Rupkatha Journal on Interdisciplinary Studies in Humanities
Vol. 13, No. 4, 2021. 1-12
DOI: https://doi.org/10.21659/rupkatha.v13n4.65
First published on December 30, 2021

www.rupkatha.com

Modern Linguistic Technologies: Strategy for Teaching Translation Studies

Bilous O[1], Mishchenko A[2], Datska T[3], Ivanenko N[4], Kit L[5], Piankovska I[6] & Vereshchak Y[7]

[1]PhD, Full-Professor of German at the Department of Translation, Applied and General Linguistics, Dean of the School of Foreign Languages, Department of Translation, Applied and General Linguistics, the School of Foreign Languages, Volodymyr Vynnychenko Central Ukrainian State Pedagogical University, Ukraine.
[2]Phd, Full-Professor of German, Department of Translation, Applied and General Linguistics, the School of Foreign Languages, Volodymyr Vynnychenko Central Ukrainian State Pedagogical University, Ukraine.
[3]Doctor of Philosophy, Associate Professor, Department of the English Language and Methods of Teaching, Volodymyr Vynnychenko Central Ukrainian State Pedagogical University, Ukraine.
[4]Doctor of Philosophy, Associate Professor, Department of the English Language and Methods of Teaching, Volodymyr Vynnychenko Central Ukrainian State Pedagogical University, Ukraine.
[5]German Language Instructor, Department of Translation, Applied and General Linguistics, Volodymyr Vynnychenko Central Ukrainian State Pedagogical University, Ukraine.
[6]Doctor of Philosophy, Associate Professor, Head of the Department of the German Language and Methodology of Teaching, Volodymyr Vynnychenko Central Ukrainian State Pedagogical University, Ukraine.
[7]English Language Instructor, Department of Translation, Applied and General Linguistics, the School of Foreign Languages, Language Center Director, Volodymyr Vynnychenko Central Ukrainian State Pedagogical University, Ukraine.
Corresponding e-mail: gridina-98@bk.ru

Abstract
How often students use IT resources is a key factor in the acquisition of skills associated to the new technologies. Strategies aimed at increasing student autonomy need to be developed and should offer resources that encourage them to make use of computing tools in class hours. The analysis of the modern linguistic technologies, concerning intellectual language processing necessary for the creation and function of the highly effective technologies of knowledge operation was considered in the paper under consideration. Computerization of the information sphere has triggered extensive search for solving the problem of the use of natural language mechanisms in automated systems of various types. One of them was creating Controlled languages based on a set of features which made machine translation more refined. Triggered by the economic demand, they are not artificial languages like Esperanto, but natural simplified languages, in terms of vocabulary, grammatical and syntactic structures. More than ever, the tasks of modern computer linguistics behold creating software for natural language processing,

information retrieval in large data sets, support of technical authors in the process of creating professional texts and users of computer technology, hence creating new translation tools. Such powerful linguistic resources as corpora of texts, terminology databases and ontologies may facilitate more efficient use of modern multilingual information technology. Creating and improving all methods considered will help make the job of a translator more efficient. One of the programs, *CLAT* does not aim at producing machine translation, but allows technical editors to create flawless, sequential professional texts through integrated punctuation and spelling modules. Other programs under consideration are to be implemented in Ukrainian translation departments. Moreover, the databases considered in the paper enable studying of the dynamics of the linguistic system and developing areas of applied research such as terminography, terminology, automated data processing etc. Effective cooperation of developers, translators and declarative institutes in the creation of innovative linguistic technologies will promote further development of translation and applied linguistics.

Keywords: computer linguistics, linguistic technology, teaching, machine translation, integrated translation tools, CLAT technology.

1. INTRODUCTION

Modern society is defined as "Information Society" or "Knowledge-based society". Innovative technologies are the priority areas of such a society, and information, knowledge and social intelligence are the driving force behind dynamic development. "In real production and technological situations, at the present stage, whole institutions are created, which ensure the repeated transformation of information and its transportation, until it reaches the socio-technical system at the right moment and in the right form, adapted to the perception of the specified system" (Chyrokov 2005). Any knowledge in the form of information is verbalized and structured according to the laws of the linguistic system, so the creation of information-linguistic tools, tuned to the intellectual processing of the language and necessary for the creation and operation of highly efficient technologies of knowledge management, is an urgent need of modern times.

2. METHODS

Computerization of the information sphere, on the one hand, stimulated the development of computational linguistics aimed at solving the problems of the use of natural language mechanisms in automated systems of various types. On the other hand, the possibility of using a computer in linguistic research allows the analysis of large arrays of textual data to confirm or refute the theoretical achievements of modern linguistics and to study the phenomena of natural language (Knapp 2007). This is why computer linguistics, such as automatic natural language recognition (Allan 2006), is at the forefront; creation of formalized analyzers and synthesizers of natural language; development of automated information retrieval and data management systems; creation of automated ontologies capable of supporting research work, publishing specialist encyclopedic knowledge and bibliographic information, helping to create multidimensional classifications of material; machine translation; generation of electronic

corpora of text material; automatic annotation and abstracting of texts; extraction of information from texts, et.

The applied tasks of computational linguistics as a related field of linguistics and informatics are to create software products for intelligent language processing and technologies for creating, structuring, operating information arrays that can be transformed into knowledge. These technologies enable the transformation of the information resource into an instrument of economic growth of the developed countries and are an important means of preserving and spreading their cultural heritage. For this reason, linguistic technologies have been chosen as one of the priority areas of European Union policy (Pfister & Kaufmann 2008) and the national courses of the European countries, for the creation of which attract the best specialists of public and commercial research institutions from all over Europe through projects.

The main tasks of these projects are the pooling of resources of European countries and the exchange of technological advances to ensure the competitiveness of European Union countries in today's globalized world. Operating information resources in European Union languages using innovative linguistic technologies not only saves money on European firms and institutions but also further nurtures and develops the internal resources of the national languages of the European Community member states.

Due to the multilingualism of the EU organizations and the openness of the borders of European countries, special attention was focused on projects aimed at solving problems of machine translation and terminography.

3. FINDINGS

The history of machine translation systems was started by IBM in 1946 in Georgetown, USA. The purpose of the project was to create a Russian-English machine translation system for the Pentagon ("Maschinelle Übersetzung," n.d). However, due to the poor quality of such translations, it was concluded in 1966 that machine translation was not possible (ALPAC report) and the project was closed. They returned to the problems of machine translation only 10 years later in Japan and Europe, and 10 years later, after the first successes in the field of computational linguistics and artificial intelligence, there was an opportunity to create commercial systems of machine translation that allow: 1) written and oral automated translation of language pairs among a large number of languages (Promt, integrated translation systems with elements of machine translation SDL Trados, MemoQ, Across, DejaVu); 2) diversification of specialist work (eg, project management in integrated translation systems; creation of documents by technical editors; translation of text in SGML, XML, etc.; language control modules, for example, DUDEN-Korrektor; automated terminology management, e.g., Multiterm in SDL Trados or TermStar in Transit). ***Verbmobil*** project research on the possibilities of automatic translation. Modern machine translation systems are based on algorithmic and statistical methods, where the former are the result of formalizing the knowledge of the translator and are created as a result of cooperation of linguists and programmers; the second is the automatic extraction of similar language pairs after processing a large number of bilingual corpora of text material ("Maschinelle Übersetzung, "n.d; Lobin 2010)

The linguistic aspects of machine translation systems are based on such modules as 1) morphological (lemmatization of lexical units, search of lexical units in the dictionary, analysis of morphemes, recognition of context grammatical class of lexical units, distinctions, flexions, etc.); 2) syntactic (recognition of types of syntactic structures, relational relationships between individual elements of syntactic structure, etc.); 3) semantic (separation of the lexical meaning of multivalued lexical units and affixes, definition of their semantic function, synthesis of their syntactic uniqueness based on semantic analysis). Thus, the sequential stages of the operation of machine translation systems are analysis, transfer, synthesis, and interlingua ("Interlingua", n.d), which can be schematically illustrated in Figs. 1:

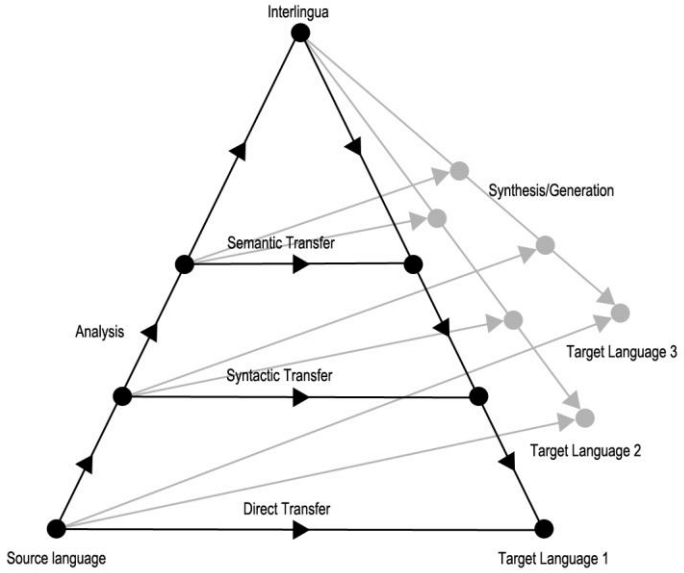

Fig. 1. Stages of operation of machine translation systems ("Prolog," n.d)

In 1976, the US-based platform SYSTRAN ("Systran," n.d)was acquired to launch research projects in the field of machine translation in some developed European countries such as Germany, England, France, Italy which was created to translate Russian-English texts in the military industry and NASA. Based on this platform a system of machine translation was created for pairs of languages *EN-FR, EN-IT, FR-EN.* Since 1978, SYSTRAN has been used in the Institutes of Germany and Canada, and for two years it has been the basis of the *JAP-EN, EN-JAP* machine translation system.

Based on the SYSTRAN platform, an international EUROTRA project was created in 1978, with the participation of working groups of researchers from Grenoble, Pisa, Saarbrücken, Manchester, etc. The project created a prototype machine translation system for 9 working languages of the European Union (72 pairs of languages), which worked on a limited vocabulary – 20,000 vocabulary articles in the telecommunications industry. For the interpretation of dictionaries and grammatical rules, the system architecture was created based on the logical programming language of the predictors of the mathematical logic of Horn's dispositions – Prologue (Freigang & Hellar 2010-11). This project became the prototype of the following projects, a detailed description of which does not allow the scope of this article.

The most famous experiment in the field of machine translation of verbal natural language, funded by the German Federal Ministry of Research and Technology (BMFT), known as ***Verbmobil*** (DFKI, Saarbrücken, 1993-2000). The purpose of the project was to create an automated oral translation system (*DE, ENG, JAP*) for translating simple dialogs into conversational topics such as "Hotel", "Appointment", "Getting Help", "Travel Planning" and more. To create the system, 10,000 of the most common German language tokens have been isolated, but individual features of speech, intonation, pronunciation, etc. have hindered the successful implementation of the project and have been discontinued ("Prolog," n.d).

4. DISCUSSIONS

Controlled languages based on simple syntax constructions, the predominant use of one-to-one tokens, unique synonyms and antonyms, and high-quality automated analysis + text synthesis allow greatly improve the quality of machine translation (Mishchenko 2012).

Controlled languages are not artificial languages like Esperanto, but natural simplified languages, artificially created based on a limited amount of vocabulary, grammatical and syntactic structures. The first controlled language, numbered 850, was created based on English in 1930 (Basic English). In the 70s of the twentieth century with the onset of the globalization process, controlled languages began to be generated as "corporate" languages to create and improve the quality of technical documentation, as well as to save money on its translation. This is achieved through the comprehension of professional texts, standardization of terminology, and improved quality of machine translation.

The standard for creating such languages was initiated by ***Caterpillar's Caterpillar Fundamental English (CFE)***. In 1986, ASD Simplified English (800 lexical units) was developed to meet the specific needs of the aircraft and rocket industry. Today, most international corporations create controlled languages to meet internal needs, eg, *Association Européenne des Constructeurs de Matériel Aérospacial (AECMA)*, *AeroSpace and Defense Industries Association of Europe (ASD)*, *Simplified Technical English (STE100, Boeing, Rolls-Royce, Saab Systems)* also used for *Dassault Aerospace: Français Rationalisé (FR), Ericsson: Ericsson English, General Motors (GM): Controlled Automotive Service Language (CASL), IBM: Easy English, Kodak: International Service Language, Siemens: Siemens Dokumentations Deutsch (SDD), Sun Microsystems: Sun Controlled English, Xerox: Xerox Multilingual Customized English* ("Prolog," n.d).

Controlled languages allow the use of additional tools that optimize the performance of technical editors through the use of controlled language verification programs and the translation of professional texts. Software for creating professional texts based on a controlled language is not a translation program. The main function of such programs is to monitor the observance of the rules of "corporate" language, for example, automatic search for "unauthorized" syntax constructions, terms and term elements, violations of stylistic characteristics of the text, automatic notification of "error" and alternative suggestions for its

elimination (CLAT technology Saarbrücken 2008, Manual; Saarbrücken 2008, Intro; Saarbrücken 2008, In-for-word, Acrocheck).

Let us dwell on the functional components of *CLAT (Controlled Language Authoring Technology)* technology. *CLAT* technology consists of a *CLAT server* and a *CLAT client* that can function independently (*Java CLAT-Client*) or in a text editor (*CLAT-Ins for PTC Arbortext and Microsoft Word*):

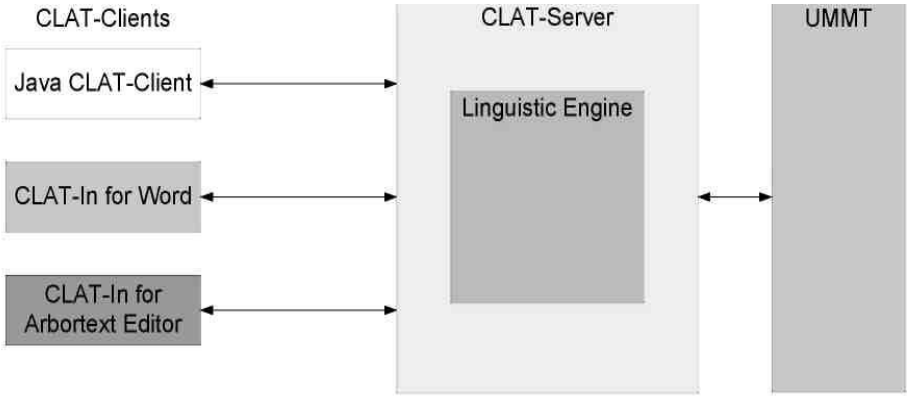

Fig. 2. Components of *CLAT* technology (Saarbrücken 2008, Intro).

CLAT allows technical editors to create flawless, sequential professional texts through integrated punctuation and spelling modules (*extra spaces, unnecessary or missing punctuation, old/new spelling, capital/lowercase, composing together: individually / hyphen, typographical errors*). Word borders are punctuated by spaces or spaces, and sentence boundaries, unless it is a structured file language (*SGML-/XML*), are punctuated and capitalized. Each sentence is decomposed into separate clauses of the sentence, depending on the position of the verb is determined by the type of sentence (narrative, interrogative, infinitive, etc.). At the heart of the *CLAT* control function is the correct definition of the sentence members.

Each word is analyzed by the program as a linguistic category: a grammatical class (*noun, verb, adjective, article*) is defined. For analysis, the word is decomposed into morphemes, each of which is compared with the morphological dictionary of the German language and automatically determines which component of the word in question (*root, morpheme, flexion*) and with what other components it can be combined. From the components according to the rules of German word formation, the word is synthesized by the program as a whole, while for one word the program can synthesize several variants, for example, the possible components of the word "*Weichen*" may be "*weich*" and "*en*", and also "*weiche*" and "*n*". From the components „*weich*" and „*en*" can be formed a definition (*die weichen Knie*), a verb (*er soll weichen*) or a noun (*beim Weichen auf den Vordermann achten*). Of the components "*weiche*" and "*n*" can be formed the noun (*die Weichen*). The grammatical analysis involves the control of the matching of the subject and the adverb, the article and the noun, the definition and the noun according to the norms of Duden.

The terminology control module allows consistent use of only a specialist corporation language, and also allows standardization of terminology and semiautomatic generation of terms in a narrow visual area (*filters of synonymous terms priority, reports of inconsistent term usage or use of "unknown" abbreviations, acronyms, etc.*). The *UMMS* terminology management component provides for the installation of additional terminology control filters and allows administrators to add new terms to the terminology database (Voltmer 2006).

The style control module documents the author's compliance with the stylistic norms of the controlled language (*active state instead of passive, hypotaxis instead of parataxis, modal words instead of equivalent forms of expression of modality, too long sentences, elliptical sentences, ambiguous sentences, etc.*). Error messages are displayed in different Word text editor windows:

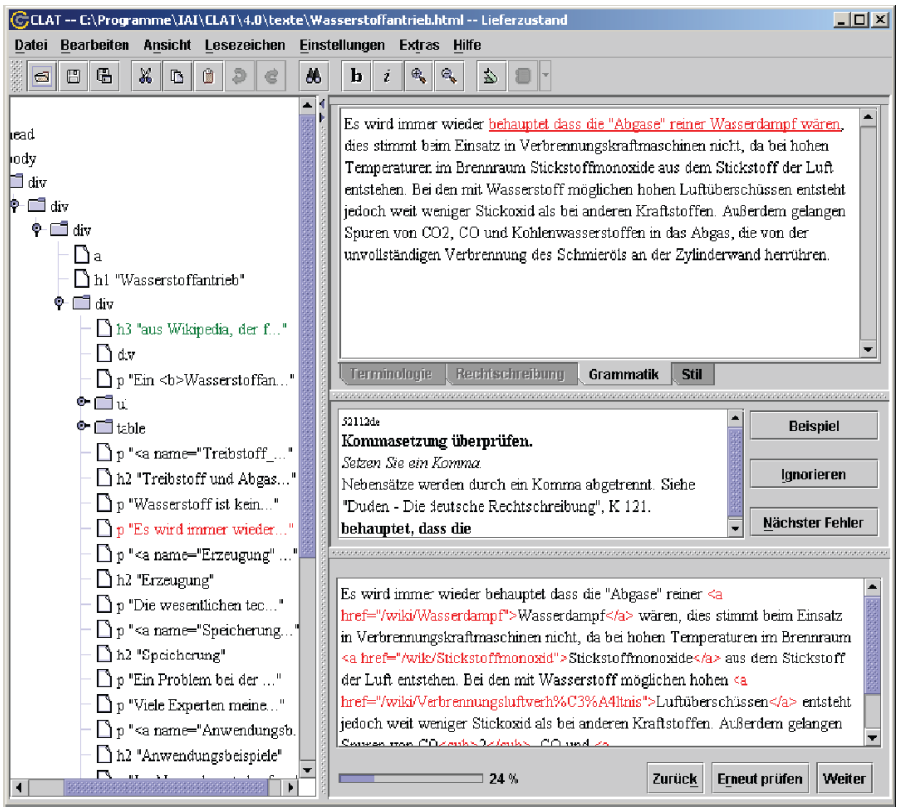

Fig. 3. *CLAT* technology interface with Word text editor (Saarbrücken 2008, In-for-word).

An additional *CLAT* function is an automatic extraction of "terms-candidates": the program displays in a separate window a text segment with a noun with the term attributes, which is missing from the terminology database both in terms and in non-consecutive terms. The technical author has the opportunity to evaluate whether the program should include a prototype term in the corporation's terminology database or send it for further analysis of the corporation's linguistic department terminologists.

The Sequence Control feature alerts the author to the inconsistent use of terminology or other lexical material in the text creation process:

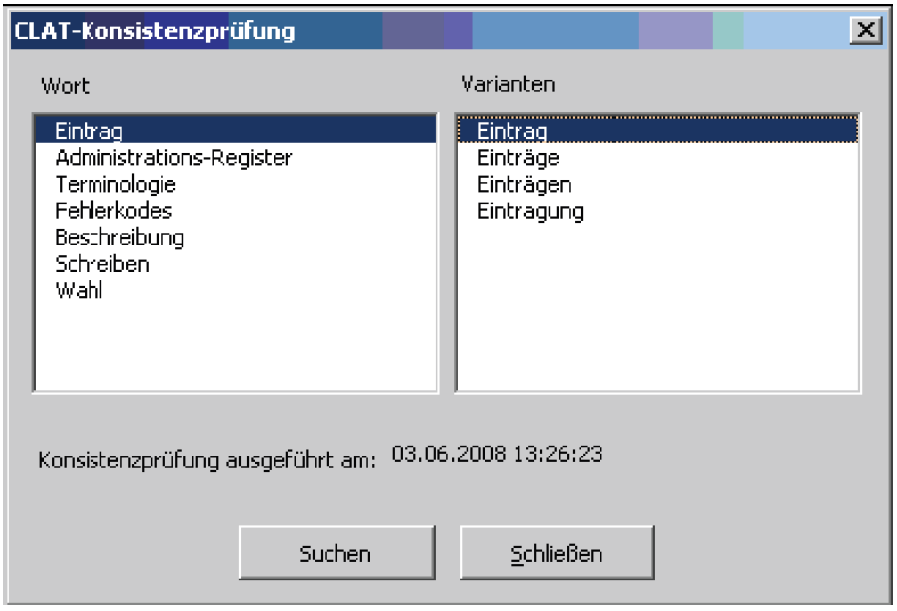

Fig. 4. The window for sequential use of lexical material (Saarbrücken 2008, In-for-word).

For example, words that are added to the same basis and may have the same meaning, but their form is slightly different: *Wasserstoff-Fahrzeug*, *Wasserstofffahrzeuge*, *Wasserstoff-Autos*. Consistent words are displayed in a separate window and highlighted in a specific color. Also, the author can use the search function to check any word or term in the document for inconsistent use.

Structured documents are reproduced based on *ISO* (*SGML, XML*, or *Word formatted documents*), and the structure of such documents is reproduced in a separate window – *DTD* (*Dokument-Type-Definition*). Markup elements of this type of document significantly improve the quality of control.

Such software products allow you to create quality texts, and clear syntactic, morphological and stylistic rules prevent variability, which in turn increases the likelihood of reusing text segments and reduces the cost and time of translating professional texts of this type.

The most common way to use a computer at a translator's workplace while working on a written translation is to use dictionaries, glossaries, terminology databases, and integrated translation systems. At the heart of the translation memory component is the "*Fuzzy-Match*" algorithm, the key function of which is to automatically save identical segments of the original text and the translation text in the translation process and to automatically identify identical or similar segments when translating other texts of this type that are displayed on the screen during the pre-analysis of the *Pretranslation*. Thus, the use of an integrated translation system facilitates the translation of professional texts of a certain field, which are characterized by the repetition of textual material (Abel & Glaznieks 2017).

The main components of commercial integrated systems are translation memory (*the so-called archive of previously translated texts*), which consists of modules such as **a)** Alignment – creating a parallel body of texts; b) extraction of terminology; **c)** exchange of translation memory

databases between different integrated systems; **d)** support for various data formats, electronic publishing systems, and data processing; **e)** project management; **f)** linguistic – terminological base and archive of translation. These state-of-the-art tools help you save costs, optimize your translator's work qualitatively and quantitatively, and free him from performing routine operations, such as creating a text-translation layout, copying illustrations, re-translating, and more (Reinke 2004; Kapanadze & Mishchenko 2013 a).

Components of terminology management recognize terms, the morphological paradigm of terms, their derivatives, and terms appearing in the constituents of composites, and allow their search in the terminology database, and then output the search result in the form of identified data (*Fuzzy-Match*).

Some components of terminology management are equipped with additional functions, for example, the function of automatically registering terms to a specific project, the function of automatic extraction of terms from mono or bilingual texts, the function of terminology control. The schematic functioning of integrated translation systems is illustrated in Fig. 5:

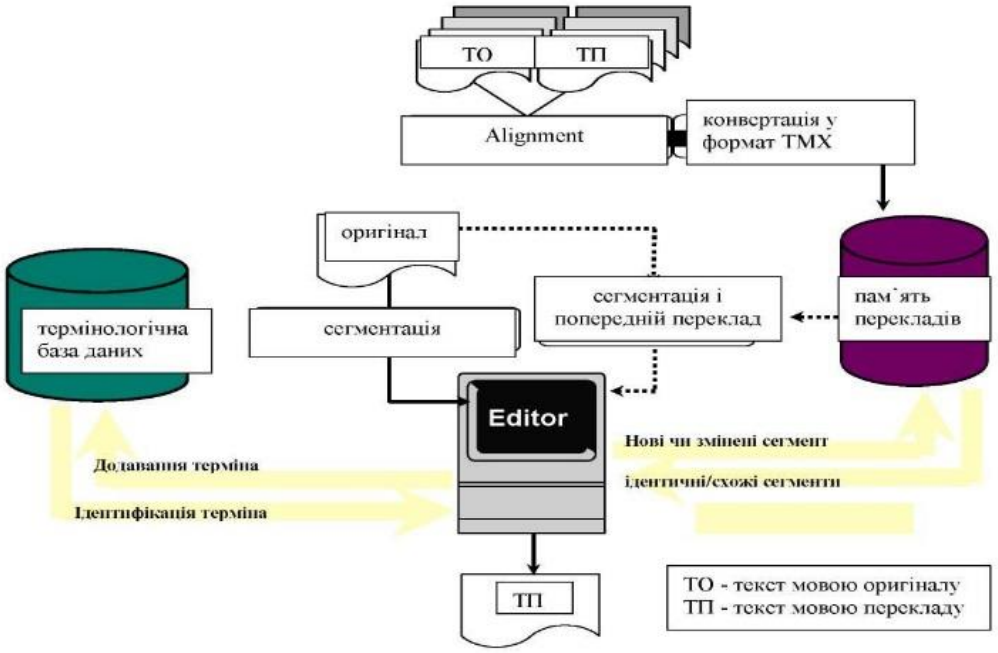

Fig. 5. The functioning of integrated translation systems ("Prolog," n.d).

Software localization uses specialized programs, such as *Passolo* (Reinke 2003), which allow the translation of dialog messages directly into software products and localization testing. For the translation of audiovisual material, they use specialized tools, such as Swift, which integrate separate components of translation memory but allow the creation of subtitles, their formatting on the screen following video standards.

5. CONCLUSIONS PROSPECTS FOR FURTHER RESEARCH FUNDING

Thus, the tasks of computer linguistics are to create software for natural language processing, information retrieval in large data sets (*Information Retrieval, Data-Mining, Informationsextraktion*), support of technical authors in the process of creating professional texts and users of computer technology under working with text editors, creating modern translation tools (Kapanadze & Mishchenko 2013 b).

Attraction of the latest software products from the best manufacturers of Germany, Spain and Hungary, which at the present stage are absent in the most Ukrainian educational establishments, allow to provide qualitative training of competitive specialists of the specialties "Translation" and "Applied Linguistics", able to work in the modern globalized society and use technology Translator and linguist tools that require a specialist, in addition to thorough professional training, specific technical competence, and because of the user's automated skills. The acquired practical skills of working with modern linguistic systems from the best world manufacturers will allow future specialists to work on the creation of similar Ukrainian software products, the development of which is an urgent need of the present.

The creation of databases of professional languages in Ukraine and the study of the features of their translation, as noted in our previous exploration (Bilous , Semeniuk , Mychyda & Bilous 2019), is a pressing issue for Ukrainian terminology, since in the Soviet Union all terminological databases were created only on the basis of the Russian language. However, since independence, Ukrainian has become the only official language, and the demand for translating professional language texts into Ukrainian has grown significantly. Thanks to a quality translation, Ukrainians can learn about current political, cultural (Sturz 2016) and economic events in Europe and the world, the latest developments in the fields of medicine, technology, ecology, sociology and more. This has influenced the distribution of audiovisual channels of international information such as Euronews, BBC, France 24, Deutsche Welle in the Ukrainian language, which to some extent also affects the level of integration of Ukraine into the European space.

To use modern multilingual information technology effectively, you need to create your powerful linguistic resources in the form of corpora of texts, terminology databases, and ontologies. Consistent accumulation of electronic linguistic resources will create search engines that can replace traditional printed professional vocabularies. Besides, these databases allow you to study the dynamics of the linguistic system and to develop areas of applied research such as terminography, terminology, automated data processing.

It should also be mentioned that Central Ukrainian State Pedagogical University is the only higher educational establishment in Ukraine, the Translation Department of which applies *CLAT (Controlled Language Authoring Technology)* technology, which enables our students to master the most advanced translation techniques, and as highly qualified graduates are involved in a number of national and world-wide projects, aimed at honing the profession of a translation to perfection.

The experience of involving developers, translators and declarative institutes in the creation of innovative linguistic technologies illustrates the EU, and the effectiveness of such cooperation proves that this experience is worthy to be implemented in Ukraine.

REFERENCES (TRANSLATED AND TRANSLITERATED)

Chyrokov V.A. (2005) Elementy Leksykographii . – Kyiv: Dovira.

Knapp, Karlfried et.al. (Hrsg.) (2007). Angewandte Linguistik// Ein Lehrbuch/2.Auflage. A.Francke Verlag Tübingen und Basel.

Allan, Keith (2006). Natural Language Semantics.- Malden, Massachusetts: Blackwell Publishers Inc.

Pfister, B., Kaufmann, T. (2008). Sprachverarbeitung/ Grundlagen und Methoden der Sprachsynthese und Spracherkennung.Berlin, Heidelberg: Springer-Verlag.

Maschinelle Übersetzung: http://de.wikipedia.org/wiki/Maschinelle_%C3%9Cbersetzung.

Lobin, H.(2010). Computerlinguistik und Texttechnologie. Paderborn: Wilhelm Fink.

Interlingua http://ru.wikipedia.org/wiki/Interlingua.

Prolog: http://ru.wikipedia.org/wiki/Prolog.

Systran: http://de.wikipedia.org/wiki/OpenLogos.

Some materials and drawings can be found in 2010-2011 academic year lectures of graduate translator K. H. Freigang and H. Haller - the professor of the Saarland University of (Saarbrücken, Germany).

Mishchenko A. (2012). Kontrollierte Sprachen und multilinguale technischeDokumentation / A. Mishchenko // Sprache & Sprachen. – Halle-Wüttenberg : Gesellschaft für Sprache und Sprachen (GeSuS) e.V., Martin-Luther-Universität.

CLAT-Client-Manual (2008). – Saarbrücken: IAI.

CLAT-Intro (2008). – Saarbrücken: IAI.

CLAT-In-For-Word-Manual (2008). – Saarbrücken: IAI.

Voltmer, L. (2006). Computerlinguistik für die Terrminologie im Recht.-Tübingen: Gunter Narr Verlag.

Abel, A., Glaznieks, A.(2017) Koko: Bildungssprache im Vergleich: Korpusunterszützte Analyse der Sprachkompetenz bei Lernenden im deutschen Sprachraum – ein Ergebnisbericht.- Eurac Research .

Reinke U. (2004).Translation Memories : Systeme – Konzepte – linguistische Optimierung // hrsg. Fachrichtung Angewandter Sprachwissenschaft sowie Übersetzen und Dolmetschen der Universität des Saarlandes. – Sabest : Saarbrücker Beiträge zur Sprache- und Translationswissenschaft. Bd. 2. – Frankf./M.; Berlin [u.a.] : Peter Lang.

Kapanadze O., Mishchenko A. (2013 a). Multilingual GRUG Treebank forUnderresourced Languages / Oleg Kapanadze, Alla Mishchenko // Computational Linguistics and Intelligent Text Processing. - Volume 7816 / ed. By Alexander Gelbukh. – Springer.

Reinke U. (2003). Passolo 4.0 – versteckte Highlights // Mitteilungen für Dolmetscher und Übersetzer. – Jg. 49. – № 6.

Kapanadze O., Mishchenko A. (2013 b). A parallel Treebank for less resourcedGeorgian and Ukrainian Language / Oleg Kapanadze, Alla Mishchenko // CognitiveStudies | Etudes cognitives. Vol. 13 / ed. by Roman Roszko. – Institut of Slavic Studies : Katowice.

Bilous O.M., Semeniuk O.A., Mychyda S.P., Bilous O.I. (2019). "DEUTSCHE WELLE" AS A MASS MEDIA LANGUAGE FOR SPECIALIST PURPOSES: PRAGMATICS, ANALYSIS AND TRANSLATION // Humanities @Social Sciences Reviews. – Vol 07, No 06 (2019).

Sturz, W. (2016). Kulturelle Unterschiede in technischen Terminologien// Terminologie und Kultur/ Deutscher Terminologie-Tag e.v., M[nchen/Köln/Karlsruhe 2016.