

Polish Translation Software by Lingvistica '98 Inc.

EUGENE GULAK.

Kharkov State Pedagogical University. Kharkov, Ukraine

NADIYA BYEZHANOVA

MICHAEL S. BLEKHMAN

Lingvistica '98 Inc. Montreal, Canada

ALLA RAKOVA

ANDREI KURSIN

Lingvistica '98 Inc., Ukrainian Affiliation. Kharkov, Ukraine

Computational linguistics and language engineering are embracing more and more language areas and languages. In our previous publications, we described translation software for two languages that represent the Slavic group: Russian and Ukrainian. The present paper touches upon another language that belongs to the same family - Polish.

Poland is a fast developing country in the middle of Europe, a nation with a great history and culture. Its language is more or less close to Ukrainian and, to a smaller degree, to Russian, while, unlike both of the above, it uses the Latin characters instead of Cyrillic.

In 2000-2001, on the order from France Telecom, we developed a large grammatical dictionary of approximately 120,000 Polish words and phrases, in which each Polish lexical unit has a detailed grammatical and phonetic description:

- part of speech
- subclass characteristics, such as noun gender, verb transitivity, etc.
- declension/conjugation features
- pronunciation description, or, in other words, transcription.

The dictionary is based on the newest orthographic and bilingual dictionaries published in Poland as well as on the formal descriptive grammar of the Polish language developed by Prof. Anatoly Gulak and Dr. Eugene Gulak. The grammar describes most of the morphological phenomena and major syntactic relations in modern Polish.

Both the grammar and the grammatical dictionary (excluding its phonetic description) have been used in our recent project -

"English↔Polish Translation Software". The project was partially financed by the LogoMedia Corp., a company based in Boston, Mass., and well known for its machine translation software covering a number of European and Asian languages. The resulting English↔Polish translation engine is presently used in the LogoMedia Translate system and in our company's newest product - PARS/Polish.

This paper describes the first two products in our family of Polish software developed in 2002:

- PARS/P - an English↔Polish translation system
- EleGloss Polonaise - an English↔Polish general dictionary

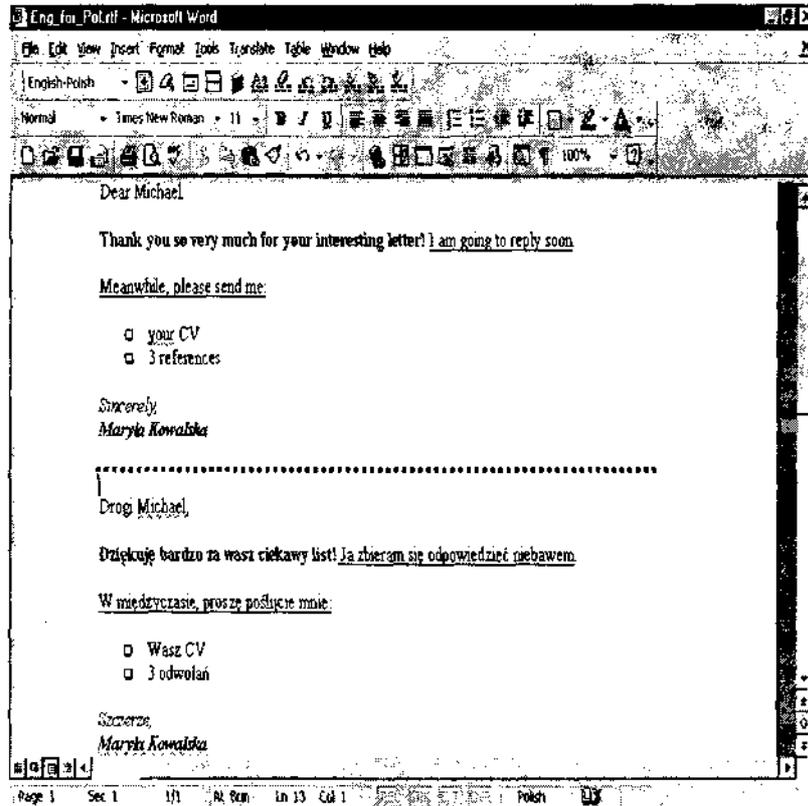
WHAT IS PARS/POLISH

PARS/P belongs to a well-known family of machine translation systems capable of translating between English and Polish, Ukrainian, and Russian as well as between German and Russian and Ukrainian.

As any machine translation software, PARS/P is not "artificial intelligence": it can only generate what is called "draft translation". In other words, the grammar laid in the foundation of the system only includes the major descriptive rules, which sometimes results in poor translation. Besides, which is another drawback, PARS/P makes use of very few semantic rules, which doesn't enhance translation quality either. On the other side, this comparative simplicity has made it possible for us to come up with a non-expensive product producing decent translation level.

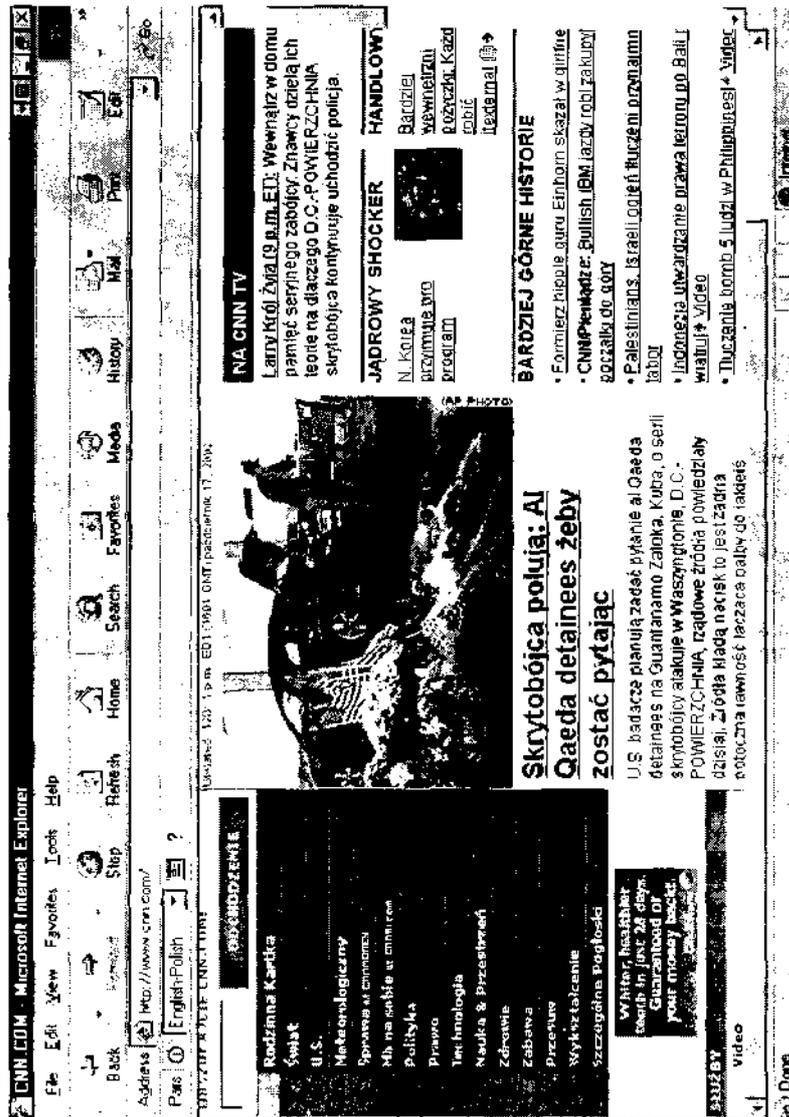
As any PARS family system, PARS/Polish has several specific features:

- It is integrated with the latest MS Word versions: Word 97,2000, XP. By integration with MS Word, we mean that PARS/P can be activated directly from Word and translate the source text into another Word window preserving text formatting including bullets and numbering, font styles (bold, italic, underlined), footnotes, etc. Sometimes, this turns out to be a very serious achievement making it possible for the end user to avoid reformatting the machine-made translation, which sometimes would not be much simpler than translating as such.



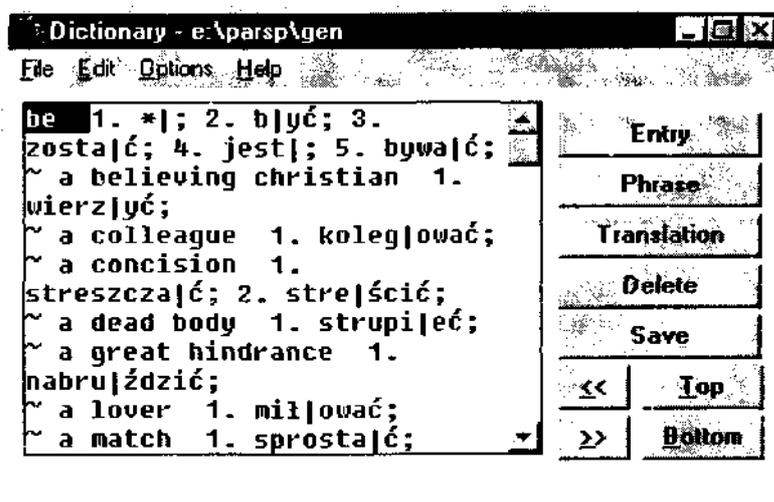
- PARS/P is also integrated with Internet Explorer 5.5 and above. This is rather convenient in those cases when the user needs "linguistically draft" translations of web pages fully preserving the page format. The result is "the same" web page in the target language - although linguistically imperfect, as any modern machine translation. Another convenience is the user-friendly design: the translation buttons and menu are located directly on the Internet Explorer toolbar.

The screenshot below shows an English web page translated into Polish by PARS/P, fully preserving the page format:

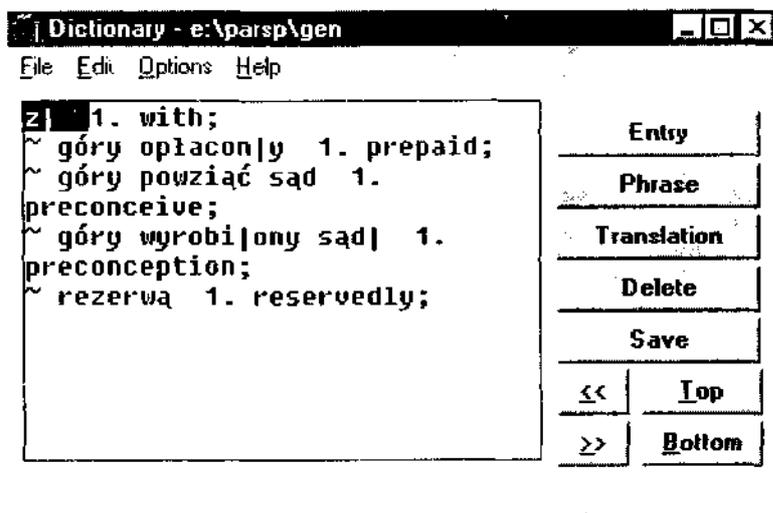


Besides translating in MS Word and Internet Explorer, PARS/P (as any PARS system) provides the so-called drag-and-drop mode: the source text is dragged to the PARS icon and dropped upon it, after which the target language translation appears in a separate PARS window.

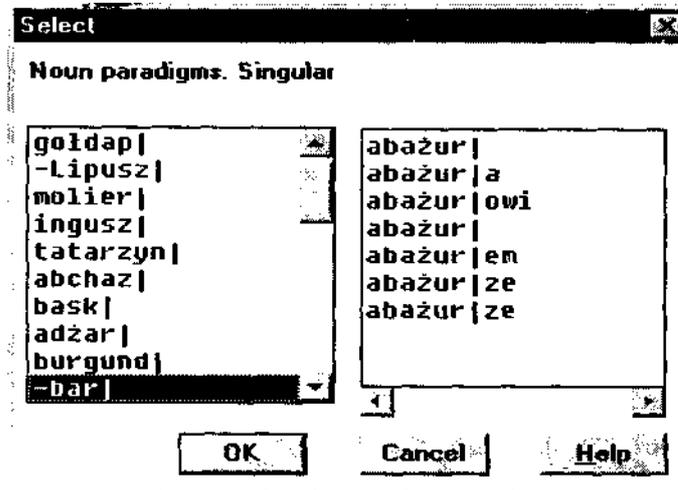
Another PARS/P peculiarity is the user-friendly dictionary editor, which lets the user enter and delete words and phrases and edit the dictionary entries. Here is a screenshot that illustrates the PARS/P dictionary editor entry: the English word *be*, its Polish translations, English phrases beginning with *be*, and their Polish translations:



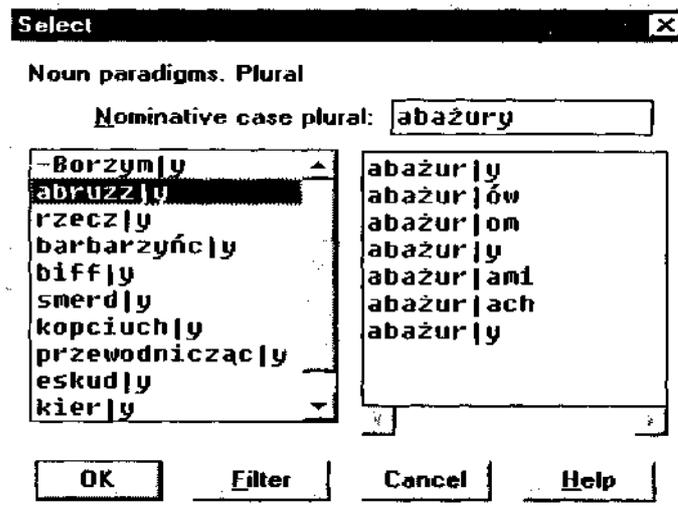
Any PARS dictionary entry, PARS/Polish in particular, is bi-directional. Here is one of the Polish-English entries:



The Polish words in the PARS/P dictionary have grammatical features "borrowed" from the above-mentioned Polish grammatical dictionary. For example, the nouns have gender and declension features; the latter are provided both for the singular and plural forms, for example, for the Polish noun *abazur*.

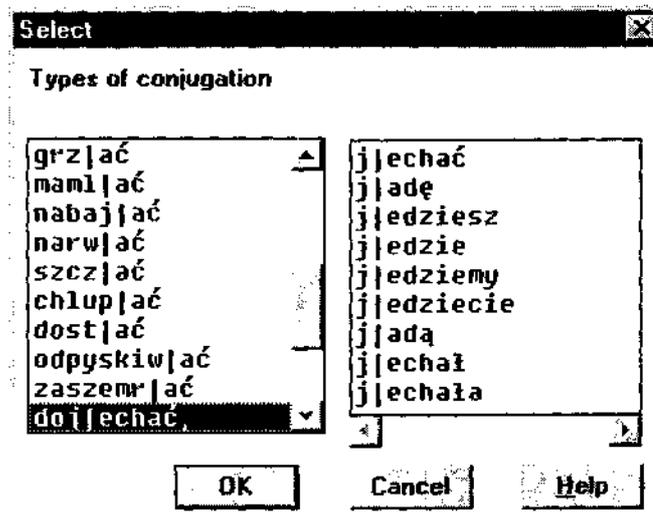


Plural declension paradigm

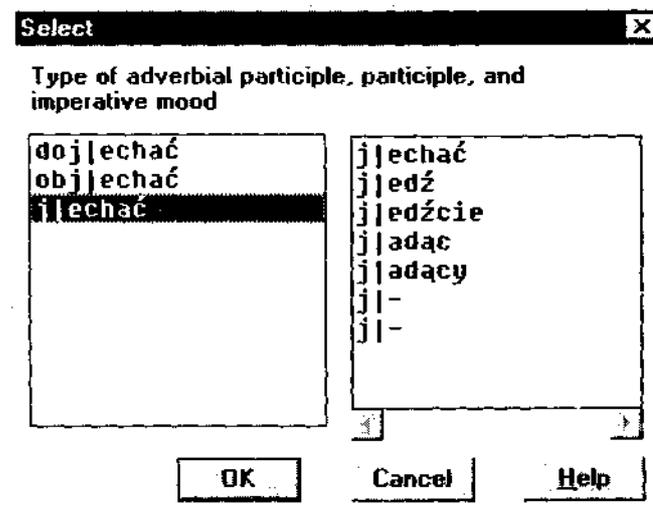


Singular declension paradigm

The verbs have conjugation features for indicative forms as well as for the verbals (excluding the infinitive, which is included in the indicative paradigm) and the imperative mood. Here is an example -the Polish verb *jechać*.



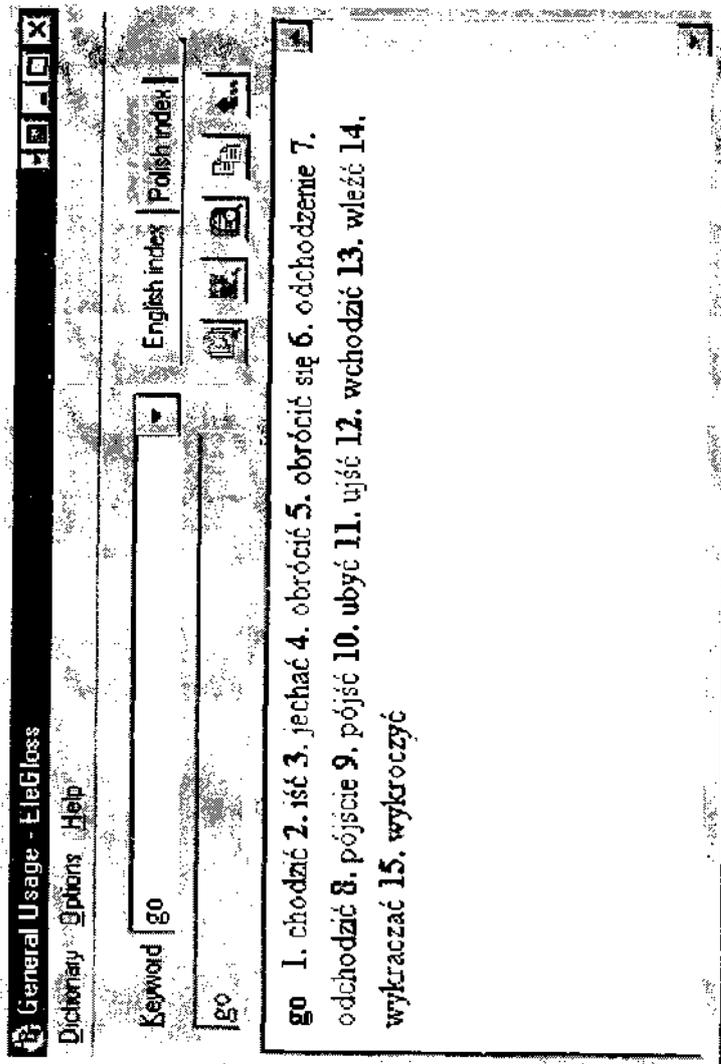
Verbals and imperative



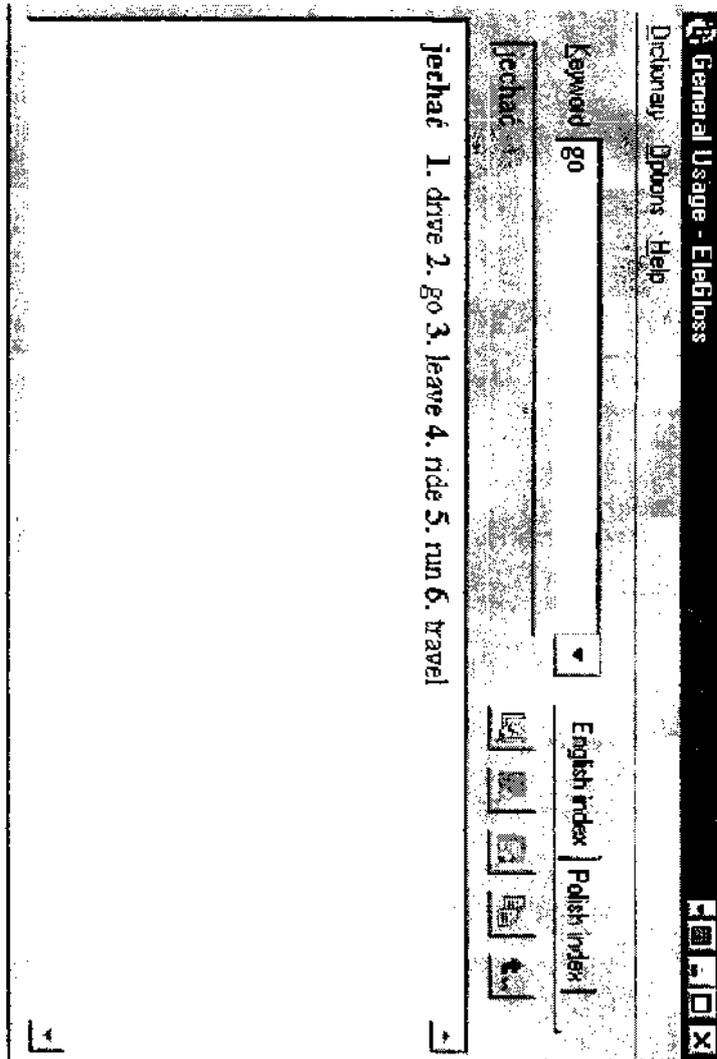
Indicative paradigm

ELEGLOSS POLONAISE

EleGloss is in a way a “byproduct” of PARS/Polish. It is a simple-to-use, user-friendly bi-directional dictionary that lets the customer find Polish translations of English words:



and English translations of Polish words:



The Copy/Paste mode lets the dictionary communicate with the text editor, such as MS Word, for pasting the translations directly into the text.

EUGENE GULAK

Kharkov State Pedagogical University. Kharkov, Ukraine

E-mail: gulak@nord.net.ua

NADIYA BYEZHANOVA

MICHAEL S. BLEKHMAN

Lingvistica '98 Inc. Montreal, Canada

E-mail: ling98@canada.com

ALLA RAKOVA

ANDREI KURSIN

Lingvistica '98 Inc., Ukrainian Affiliation. Kharkov, Ukraine

E-mail: kursin@kharkov.ua