

MAREK ŚWIDZIŃSKI, MICHAŁ RUDOLF

**Narzędzia informatyczne obsługi wielkich korpusów tekstów:
wyszukiwarka Holmes**

Summary

Large text corpora management requires sophisticated computational tools. For highly inflecting languages like Polish homonymy is a challenge computer men have to face; in Polish texts, every 42nd word per 100 is grammatically ambiguous. A search engine **Holmes**, designed by Michał Rudolf, works as a disambiguator, rather than a tagger. It operates on texts which are morphologically marked before by special programs. After the user keyboards her query **Holmes** examines sets of tags for each word, rejecting as many improper interpretations as possible. **Holmes** makes use of linguistic, not statistical methods of disambiguation. It is based upon a number of rules formalizing various contextual restrictions on words. Query results are obtainable online.

JANUSZ S. BIEŃ

Aparat pojęciowy wybranych systemów przetwarzania tekstów polskich

Summary

Several computer systems are discussed. PoMor is a commercial product (morphological analyser and spelling checker) developed by Robert Wołosz for Morphologic. Morfeusz is a morphologic analyser developed by Marcin Woliński. Poliqarp (POLyinterpretation Indexing Query and Retrieval Processor) developed by Zygmunt Krynicki and Daniel Janus is used to search in a corpus of Polish (<http://korpus.pl>). Both Morfeusz and Poliqarp are freely available for research purposes. The linguistic tools of Michał Rudolf are also mentioned.

The paper advocates the use of *lex* instead of *word* or *segment* for better precision and clarity.

MAŁGORZATA MARCINIAK, AGNIESZKA MYKOWIECKA

**Powierzchniowe przetwarzanie składniowe na potrzeby
automatycznej ekstrakcji informacji z raportów medycznych.**

Summary

The paper presents a system for automatic content extraction from mammogram reports written in Polish. The system combines general information extraction (IE) techniques with external post-processing aimed at structuralizing the results. The paper contains a characteristics of the specific type of texts as well as a description of the results obtained together with a short analysis of advantages and disadvantages of shallow text processing.

MAREK ŁAZIŃSKI, MONIKA SZEWCZYK

Słowa klucze w semantyce i statystyce. Słowa tygodnia „Rzeczpospolitej”

Summary

The article discusses the notion of *key words* in linguistics, from cultural semantics to the strictly statistical approach of determining the words that identify texts of particular authors or periods. The author then presents *words of the week* of the Polish daily newspaper *Rzeczpospolita* – lists of proper names and common words whose relative frequency of occurrence in particular weeks is distinctly higher compared to the last quarter. The key words of consecutive weeks were categorised according to their durability over longer periods of time: in a quarter as compared to the last three years and in the last three years as compared to the vocabulary of the Polish press of the 1990's to date from the PWN Corpus of Polish. It has been found that key words in the language of the press are connected with current affairs and none of them are words considered to be of key importance for Polish culture in general by researchers in cultural semantics.

MAGDALENA KURATCZYK

Narzędzia korpusowe w leksykografii dwujęzycznej Summary

The use of monolingual electronic corpora in bilingual Polish and Russian lexicography is discussed. Starting from her own experience in dictionary building the author provides evidence that electronic corpora are most useful for the description of new words (even those not yet included in monolingual dictionaries). Electronic corpora have also proved to be an effective method for selecting equivalents in existing bilingual dictionaries. A typology of available electronic corpora is proposed.

MAGDALENA TURSKA, NATALIA KOTSYBA

Polsko – ukraiński korpus równoległy (PolUKR) Summary

A fairly considerable lack of Polish-Ukrainian electronic language resources, especially large dictionaries and parallel corpora, is presently felt among the professional translators and those who learn either language. A morpho-syntactically annotated aligned corpus with a parallel concordancer can be used in many cases as a traditional dictionary. One of its possible advantages is being an up-to-date database of numerous examples of language use, which could be of use while dealing with neologisms and new lexical meanings.

Our aim is to create such a corpus and make it publicly available through the Internet. At the moment one can advise the prototype of the parallel Polish-Ukrainian corpus on-line from the website <http://corpus.domeczek.pl>. It contains ca. 50 parallel bitexts compliant with the XCES standard. Not all sentence segments have been tagged yet, the work on morphological annotation is on-going as well. The corpus is aligned presently only at the paragraph level.

We are convinced that such a parallel corpus with an efficient search engine, a flexible query language and a friendly user interface will be one of the most powerful tools for all Polish and Ukrainian researchers, translators and students.

IWONA LOEWE

Internet i jego zasoby w polskich badaniach lingwistycznych. Rekonesans Summary

The Internet's mediality, geography, and communicative specificity make it a socially interesting domain of study. Linguistics as a humanistic discipline may contribute significantly to the research on the new media and their impact on contemporary society and it has the necessary tools for the task.

The article discusses the following areas of linguistic research on the Internet: secondary orality or oralisation of writing, understood as "transformation of verbal expressions by electronic means" (a secondarily oral written text may be sent and received simultaneously in real time); tendency to abbreviation, evident in acronyms; emoticons as a fashionable means of conveying information about the attitude of senders of messages; netiquette, i.e. the emergence of new conventions of politeness observed by chat users; the role of the Internet in language counselling and guidance; creating a new type of dictionary (e-dictionary) as a task for lexicographers and metalexigraphers; hypertext as a representation of a net text *sensu stricto*; genres on the internet.

ANNA TERESZKIEWICZ

Analiza gatunkowa encyklopedii internetowej Wikipedia

Summary

The paper presents the results of a genre analysis of an online encyclopedia – Wikipedia, whose form is completely shaped by the web. The aim of the analysis was to determine characteristic features of the encyclopedia and point out the ways in which it differs from other available online encyclopedias. Analysis covered structural and stylistic features of the articles. Structural analysis defined the characteristics of the visual side of the encyclopedia, graphics, constructional elements of the site as well as systemic mechanisms. Stylistic analysis verified the language of the encyclopedia, language of articles and means of communication with the users. Linguistic analysis verified the degrees of formality, lexis as well as the most frequent syntactic structures. The results show how users acting through mechanisms provided by the system can shape the features of the content in particular ways. The paper also includes a comparative analysis of the Polish version of Wikipedia with its English and German equivalents.

MARTA DĄBROWSKA

(Nie)grzeczność w mediach elektronicznych

Summary

The paper describes deviations from the conventional norms of politeness in the process of communication via e-mail between students (as subordinates) and lecturers (as their superiors) in the English Department. It has been noted that the students are to a large extent unable to use an appropriate register in their e-mail correspondence, focusing more on the aspects of solidarity between interlocutors rather than power distance between them. These modifications manifest themselves through a variety of elements, e.g. a disregard for the norms of punctuation and spelling, the use of forms of address typical of the spoken interaction with peers, code-mixing between Polish and English, or the choice of vocabulary and style unsuitable for the type of relationship. All these violate the traditional linguistic etiquette. This is additionally augmented by the inappropriate phrasing of requests which formally often employ the markers characteristic for orders. These changes are attributed to

the student's predominant exposure to electronic media communication in informal contexts which consequently acts as the prototype for the written communication process in the modern era.

ALEKSANDER KIKLEWICZ

Forma Summary

The author discusses the problem of the form of linguistic signs in different linguistic-philosophical paradigms: rationalistic (formalistic), postrationalistic, structural and poststructural. According to the author the dynamics of the linguistic-philosophical paradigms has depended on two factors: the divisibility or indivisibility of the form and the content, and also the domination of the form or the content in the structure of linguistic signs as objects of linguistic description. The author distinguishes three kinds of linguistic forms: explicative (monomorphous), operative, and implicative, and discusses them in detail with reference to the stylistic function of the sentence in Polish and Russian. Special attention is given to zero forms, which are treated as a kind of synsemantic (index) signs, equivalent to anaphoric and cataphoric pronouns. The author considers also the problem of nonverbal forms, i.e. those which result from the cooperation of the utterance and the text with the relevant situational context. He introduces the notion of the principle of optimality, which defines the cooperation of the different factors of linguistic communication.

ZUZANNA TOPOLIŃSKA

Two levels of the grammatical (= morpho-syntactic) contact-induced change Summary

The author argues that contact-induced morpho-syntactic changes should be discussed and interpreted separately at the functional and at the formal level. She argues that pragmatic and/or semantic changes in every multilingual environment are triggered by the same factors: the need for successful communication between interlocutors speaking different languages and possessing insufficient knowledge of the language of the 'other'. These changes are defined by the universal hierarchy of information needed for successful communication. Referential information enabling identification of the spoken of events and their protagonists comes to the fore. On the other hand, formal changes depend directly of the inherited grammatical structure of the languages in question.

In order to prove her thesis the author analyses the so-called Balkanisms in Slavic-Balkan languages, i.e. contact-induced changes deriving from the contact with the non-Slavic members of the Balkan Sprachbund.

PIOTR STALMASZCZYK

Przyimki zaimkowe i konstrukcje posesywne w nowożytnych językach celtyckich Summary

One of the characteristic features of Celtic languages is the absence of a singular verbal form with the meaning 'to have'. The principal way of expressing possession is

through phrases with prepositions (especially Irish *ag*, Scottish Gaelic *ag* ‘at’, Welsh *gan*, Breton *gant* ‘at; with’). Pronominal prepositions (also known as conjugated prepositions or prepositional pronouns), another distinctive feature of Celtic languages, consist of a preposition and a suffixed pronoun, or rather a pronominal personal ending. Thus the Irish and Welsh equivalents of the English sentence ‘I have money’ are *Tá airgead agam* or *Mae arian gen i*, respectively, both literally meaning ‘is money at/with me’. The paper discusses the properties of pronominal prepositions in modern Celtic languages and comments on some historical developments leading to this formation.

WITOLD MAŃCZAK

Nieregularny rozwój fonetyczny spowodowany frekwencją
Irregular frequency motivated phonetic development

CYNTHIA M. VAKARELIYSKA

**Zróźnicowanie tożsamości językowej i kulturowej
niemieckojęzycznej mniejszości ewangelicko-augsburskiej
zamieszkującej tereny byłego zaboru rosyjskiego w Polsce w XIX i XX wieku**
Summary

The article is a sociolinguistic study of the long-standing German-speaking Lutheran minorities in two adjacent provinces of the 19th-century Russian Empire (Mazowsze, in present-day Poland, and Suwalszczyzna, in present-day southwestern Lithuania). Based on archival and field research, the study examines how members of this ethnoconfessional minority and their descendants perceived their own language, cultural, and ethnic identity over the 200+ years that they lived, as Russian citizens, in a community with multiple dominant languages (Polish, Russian, and sometimes Lithuanian).

The study focuses on the Lutheran settlers and their descendants, who tended not to lose the language as quickly as their Roman Catholic counterparts, as the Lutheran liturgy, Sunday School lessons, and Confirmation training all required an ability to read and speak German. There are indications, however, that the cultural and language self-identification of the Lutheran Russian Germans was more complex than the historical scenario suggests. The most striking factor is their multilingualism: by the mid-19th century, most Russian Germans were native-level speakers of Polish, and many were trilingual in Lithuanian. Many had at least an elementary knowledge of Yiddish, and after 1868, when Russian replaced Polish as the state language, those who went to school learned to read and write in Russian. Comparing data from both before and after 1868 on language choice, language interference between Russian and Polish, and individual naming and signature practices, the article draws the conclusion that most German-speaking Lutherans in Mazowsze and Suwalszczyzna identified themselves linguistically and culturally not only with German, but also, to a significant extent, with Polish and, to a lesser extent, with Russian.

DANUTA STANULEWICZ

**Zróźnicowanie regionalne prototypowych odniesień sześciu podstawowych
nazw barw w języku polskim (*biały, czarny, czerwony, zielony, żółty, niebieski*)
– na tle porównawczym wybranych języków słowiańskich i germańskich**
Summary

The aim of this paper is to present the results of a questionnaire the purpose of which was to investigate whether the primary basic colour terms in Polish (i.e. the equivalents of English *white, black, red, green, yellow* and *blue*) have the same reference-points in four regions of Poland. It appears that the reference-points are different for *niebieski* 'blue': in Central and Southern Poland it is associated mainly with the sky, whereas the inhabitants of Northern Poland associate it with the sky and water, just like speakers of English and Swedish. The vicinity of the Baltic Sea is undoubtedly the factor which influences this treatment of *niebieski* by the inhabitants of Northern Poland.

REVIEWS

ALEKSANDER KIKLEWICZ

Eugenio Coseriu, *Der Physei-Thesei-Streit. Sechs Beiträge zur Geschichte der Sprachphilosophie*. Hrsg. von R. Meisterfeld. Tübingen: Gunter Narr Verlag, 2004. 212 p.

WITOLD MAŃCZAK

Zagadnienie praojczyzny Słowian w encyklopediach PWN

[The question of the Slav's common original homeland in PWN encyclopedias]

BIBLIOGRAPHY

JANINA OZGA

Bibliography of Polish works on general linguistics, Indo-European and non-Indo-European languages published in 2004, with addenda