

3.2.2 PECO-COPERNICUS project GLAMLEX (1995-1998)

The aim of the COPERNICUS Project 621 GRAMLEX was to facilitate the initiation, coordination and standardization of the construction of morphological dictionary packages for the following European languages: French, Hungarian, Italian and Polish, including detailed formal description of the morphology of the languages. The

[1] Resarch project „POLEX - Polska Leksykalna Baza Danych No KBN8S50301007" realised by Z. Vetulani, B. Walczak, T. Obrębski, G. Vetulani and other team members (1994-1996).

[2] The resource is distributed through ELRA. ISLRN: 147-211-031-223-4; ID: ELRA-L0047

[3] On both syntaxic and semantic levels.

[4] GENELEX was continued by LE-PAROLE (1996-1998), LE-SIMPLE (1998-2000) and GRAAL (1992-1996) projects.

intention of the GRAMLEX tasks as to Polish was to contribute to the improvement of the situation concerning language engineering tools and resources for Polish. Among the main achievements was a corpus-based SGML-encoded (in format GRAMCODE) morphological dictionary of Polish (over 22.500 entries) and related tools and applications (lemmatizer, inflected form generator, concordance generator and other5). The project GRAMLEX was closely connected with two projects mentioned above POLEX and CEGLEX.

3.3 Further steps towards Lexicon grammars for Polish. PolNet 3.0 as Lexicon-Grammar

The IT applications with language competence we were able to develop until 1990 were all of the category of toy-systems. This was, first of all, because of scarcity of real-size digital, easily machine processable electronic resources. This problem was addressed in the R&D grant "Text Processing Technologies for Homeland Security Purposes" that we realized during 2006-2010[6]. Within this grant we created a prototype of POLINT-112-SMS system designed to assist the monitoring process of mass events and to enhance real-time identification of processes in the crowd of fans in order to discover potentially dangerous situations with a high degeneration risk (early prevention).

The POLINT-112-SMS project required a robust natural language competence to understand and process SMS messages exchanged between the security staff agents in uncontrolled natural language (cf. Vetulani and Osiński, 2017). The understanding module of the system is rule-based because of necessity to obtain very precise representation of utterance content which is crucial in processing neuralgic information. Messages were supposed to be written in standard, correct and unconstrained Polish.

3.3.1 PolNet - Polish Wordnet as lexical ontology (since 2006)

Within the project POLINT-112-SMS (Vetulani et al., 2010) we made extensive use of ontology to represent meaning of text messages. Absence on the market of lexical ontologies reflecting conceptualization typical of Polish speakers inspired us to develop PolNet Polish Wordnet – a lexical database of the type of Princeton WordNet[7]. We built it from scratch for Polish following the so called "merge model" methodology[8] PolNet design started in 2006 and its progress continues. The resource development procedure was based on the exploration of good traditional dictionaries of Polish and the use of available language corpora (e.g. IPI PAN Corpus; cf. Przepiórkowski, 2004). Development of PolNet was organized in an incremental way, starting with general and frequently used vocabulary[9].

We decided to selected the most widely used words found in a reference corpus of Polish language (ibid.) with however one important exception made for methodological reasons. Even though we wanted the core of PolNet to be a resource of general interest, we also assumed its possibly early validation in the real-size applications.

By 2008, the initial PolNet version based on noun synsets related by hyponymy/hyperonymy relations was already reach enough to serve as core lexical ontology for real-size applications. However, to develop a POLINT-112-SMS system prototype, an extension of the core set of nouns with domain terminology was necessary. Further extension with verbs and collocations transformed PolNet in a lexicon-grammar facilitating implementation of parsers.

3.3.2 From PolNet to Lexicon-Grammar for Polish. PolNet 1.0

Already in early 1980s information typically contained in lexicon-grammar entries for predicative words, simple or compound, was considered useful for parsing and generating natural language sentences. Lexical entries used in the PROLOG code of the demonstration system ORBIS[10] (Colmerauer and Kittredge, 1982), were in fact lexicon-grammar units describing syntactic and sematic valency of words. The syntactic/semantic valency was used as constraints permitting avoiding producing incorrect sentences by a generation algorithm, to avoid accepting incorrect sentences by a parser, and to build error-correcting software (ibid, see also Vetulani, 1988). Qualitative evolution of PolNet, initially conceived as lexical ontology, towards a lexicon-grammar of Polish took place at the passage from the release of PolNet 0.1 (2009) to the version PolNet 3.0. (2014). The pragmatic reason to substantially enrich PolNet was the need of an efficient parsing engine to support the understanding module.

In addition to noun synsets that make of PolNet 0.1 (2009) a lexical ontology, we decided to enrich PolNet with verb synsets containing syntactic/semantic information in form of valency structure. The valency structure of a predicative word provides the morpho-syntactic and semantic constraints on the acceptable fillers of the argument positions opened by this word (like *case, number, gender, preposition, register* etc. for morpho-syntactic constraints and semantic roles like *agent, patient, beneficient,* etc. for semantic ones). In (Vetulani, Z. & Vetulani, G., 2014b) we

[5] These tools and applications were:
1) a lemmatizer/tagger (LEXAN) (Vetulani et al., 1997;1998),
2) a generator of inflected forms for simple and compound lexemes (Vetulani et al. 1998),
3) a syntactic concordance generator (SCON) (Vetulani et al. 1998),
4) a tool for extraction of compound terms and terminology from texts (EXTRACT) (Vetulani et al. 1998),
5) an application for structure analysis of dictionary entries (VERBAN) (Vetulani et al. 1998),
6) an application for acquisition of the lexicon from dictionary definitions (NOUNAN) (Vetulani et al. 1998),
7) an application for interactive analysis of dictionary definitions (NOUNDAN) (Vetulani et al. 1998).

[6] Grant of Polish Ministry of Science and Higher Education (MNiSzW) Nr R0002802.

[7] In the Princeton WordNet (and similar systems) the basic entities are synsets, i.e. classes of synonyms related by some relations of which the most important are hyponymy and hyperonymy.

[8] Princeton WordNet (Miller et al., 1990) was used as a formal ontology to implement systems with language understanding functionality. In order to respect specific Polish conceptualization of world, we decided to build PolNet from scratch rather than merely translate Princeton WordNet into Polish.

[9] See (Vetulani et al., 2007) for PolNet development algorithm.

[10] ORBIS, an interface to a database on planets, was entierly implemented in PROLOG to show the qualities of declarative programming paradigm.

presented the idea of a verb synset as follows: "In opposition to nouns, where the focus is on the relations between concepts (represented by synsets), and in particular on hiperonymy/hyponymy relations, for verbs the main interest is in relating verbal synsets (representing predicative concepts) to noun synsets (representing general concepts) in order to show what connectivity constraints corresponding to the particular argument positions are. This approach imposes granularity restrictions on verbal synsets and more exactly on the synonymy relation. Synonymous will be only such verb+meaning pairs in which the same *semantic roles* take the same concepts as value (this is necessary but not sufficient). In particular, the valency structure of a verb is one of formal indices of meaning (members of a sysnset share the valency structure)." (Vetulani, Z., Vetulani, G. 2015).

Verb synsets appeared already in the first public release of PolNet in 2011 (PolNet 1.0) (Vetulani et al. 2016). This opened a new generation of PolNet systems that we call now "PolNet – Polish Lexicon-Grammar systems". The expansion of PolNet to Lexicon Grammar of Polish was based on the results of theoretical research on predicative verbs assembled in Dictionary of Polish Verbs (Polański, 1980-1992) where linguistic descriptions were provided for 7000 Polish predicative verbs.

The valency information permitted us to make a smart use of PolNet enriched with lexicon-grammar features when implementing the POLINT-112-SMS system. In addition to using PolNet as lexical ontology in the World Knowledge and Situation Analysis Modules, we made use of the valency information to enhance efficiency of the parser being a part of the Text Understanding Module. In this module, syntactic/semantic valency information stored in lexicon-grammar rules was used to control parsing execution by heuristics[11] in order to speed-up parsing due to additional information gathered at the pre-analysis stage. The effect of substantially reducing the processing time was due to the reduction of search space.

3.3.3 Collocations in PolNet 2.0 - PolNet 3.0.

Usefulness of lexicon-grammar approach confirmed through successful implementation of the POLINT-112-SMS system's prototype motivated us to successive extension of PolNet towards a full lexicon-grammar. The versions PolNet 2.0 and PolNet 3.0 are important milestones in this process.

The passage from PolNet 1.0 to PolNet 2.0 was marked by inclusion of an important set of verb-noun collocations from the "Syntactic dictionary of verb-noun collocations in Polish" (Vetulani, G. 2000 and 2012)[12] or directly from corpora (Vetulani et al. 2010)[13]. Adding verb-noun collocations to PolNet appeared a non-trivial task because of specific morpho-syntactic phenomena related to collocations as for example syntactic synonymy[14] (Vetulani et al. 2016), as well as the problem of (optimal) granularity of verbal synsets. In (Vetulani, Z., Vetulani, G.,2014b) we noticed: "The challenging issue of verb synsets granularity is closely connected with synonymy which is fundamental for the concept of wordnet. Let us notice the fact (…) that while there is consensus through the wordnet community concerning the principle that *synonymy is the basis of organization of the (wordnet) database in synsets*, (i.e. synonyms should belong to the same synsets), there is no consensus among linguists on the concept of synonymy. Miller and Fellbaum (in Vossen et al., 1998) postulate a very weak understanding of this concepts (based on *invariability test with respect to just one linguistic context*) often leading to very large synsets."

The version 3.0 of PolNet was meticulously cleaned and extended[15] with respect to the version 2.0. It has been user-tested as a resource for modeling semantic similarity between words (Kubis, 2015).

4. Conclusions

This article is a case study to illustrate the challenges on way to achieve ambitious technological goals and to give an idea of the effort to be invested in the situation of initial scarcity of language resources and tools – typical of less-resourced-languages. We presented our long term works resulting with implementation of a sophisticated ICT system and development of significant language technology workbench. To achieve success it was necessary to first collect or produce basic instruments. Therefore the urgent need was to produce ready-to-use resources like processable text and speech corpora, electronic dictionaries, computer-readable grammars.

By no means we pretend to claim that the solutions we present here are the only good and sufficient measures to reach the goal which is to obtain for Polish the status of a well-resourced-language. In fact, the totality of good effects obtained on this way were due to the individual or collective effort of the LT community in Poland.[16]

[11] A well-constructed heuristic permits – on the basis of morphological and valency information combined with the switch technique Vetulani (1994) – to to reduce the complexity of parsing down to linear in an important number of cases.

[12] This dictionary is described in two monographs. The first one (2000) describes the initial phase of works on a dictionary of verb-noun collocations together their usage in sentences as predicates (2862 predicative nouns). This work was done manually. Extension of resource to 14.600 collocations was described in the second book (2012) reporting further, computer-assisted work. A part of this resource was integrated with PolNet.

[13] In (Vetulani et al. 2010) we described a computer-assisted algorithm to extract collocations directly from text corpora. The algorithm requires involvement of qualified lexicographers.

[14] In Polish we observe the phenomenon of syntactic synonymy (Jędrzejko 1993) where for some predicative verbs their morpho-syntactic structure is different from the morpho-syntactic structure of their semantic synonyms in form of verb-noun collocation (e.g. for the direct complement). Therefore to be consistent with our methodological assumption, we will range these synonymous forms in distinct synsets interconnected by a special semantic similarity relation.

[15] From14.400 in PolNet 2.0 to 17.564 in PolNet 3.0.

[16] We applied some outcomes of this effort., as e.g. the IPI PAN Corpus of Polish teksts (Przepiórkowski 2004), further extended to a National Corpus of Polish Language. Other crucial resources as wordnets were still non-existing or not available at the time we needed them. For example another wordnet for Polish – Słowosieć (also known as plWordNet; see Piasecki et al. 2009) was developed independently of PolNet at about the same time and according different methodological bases.

A particular attention should be attracted to make the developed resources and tools sustainable, reusable, open for further development, and – last but not least – easy to maintain. All these require an additional effort.

Three kinds of circumstances may considerably speed-up the transition of the status of the language from "less-resourced" to "well-resourced":

- for language: to have solid traditional linguistic description,
- man-power: dispose of well-formed staff and students: linguists and computer engineers,
- technology: being eligible for receiving international assistance, participate in international development programs and projects, possibly as full partners.

5. References

Colmerauer A. and Kittredge R. (1982): ORBIS. In J. Horecký (Ed.) *Proceedings of the 9th COLING Conference*.

Gross, M. (1975): *Méthodes en syntaxe,* Paris: Hermann.

Jędrzejko, E. (1993): *Nominalizacje w systemie i w tekstach współczesnej polszczyzny,* UŚ Katowice.

Kubis, M. (2015): A semantic similarity measurement tool for WordNet-like databases. In Z. Vetulani, J. Mariani (Eds.), *Proceedings of the 7th Language and Technology Conference*, Poznań, Poland, 27-29 November 2015. FUAM, Poznań, 150 – 154.

Miller, G. A, Beckwith, R., Fellbaum, C.D., Gross, D., Miller, K. (1990): WordNet: An online lexical database. Int. J. Lexicograph. 3, 4, 235–244.

Piasecki M., Szpakowicz S., Broda B. (2009) *A Wordnet from the Ground Up*, Oficyna Wydawnicza Politechniki Wrocławskiej,Wrocław.

Polański, K. (Ed.) (1980-1992): *Słownik syntaktyczno-generatywny czasowników polskich,* vol. I-IV, Ossolineum, Wrocław, 1980-1990, vol. V, Instytut Języka Polskiego PAN, Kraków, 1992.

Przepiórkowski, A. (2004): *Korpus IPI PAN. Wersja wstępna* (The IPI PAN Corpus: Preliminary version). IPI PAN, Warszawa.

Vetulani, G. (2000): *Rzeczowniki predykatywne języka polskiego. W kierunku syntaktycznego słownika rzeczowników predykatywnych na tle porównawczym*, Adam Mickiewicz University Press: Poznań.

Vetulani, G. (2012): *Kolokacje werbo-nominalne jako samodzielne jednostki języka. Syntaktyczny słownik kolokacji werbo-nominalnych języka polskiego na potrzeby zastosowań informatycznych. Część I,* Adam Mickiewicz University Press: Poznań.

Vetulani, G. (2017): Próby formalizacji zdań opartych na predykatach rzeczownikowych języka polskiego, *Linguistica Copernicana*, 14/2017, Wydawnictwo Naukowe UMK, Toruń, 127-143.

Vetulani, Z. (1988): PROLOG Implementation of an Access in Polish to a Data Base. In *Studia z automatyki*, XII, PWN, 5-23.

Vetulani, Z. (1991): Lexical preanalysis in a DCG parser of POLISH. In E. Klein et al. (Eds.), *Betriebslinguistik und Linguistikbetrieb. Akten des 24 Linguistischen Kolloquiums, Bremen 1989,* (Linguistisches Arbeiten 260/261), Max Niemeyer Verlag, Tübingen, 389 - 395.

Vetulani, Z. (1994): SWITCHes for making Prolog more Dynamic Programming Language, Logic Programming, *The Newsletter of the Association for Logic Programming*, vol 7/1, February 1994, page10.

Vetulani, Z., Martinek, J., Vetulani, G. (1995): The CEGLEX dictionary model for Polish. In R. Bazylewicz, O. Kossak (Eds.), *Proc. of the 4th and 5th International Conferences UKRSOFT (Lviv, 1994, 1995),* SP «BaK», Lviv, 1995, 144 - 150.

Vetulani, Z., Martinek, J., Obrębski, T., Vetulani, G. (1997): Lexical Resources and Tools for Tagging Polish Texts within GRAMLEX. In *Investigationes Linguisticae, XXI:2, 1997,* 401-416.

Vetulani, Z., Walkowska, J., Obrębski, T., Konieczka, P., Rzepecki P., Marciniak, J. (2007): PolNet - Polish WordNet project algorithm. In Z. Vetulani (ed.) *Proc. of the 3rd Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, October 5-7, 2007, Poznań, Poland*, Wyd. Poznańskie, Poznań, 172-176.

Vetulani, Z., Marcinak, Obrębski, T., Vetulani, G., Dąbrowski, A., Kubis, M., Osiński, J., Walkowska, J., Kubacki, P., Witalewski, K. (2010): *Zasoby językowe i technologie przetwarzania tekstu. POLINT-112-SMS jako przykład aplikacji z zakresu bezpieczeństwa publicznego (Language resources and text processing technologies. POLINT-112-SMS as example of homeland security oriented application),* Adam Mickiewicz University Press: Poznań.

Vetulani, Z. (2012): Language Resources in a Public Security Application with Text Understanding Competence. A Case Study: POLINT-112-SMS, *Proc. of the LRPS Workshop at LREC 2012, May 27, 2012. Istanbul, Turkey*, ELRA: Paris, 54-63.

Vetulani, Z. (2014): PolNet - Polish WordNet, in: Z. Vetulani, and J. Mariani (Eds.): *Human Language Technology. Challenges for Computer Science and Linguistics. LTC 2011. Revised Selected Papers.* LNAI 8387, Springer-Verlag Berlin Heidelberg, 408-416.

Vetulani, Z., Vetulani, G. (2014a): Through Wordnet to Lexicon Grammar, in. F. Kakoyianni Doa (Ed.). *Penser le lexique grammaire: perspectives actuelles*, Editions Honoré Champion, Paris, 531-543

Vetulani, Z., Vetulani G. (2014b): Verb-Noun Collocations in PolNet 2.0. In V. Henrich and E. Hinrichs (Eds.): *Proceedings of the Workshop on Computational Cognitive, and Linguistic Approaches to the Analysis of Complex Words and Collocations (CCLCC 2014)* Tübingen, Germany, 73-77.

Vetulani, Z., Vetulani, G. (2015): Synonymie et granularité dans les bases lexicales du type Wordnet, *Studia Romanica Posnaniensia*, Vol. 42/1, Wydawnicywo Naukowe UAM, Poznań, 113-127.

Vetulani, Z., Vetulani, G., Kochanowski, B. (2016): Recent Advances in Development of a Lexicon-Grammar of Polish: PolNet 3.0. In N. Calzolari et al. (Eds.), *The 10th Conference on Language Resources and Evaluation*, ELRA, Paris, France, 2851-2854.

Vetulani, Z., Osiński, J. (2017): Intelligent Information Bypass for More Efficient Emergency Management", *Computational Methods in Science and Technology* 23(2), 105–123.

Vossen,P., Bloksma, L., Rodriguez, H., Climent, S., Calzolari, N., Peters, W. (1998): *The EuroWordNet Base Concepts and Top Ontology, Version 2, Final, January 22, 1998 (Euro WordNet project report).*