URGENT PROBLEMS OF THE MACHINE-OPERATED SEMANTIC ANALYSIS

Prof Dr Iryna Sekret
Zirve University, Gaziantep, Turkey
iryna.sekret@zirve.edu.tr

Introduction

Nowadays the world market of the commercial text mining / text analytics software is among the most rapidly developing ones. By a brief net search one can come across more than hundreds of the software propositions designed to conduct the automatic analysis of the textual data.
But in spite of the significant progress in this area and a huge technological leap in designing semantic analysis software, when it comes to the practice users continue to face a number of technical and contextual problems that decrease the quality of the anticipated results.

Contextual Problems

Among contextual problems it is possible to distinguish those, which derive from the **discrepancies** between the programmed **content** of the software and the **natural language** discourse.
One can face numerous cases of the word's meaning misinterpretations, which are caused by the following factors:

1. The word's polysemantic development;
2. Homonymic lexical items;
3. Metaphoric meanings of lexical units;
4. Limited programmed repertoire of the word's co-occurrences;
5. Lack of profound studies and analysis of the lexical unit's compatibility;
6. Semantic shifts as the result of a) the word's borrowing by another language, b) the word's entering another sublanguage; c) jargon or dialectal usage, etc.
7. Homonymy of terms and terminological units regarding to the subject area of their usage;
8. Rapid development of languages as the result of a) the technology development (evolution of new notions and new terms to denote them or reconceptualization of the existing lexical units), b) globalization and intercultural mergers, c) political, economic and cultural penetrations, etc.

The problems enumerated above concern mostly the level of a word and a lexical unit. But the list can be prolonged by such linguistic facts as

9. A syntactic ambiguity which also relates to the semantic analysis of the lexical units but at the level of a phrase and a sentence;
10. Discourse analysis entailing implicatures (additionally conveyed meanings), and based on specific schemas and scripts.

All these problems are especially obvious when it comes to the necessity to translate the text from one language into another.

References

1. *Conclusions on Efficient and Innovative Education and Training to Invest in Skills1 - supporting the 2014 European Semester, 2014. EDUCATIO , YOUTH, CULTURE and SPORT Council meeting Brussels, 24 February 2014. [online] Available at http://www.consilium.europa.eu/homepage/highlights/council-discusses-education-investing-in-skills?lang=en*
2. *Contemporary Linguistics, Ed. by W. O'Grady & J. Archibald, 6th Ed. 688 P., 2010.*
3. *European Commission. Europe 2020: A strategy for smart, sustainable and inclusive growth, 2010. COM (2010) 2020.*
4. *Digital Agenda for Europe. A Europe 2020 Initiative. [online] Available at http://ec.europa.eu/digital-agenda/en/our-goals/pillar-vi-enhancing-digital-literacy-skills-and-inclusion*
5. *Sekret I. Developing Students' Foreign Language Professional Competence at the Higher Technical Educational Establishments by Means of Distance Education: Monograph. Dniprodzerzhinsk, 2011, 384 p.*
6. *Sekret I. Enhancing Students' Foreign Language Professional Competence by Means of E-learning. In Global e-learning. Universidad a Distancia de Madrid. pp. 417- 428. (Spain).*
7. *Sekret, I., Feiner, F. The Vision of the Inclusive Society – Is it Possible to Realise? In: FEINER, Franz, PICKL, Gonda, SCHWETZ, Herbert, STRAßEGGEREINFALT, Renate, STRAßEGGER, Ulrich, SWOBODA, Birgit (Hg.): Jugend in ihrer Vielfalt - ihre Werte, ihre Perspektiven. Ergebnisse einer Studie der KPH Graz zu Erlebensund Befindenswelten Jugendlicher, 177 – 183, Graz: Leykam, 2013. ISBN 3-7011-0267-9*
8. *Sekret I., Kommers P. Conceptual issues of the digital competence development in the framework of the council of the European Union pp, 183-194 DIVAI 2014, The 10th International Scientific Conference on Distance Learning in Applied Informatics. 183 ISBN: 978-80-7478-497-2*
9. *Shepel Yu., O., Sekret I.V. Intersectoral Terminological Homonymy in Modern English. Monograph. Dnieprodzerzhinsky State Technical University, 2010. - 335 P.*
10. *Yule G. The study of language. 3d edition. Cambridge University Press, 2006.*

Technical Problems

Another group of problems are related to the fact that semantic analysis software remains to seem very much technically complicated for an ordinary user to apply it at least to some extent. Huge armies of interpreters, translators, language teachers, linguists, etc. are unaware of the full potential of the semantic analysis software or remain to be reluctant to implement it for their professional purposes because of lack of necessary skills and competences.
As the result, only small circles of specialists whose occupational needs are directly related to the discourse analysis come to implementing the textual data analysis software in their routine. So, even thoroughly elaborated software stays away from a wide public interest not because of the drawbacks of its promotion but because of the lack of the users' competences to consume it.
Due to this fact in order to overcome a gap between the semantic analysis software and its consumers, efforts should be made to design such programs which would be maximally directed to the needs of the potential user, simplified in its operations, user-friendly in its design, and accompanied with appropriate trainings.
Therefore, the decision of this group of problems rests in the area of the analysis of the users' professional needs, psychology, beliefs, positive / negative experience of implementing such kind of software, and introducing accompanying trainings to enhance required competences within the targeted consumers.

Problems of the Applied Character

Within this paper the problems of the applied character refer to the area of the language e – learning, namely vocabulary teaching, and developing skills in speaking and writing by the language learning software.
In spite of the profound achievements in the area of the human cognition, namely, functions of memory (memorization – recalling – forgetting), mental processes, image development, mechanisms of turning knowledge into practical skills, etc, when it comes to the programmed language teaching we can face very limited strategies and scenarios implemented to enrich a learner's vocabulary, which are mostly based on pure memorization, and provide very poor chances to convert obtained knowledge into speaking / writing skills.

Inclusiveness of the Information Society – Designing Appropriate Semantic Analysis Software

Another group of challenges posed by the modern society is the idea to make information and appropriate tools accessible to all social groups. The idea of the inclusiveness of the society is being spread to all spheres of the human life, starting from access to information, learning and including means facilitating those processes. It especially concerns language minority groups, and people with disabilities, having special needs in their life, learning and occupation.
In this respect the design of the semantic analysis software based on those principles and entailing analysis of the sign language, symbolic language, and other alternative means of the human communication could be urgent and timely.

Contact Information

Prof Dr Iryna Sekret

English Language Teaching Department,
Faculty of Education,
Zirve University
Kızılhisar Kampüsü 27260 GAZİANTEP

Tel: 539-622-5536
Email: irenesekret@mail.ru
iryna.sekret@zirve.edu.tr

Aims and Objectives of the Project

A brief overview of urgent problems of the machine-operated semantic analysis allows to suggest a project proposal which would be designed to achieve the following aims and objectives:

I. To enhance the content of the semantic analysis software (English, Russian, Ukrainian, Turkish) in the following steps:

a) To study the content of the semantic analysis software with a purpose to highlight limitations and to draw out ways of their elimination;
b) To analyze models of the semantic analysis with a purpose to discriminate pros and cons of the existing approaches;
c) To conduct thorough analysis of the linguistic sources and language factual material with the focus on the problems under concern;
d) To conduct contrastive studies of the focused language facts from the point of view of social, linguistic, and psychological perspectives, synchronically and diachronically;
e) To present findings as a bank of factual materials appropriate for elaborating new semantic analysis software;
f) To suggest an innovative model of the semantic analysis which would consider present achievements and overcome highlighted limitations.

2. To elaborate technical issues of the semantic analysis software in order to simplify its operations, make it user-friendly, and satisfying needs of the targeted consumers.
It is expected to be achieved by the following procedures:

a) To analyze technical issues of existing text data analysis tools;
b) To make research on users' experience, practices, beliefs, and needs about such kind of software;
c) To evaluate technical pros and cons of semantic analysis software by the method of the focus group;
d) To present a technical model of the semantic analysis software with the regards of the findings obtained.

3. To elaborate semantic analysis software adaptable for special needs of minority groups and people with disabilities:

a) To analyze sign language from the point of view of the correlation of the sign and its meaning;
b) To study the repertoire of the sign language and conduct intercultural investigations of its means and semantics;
c) To implement the findings based on the literature review, interviews, observations, teaching and communication practices, etc into designing new tools of semantic analysis software applicable for the usage of people with disabilities.

4. To design training programs enhancing users' competences required for operating semantic analysis software.

5. To enweave skills and competences of operating semantic analysis software into the professional competences of specialists by introducing special courses into the the training programs of interpreters, translators, language teachers, ICT engineers, etc.

Materials and Methods

The aims of the project can be achieved via an overall study of a) the semantic analysis software, b) models designed to elaborate text data analysis tools, c) existing practices of their implementation, and d) users evaluations.
Materials: semantic analysis software available for users' open access, language textual materials of different registers and modes
Methods: literature review, analysis and evaluation of the software, method of focus groups, observation, interview, discourse analysis, semantic analysis, syntactic analysis, contrastive analysis, survey, case studies, modeling, experimental method.