

Ewa Gruszczyńska  
Uniwersytet Warszawski

Agnieszka Leńko-Szymańska  
Uniwersytet Warszawski

Ruprecht von Waldenfels  
University of California, Berkeley

## The Polish-Swedish and Swedish-Polish Parallel Corpus for exploring language contacts through translation

---

### Polsko-Szwedzki i Szwedzko-Polski Korpus Równoległy w badaniach kontaktów językowych poprzez tłumaczenie

#### Abstract

Artykuł ma na celu zaprezentowanie prac związanych z powstawaniem korpusu równoległego współczesnych polskich i szwedzkich tekstów literackich. Przedstawia także wyniki badania pilotażowego porównującego za pomocą danych równoległych językowe wykładniki emocji w obu językach i ich wzajemnych tłumaczeniach.

Polsko-szwedzki i szwedzko-polski korpus równoległy powstaje w Pracowni Badań Skandynawistycznych na Wydziale Lingwistyki Stosowanej Uniwersytetu Warszawskiego. Planowany jest na około 10 milionów tokenów i wykorzystywany będzie w badaniach dotyczących powiązań językowych w tłumaczeniach oraz wpływu przetłumaczonych tekstów na wzajemne postrzeganie języków i kultur. Zawierać będzie polskie i szwedzkie teksty literackie opublikowane w obu językach w ostatnich 20 latach wraz z ich tłumaczeniami na oba języki. Wersja pilotażowa korpusu liczy obecnie około 750 000 wyrazów i obejmuje trzy współczesne szwedzkie powieści przetłumaczone na język polski oraz jedną powieść i 14 opowiadań w języku polskim wraz z ich szwedzkimi przekładami. Minikorpus został zrównoleglony na poziomie zdań przy użyciu pakietu LFAAligner 4.0, a jego polska część została otagowana przez Treetagger. Interfejs został oparty na pakiecie PARAVoz, oryginalnie stworzonym dla projektu ParaSol.

Badanie pilotażowe z wykorzystaniem minikorpusu osadzone zostało w teorii wymiaru kultur, której autorem jest Geert Hofstede. Przeprowadzono je pod kątem sposobu tłumaczenia na oba języki wybranych jednostek leksykalnych związanych z emocjami z pola semantycznego polskiego rzeczownika *strach* oraz szwedzkiego *skräck*. Wyrazy odnoszące się w obu językach do tej emocji zostały

uszeregowane pod względem intensywności. Następnie rzeczowniki w języku polskim zostały zestawione z ich szwedzkimi ekwiwalentami tłumaczeniowymi występującymi w korpusie i porównane pod względem mocy. Ta sama procedura została zastosowana dla rzeczowników polskich i ich szwedzkich odpowiedników. Wydaje się, że wyniki badania, które ze względu na niewielką objętość korpusu należy traktować bardzo ostrożnie, potwierdzają hipotezę, według której różnice w emocjonalności wyrażanej językowo po polsku i po szwedzku mają odzwierciedlenie w przekładach. Kultura szwedzka charakteryzuje się słabszym i bardziej stonowanym sposobem wyrażania emocji w porównaniu z kulturą polską. Tłumacze polscy wybierają zazwyczaj ekwiwalenty, które (biorąc pod uwagę parametr intensywności) są silniejsze od jednostek językowych użytych w szwedzkim oryginale. Widoczna jest także odwrotna tendencja w tłumaczeniach z języka polskiego na język szwedzki. Oznacza to, że oprócz innych wymiarów wskazanych przez Geerta Hofstedeego, kultura polska i szwedzka różnią się także pod względem emocjonalności. Jednak, aby potwierdzić wiarygodność wstępnych wyników, badania zostaną powtórzone na dużym korpusie docelowym.

**Keywords:** parallel corpus, Polish, Swedish, emotions, translation

**Słowa kluczowe:** korpus równoległy, język polski, język szwedzki, emocje, przekład

## 1. Background

The Scandinavian Research Centre at the Faculty of Applied Linguistics is currently launching a research project dedicated to contemporary Polish-Swedish language contacts through translation. The data for our investigations will primarily come from a purpose-built parallel corpus of literary texts. The aim of the project is to examine Polish-Swedish and Swedish-Polish linguistic relations in translation as well as the impact of translated texts on the mutual perception of our respective languages and cultures.

A parallel corpus of Polish-Swedish and Swedish-Polish translations is being built as a means to study these issues. Many such resources have already been compiled, as corpora have become an indispensable source of data in linguistics and translation studies. However, to date there is no one dedicated Polish-Swedish parallel corpus. Multilingual corpora that include a Polish-Swedish component are insufficient. The segment in *ParaSol* (von Waldenfels, 2011) is clearly too small, while *Opus* (Tiedemann, 2012), and the *Aquis Communautaire* corpus (Steinberger et al., 2006) include specialized language such as technical or scientific documents and film subtitles (*Opus*) or the EU legislation (*JTC-Aquis*), in both cases mostly translated from third languages, and thus not suitable for investigating language and culture relations between Sweden and

Poland. The ASPAC Swedish-Polish corpus in Språkbanken (the Swedish Language Bank), which is part of the *Amsterdam Slavic Parallel Aligned Corpus*, and which consists of 1,467,368 tokens (102,146 sentences), is also too small for larger-scale comparisons of Polish and Swedish. Additionally, it is not perfectly aligned and therefore query results are not reliable. Thus, there is an apparent need for a large, reliable, representative and dedicated corpus of translations into and from both languages. This gap will be filled by the resource compiled within the framework of our project.

## 2. The Polish-Swedish parallel corpus

The corpus will consist of Swedish and Polish contemporary literary texts and their translations into Polish and Swedish. The intended size of the corpus is 10 million tokens – 5 million Swedish originals with their Polish translations and 5 million Polish originals with their Swedish translations. Therefore, we estimate that the Swedish-Polish component will include about 30 original Swedish books with their Polish translations and a similar number of volumes is foreseen for the Polish-Swedish component. The literary texts to be included in the corpus are selected from a bibliography of contemporary (i.e. last 20 years) Swedish and Polish literature which has been translated into the respective languages<sup>1</sup>, and an effort will be made to ensure the inclusion of a variety of genres, authors and translators so as the corpus is balanced and representative. Each text in the corpus will be appended with rich metadata (the information on its author/translator, its source, etc.), as well as with structural and linguistic information, such as the basic text structure and part of speech tagging. The originals and their translations will also be aligned at the sentence level. Purpose-built corpus-analysis tools will offer opportunities for multiple searches based on a range of queries (such as individual words, phrases, parts of speech, units of texts), and for direct comparisons between texts in the two languages, which will be facilitated by the option of viewing the aligned sections of texts side-by-side.

In future, the Swedish-Polish and Polish-Swedish parallel corpus may be further developed and used for other research in translation studies between Polish and Swedish languages. Multilingual text collections, in particular parallel corpora, have proved to serve not only as an excellent resource for the descriptive study of translation (Baker, 1995; Kenny, 1998), but also as a basis for professional pedagogical applications in the field of translator training (Pearson, 2003;

<sup>1</sup> A bibliography of Swedish-Polish contemporary literary translations (2000 -2015) has been already compiled by Anna Sworowska (Gruszczyńska, Sworowska, 2015) and is part of the monograph: Ewa Gruszczyńska (2015) *Polsko-szwedzkie spotkania językowe za pośrednictwem przekładu*. The earlier bibliography of Swedish-Polish literary translations prepared by Hieronim Chojnacki (2003) *Szwedzka literatura piękna w Polsce 1939-1996* does not include the period of the last 20 years.

Bowker, 1998; Zanettin, 1998). The texts gathered in the parallel corpus will initially be available to the research team only. Sections of the corpus will gradually be made publically available as the copyright issues are cleared.

### 3. The mini-corpus

Before embarking on the large-scale compilation of the Polish-Swedish parallel corpus of literary texts, a decision was made to build a mini-corpus of a few hundred thousand words. This was done with the aim of verifying the feasibility of the project, testing its individual procedures and assessing its technical demands. In addition, using the mini-corpus for the pilot study described in the next section was considered an important step in testing if the architecture of the final resource will be optimal for the kinds of tasks envisaged within the research project. Finally, it was also hoped that the compilation of the mini-corpus will enhance the credibility of the project and thus help us raise necessary funding.

The mini-corpus was compiled in January-March 2015. It includes three contemporary Swedish novels (by Sven Delblanc, Stig Larsson and Kerstin Ekaman) with their Polish translations, as well as one Polish novel (by Olga Tokarczuk) and a selection of Polish short stories with their Swedish translations. The number of tokens in the mini-corpus and in its individual sections is presented in Table 1:

Polish-Swedish		Swedish-Polish	
Polish	Swedish	Swedish	Polish
81,827	98,704	320,768	284,174
<b>Polish</b>		<b>Swedish</b>	
366,001		419,472	
<b>Total</b>			
785,473			

Table 1. Number of tokens in the Polish-Swedish parallel mini-corpus

The procedures and the tools applied for the construction of the mini-corpus were adopted from the compilation project of the German-Polish parallel corpus (see Chapter 6). After scanning and OCR conversion performed with ABBYY FineReader, the text files were checked manually. A header containing meta-data was produced for each document and inserted manually. Subsequently, the texts were aligned with LFAaligner 4.0<sup>2</sup> and the accuracy of the procedure was verified by two researchers speaking both languages. The aligned documents in the TMX format were then converted to two separate text files containing XML annotation, one for each language. The Polish corpus file was tagged using Treetagger<sup>3</sup> (Schmid, 1995). Unfortunately, Treetagger does not offer a para-

<sup>2</sup> <http://sourceforge.net/projects/aligner/>

<sup>3</sup> <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

meter file for Swedish, so it could not be used for tagging the Swedish part of the corpus. Several other available taggers were tested – the Stockholm Tagger<sup>4</sup> (Östling, 2013), TnT<sup>5</sup> (Brants 2000; Megyesi, 2001), HunPos<sup>6</sup> (Halacsy, Kornai, Oravecz, 2007; Megyesi, 2008), but none of them appeared to work well with files containing XML annotation. Given the pilot nature of the current project, we did not adapt these tools for our purpose and abandoned tagging the Swedish data. Finally, the two files containing Polish and Swedish texts separately were converted to the CWB format required by the IMS Open Corpus Workbench<sup>7</sup> (Evert, Hardie, 2011) – a set of tools for managing and querying large text corpora with linguistic annotation. The interface for querying the data and viewing the results was based on PARAVoz<sup>8</sup>, (Meyer et al., 2006-2015, see also Chapter 5), initially developed for *ParaSol* (von Waldenfels, 2011). It is a simple CWB-based interface for parallel corpora operating through a web browser. At the moment the mini-corpus is running on our local server. Figure 1 presents screenshots from the query interface and the result-viewing panel.

The compilation of the mini-corpus has pointed to several problems which will need to be taken into account in the proper compilation phase. First, in order to ensure the dependability of query results, manual checks have to be foreseen in the project after the OCR conversion and alignment stages. It has become clear that the automatic tools alone do not produce sufficient quality, as too many errors occurred at both stages to be left unedited without compromising the accuracy of the resource. Thus, it is necessary to secure adequate time and financial resources for this purpose. There is also a need for a simple script for automatic generation of headers from the information gathered in a separate database. Next, it is essential to solve the problem with tagging the Swedish data by developing a tool stripping XML annotation before tagging and restoring it into the tagged files. Finally, using the mini-corpus for the pilot study has revealed that the one-sentence context available at the moment is sufficient only for an initial examination of the data. More in-depth analyses planned in the project require access to larger – at least one-paragraph-long – contexts, which is not supported by the current interface. An option of viewing a larger context has to be included in the new version of the interface. Addressing these problems will have a positive impact on the efficiency of work done within the projects and the quality of its final result.

---

4 <http://www.ling.su.se/english/nlp/tools/stagger>

5 <http://www.coli.uni-saarland.de/~thorsten/tnt/>, <http://stp.lingfil.uu.se/~bea/resources/tnt/>

6 <https://code.google.com/p/hunpos/>, <http://stp.lingfil.uu.se/~bea/resources/hunpos/>

7 <http://cwb.sourceforge.net/>

8 <https://bitbucket.org/rvwfels/paravoz>

Kłosa <b>nigdy</b> nie chciała położyć się po bożemu. [Olga Tokarczuk: <i>Prawiek i inne czasy</i> ]	Axa ville aldrig lägga sig ned som Herren anbefallt.
Popatrzyło na nią tak, jak nikt jeszcze <b>nigdy</b> na nią nie patrzył: z ogromną, niewyraźną miłością. [Olga Tokarczuk: <i>Prawiek i inne czasy</i> ]	Barnet såg på henne som ingen förut hade sett på henne: med en gränslös, ousäglig kärlek.
Taka miłość <b>nigdy</b> nie mija. [Olga Tokarczuk: <i>Prawiek i inne czasy</i> ]	Sådan kärlek går aldrig över.
<b>Nigdy</b> nie mogła się zdecydować, w jakim świecie wolałaby żyć: w zielonym, rubinowym, błękitnym czy złotym. [Olga Tokarczuk: <i>Prawiek i inne czasy</i> ]	Hon kunde aldrig bestämma sig för vilken värld hon helst ville leva i: den gröna, den rubinröda, den blå eller den gula.
Zwierzęta <b>nigdy</b> nie wchodziły sobie w drogę. [Olga Tokarczuk: <i>Prawiek i inne czasy</i> ]	Djuren gick aldrig i vägen för varandra.
Gdziekolwiek Michał jechał, zawsze myślał o Misi, zawsze do niej tęsknił. <b>Nigdy</b> nie podniósł na nią głosu. [Olga Tokarczuk: <i>Prawiek i inne czasy</i> ]	Vart Michał än åkte tänkte han på Misa och längtade alltid till henne, höjde han rösten mot henne.
Dziedziec Popielski <b>nigdy</b> nie widział takiej kostki do gry. [Olga Tokarczuk: <i>Prawiek i inne czasy</i> ]	En sådan speltärning hade godsägare Popielski aldrig sett.
Misia <b>nigdy</b> przedtem nie myślała, że jest maleńka. [Olga Tokarczuk: <i>Prawiek i inne czasy</i> ]	Aldrig tidigare hade Misa tänkt att hon var liten.

Figure 1. Screenshots from the Swedish-Polish mini-corpus

#### 4. A pilot study

The pilot study described below belongs to the area of research which deals with the so-called “linguistic images of the world”. These “images” are generally defined as a set of language properties related to grammatical categories (morphological and syntactic) as well as lexical devices which reveal specific images of the elements of the world typical for a certain language and culture (cf. Wierzbicka, 1999b). The study focuses on exploring the expression of selected emotions in both languages and it is based on the Swedish-Polish and Polish-Swedish parallel mini-corpus described in the previous section. Its aim is to investigate if there is a difference in the conceptualization of emotions in the Swedish and Polish cultures and languages, and if this difference influences the way in which lexical units denoting emotions are translated into the respective languages.

Emotions<sup>9</sup> are a significant part of the world. As they are language- and culture-specific it is interesting to investigate how they are conceptualized in Swedish and Polish and how they are translated between these two languages. Although the problem of transferring emotions has always been present in some way in the literature on translation, most studies have been concerned with a general assessment of “the spirit” of a text and the impression a text makes on the reader rather than with specific emotions (cf. Bassnett-Mc Guire, 1980: 63). A greater interest in emotions within translation studies has been prompted by contemporary semantics and a number of studies devoted to this issue has gradually increased in recent years. But unlike linguists, who have been especially interested in the affective lexicon, i.e. in words referring to emotions such as *fear* or *sadness* (cf. Clore, Ortony, Foss, 1987; Johnson-Laird, Oatley, 1989; Wierzbicka, 1990, 1991, 1992a, 1992b, 1998), most translation theorists have focused on emotionally-loaded lexical units. Thus, research on lexical units referring to emotions is still scarce in translation studies (cf. Gruszczyńska, 2001). The question about what happens to the affective lexicon in the process of translation from Swedish into Polish and vice versa seems pertinent and interesting. The subject matter of this pilot study has been limited to the emotions from one semantic field: ‘fear’ i.e. Polish *strach* and Swedish *skräck*. We analyse the occurrences of lexical units belonging to this semantic field in the parallel mini-corpus, thus focusing on the textual realisations of these sentiments. As the differences between Polish and Swedish cultures are significant<sup>10</sup>, it can be expected that the image of these particular emotions is not the same in the source and the target texts, not only because of the differences between the respective languages, but also because of a cultural difference concerning Polish and Swedish emotionality which has its influence the outcome of the translation process.

The phenomenon called *emotion* is usually defined as a post-cognitive phenomenon whose crucial aspect is the experiencer’s cognitive process leading to his/her own evaluation of the situation. Some researchers argue (Ortony, Clore, 1989: 127) that “to be an emotion, the feeling must signify the results of an appraisal of some kind. Thus, sadness is not simply a particular kind of feeling, but a particular kind of feeling for a particular reason”. Some linguists have questioned the idea that the element of appraisal is always present in the process of conceptualizing emotions. According to Wierzbicka, for example, one can

9 It is not easy to determine what is the phenomenon behind the English term emotion. The issue is complex and there is still no consensus about what emotions are like and how to describe them (cf. among others: Clore, Ortony, Foss 1987; Ekman 1992; Fries 1992; Johnson-Laird & Oatley 1989; Wierzbicka 1994, 1999a).

10 Dutch sociologist Geert Hofstede has shown (2001) that Polish and Swedish culture differ significantly from each other in terms of three dimensions: POWER DISTANCE, UNCERTAINTY AVOIDANCE and MASCULINITY.

say, *I am sad/happy today – I don't know why*, although certainly not *\*I am disappointed/disgusted today – I don't know why*. That is why she claims that for some concepts of emotions we do need a reference to a particular thought, whereas for others we do not – although we always need a reference to a prototypical scenario, which identifies, indirectly, the emotion in question (cf. Wierzbicka, 1992b: 291). As we will see, some emotions from the semantic field of 'fear' in Polish and Swedish have a particular motivation while others do not.

Emotions are usually divided by linguists and psychologists into two groups: basic and non-basic emotions (e.g. Ekman 1973, 1989, 1992; Fehr, Russel, 1984; Frijda 1986; Ortony, Clore, Collins, 1988; Plutchk, 1994; Russel, Bullock 1986; Wierzbicka, 1999). It was Paul Ekman and his co-workers who laid the foundations for the research in this field. On the basis of their detailed studies of physiological correlates of emotions they came to the conclusion that of all the emotions that people around the world feel, certain emotions have consistent correlates in facial expressions across cultures and these are so-called basic emotions:

The evidence now proves the existence of universal facial expressions. (...) Regardless of the language, of whether the culture is Western or Eastern, industrialized or preliterate, these facial expressions are labelled with the same emotion terms: happiness, sadness, anger, fear, disgust, and surprise (Ekman, 1973: 219-220).

The evaluation of all emotions (basic as well as non-basic) is carried out according to two main parameters which are considered primary, i.e. 'good'/'bad', and 'strong'/'weak'. All emotions can be defined by their positions in a two-dimensional space formed by these parameters. (cf. Fries, 1992; Gruszczyńska, 2001).

The pilot study focused on the nouns *strach* and *skräck* ('fear') and other nouns denoting related emotions<sup>11</sup> which belong to the same semantic field. We have chosen only these items from this field which have been found in our Swedish-Polish and Polish-Swedish parallel mini-corpus. In the Polish sub-corpus, these are: *strach* (*przestrach*), *przerażenie*, *trwoga*, *lęk*, *niepokój*, *obawa* *popłoch*/*panika* and in the Swedish subcorpus, they include: *skräck* (*förskräckelse*), *fasa*, *panik*, *oro*, *rädsla*, *förfäran*, *ångest*, *ängslan*.

The semantic field of 'strach' in Polish is very rich. It is represented by about 80 one-word lexical units and 400 analytical constructions (cf. Skorupka, 1974; Tomczak, 1997: 173; Gruszczyńska, 2001). Determining semantic similarities and differences between the nouns denoting this emotion (and consequently between verbs, adjectives etc.) is not a simple task. The definitions

11 According to Paul Ekman 'fear' stands not only for a single affective state but a family of related states (cf. Ekman 1992:172).



in *Słownik języka polskiego*, (*SJP*, Dictionary of Polish) reveal some similarities and differences in meaning between the items retrieved from the mini-corpus. These definitions, however, are not very useful for a precise differentiation between the analysed items because each one is defined in terms of the other units belonging the same semantic field<sup>12</sup>:

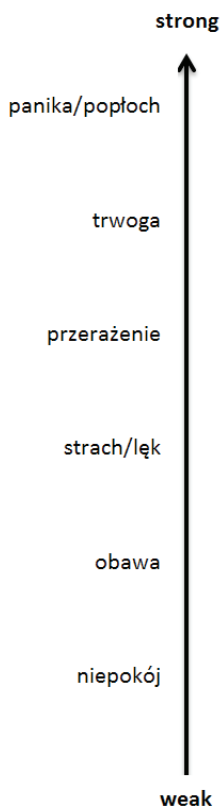
- strach* – “stan *niepokoju* wywołany przez niebezpieczeństwo lub rzecz nieznaną, która wydaje się groźna przez myśl o czymś groźącym” [a state of ‘≈ *lęk*’ evoked by a danger or something unknown that seems dangerous through thinking about a possible danger];
- lęk* – “uczucie *trwogi*, *obawy* przed czymś, *strach*”, psych. “stan emocjonalny pojawiający się jako reakcja na zagrożenie, którego źródło nie jest dokładnie znane i któremu człowiek nie może się aktywnie przeciwstawić” [a feeling of *trwoga*, *obawa* about something, *strach*, psych. an emotional state which is a response to a threat, whose source is not exactly known and which cannot be actively resisted];
- przerażenie* – “uczucie nagłego i silnego *lęku*, *przestachu*” [a feeling of a sudden and strong *lęk*, *przestach*];
- trwoga* – “stan, uczucie niepewności, *niepokoju* o to, co grozi” [a state, a feeling of uncertainty, *niepokój* of an imminent danger];
- niepokój* – “brak spokoju, równowagi” [a lack of calmness, balance];
- obawa* – “stan, uczucie niepewności, *niepokoju*, co do skutków, następstw czego” [a state, a feeling of uncertainty, *niepokój* about the results or consequences of something];
- popłoch* – “*strach* nagle ogarniający ludzi” [*strach* which suddenly overcomes people];
- panika* – “nagły, niepohamowany, często nieuzasadniony *strach*, *przerażenie*, *popłoch*, zamieszanie ogarniające zwykle większą liczbę ludzi” [a sudden, uncontrollable, frequently unjustified *strach*, *przerażenie*, *popłoch*, a confusion usually coming over a larger number of people]

The first of the defined nouns, i.e. *strach*, is one of two most frequent items among the selected words (in Polish texts in general<sup>13</sup> as well as in the analysed material; the other one is *niepokój*) and it is part of numerous phraseological

12 We quote definitions in our own translation.

13 Cf. *Słownik frekwencyjny polszczyzny współczesnej*.

constructions. It is also considered to be the core lexeme of the discussed semantic field (Tomczak, 1977: 182) and a point of reference for other items. The main differences between *strach* and the other words result from several additional semantic components defining these emotions. *Lęk* is defined in terms of *strach*, thus it seems to be equally strong but it is often connected with an unknown cause. *Przerażenie* differs from *strach* in incorporating the components [+sudden] and [+being very strong], and therefore also [+being very unpleasant] in its meaning. *Niepokój* undoubtedly belongs to the lexemes which denote weaker emotions than *strach* and therefore has the component [+weak], also, it is not necessarily evoked by a concrete cause. *Obawa*, similarly to *niepokój*, is also considered a weak emotion [+weak] but it differs from *niepokój* in always having a concrete cause. *Trwoga*, however, refers to a very strong emotion, much stronger than *strach*, which is demonstrated by all the examples in the *SJP* dictionary as well as all the citations from the mini-corpus. Therefore, the dictionary definition quoted above, which characterises this emotion as “a feeling of uncertainty” (similar to *obawa*) seems infelicitous as it omits the [+very strong] component. The last two items, i.e. *popłoch/panika*, which are equally strong, should be defined by the elements [+collective], [+mindless] and [+active], which is confirmed by the examples in the dictionaries.



In the above definitions the semantic component of being strong or weak is one of the main differentiating features. It may serve as a point of departure for an approximate ordering of the analysed lexical units according to the ‘strong’/‘weak’ parameter. The relations among them are illustrated in Figure 2, which is based on Fries’s diagram (cf. Fries, 1992)<sup>14</sup>. However, because the emotion ‘strach’ and its related feelings all belong to the group of unpleasant [+bad] emotions, only one axis is sufficient to illustrate the relations between them, as a stronger emotion is at the same time more unpleasant.

The semantic field of ‘skräck’ in Swedish is also very rich. It is represented by a similar number of one-word lexical units and analytical constructions as Polish ‘strach’ (cf. Gruszczyńska, 2001). For our analysis we have

The semantic field of ‘skräck’ in Swedish is also very rich. It is represented by a similar number of one-word lexical units and analytical constructions as Polish ‘strach’ (cf. Gruszczyńska, 2001). For our analysis we have

Figure 2. An approximate ordering of the lexical units from the semantic field of ‘strach’ according to the ‘strong’/‘weak’ parameter

<sup>14</sup> The differences in distance between the words depicted in Figure 2 are not proportional to differences in strength between them.

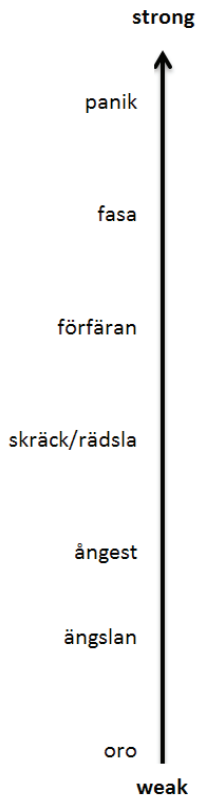
selected only several of them: *skräck*, *fruktan*, *fasa*, *panik*, *oro*, *rädsla*, *förfäran*, *skrämsel*, *ångest*, *ängslan*, i.e. only the nouns which occurred in the mini-corpus. The definitions provided by Swedish dictionaries (see References) reveal some semantic similarities and differences in meaning between the analysed items, however, in this case again, they are not very helpful in differentiating precisely between the individual emotions because each feeling is defined in terms of other emotions,<sup>15</sup> as it was the case in the Polish dictionary:

- skräck* – “mycket stark *rädsla* ofta i viss akut situation jfr *fasa*” [very strong *rädsla* often in an acute situation; cf. *fasa*]; in *SOB* it is also defined in terms of *rädsla* and *fasa* but two semantic components are emphasised: [+strong] and [+acute];
- fasa* – ”dels om mera bestående l. djupgående *ångest* l. *fruktan*, dels om mera tillfällig l. plötslig *förfäran* (*förskräckelse*)” (*SAOB*) [partly about more complex deep *ångest*, *fruktan*, partly about sudden *förfäran*];
- fruktan* – “1. *rädsla*, *skräck*, *bävan* 2. *ängslan*, *oro*, *farhåga* (att ngt obehagligt)” [*rädsla*, *skräck*, *bävan* 2. *ängslan*, *oro*, *farhåga* about something uncomfortable];
- panik* – “(plötslig o.) besinningslös *skräck* (som orsakar förvirring o. tumult o. hämningslösa försök att undkomma), *panisk förskräckelse* (som griper en samling människor)” [a sudden, foolish *skräck*: (which causes confusion and tumult, an unrestrained attempt to escape,) a *panic* seizing a group of people];
- oro* – “saknad av brist på ro, lugn, vila; tillstånd, förhållande som utmärkes av (tendenser, möjligheter till) störningar, förändringar, växlingar (i den normala tillvaron); särsk. om (tillstånd av) rörelse som stör ngts stillhet och vila; [lack of peace, tranquility, rest; a state characterized by (tendencies, possibilities of) disturbances, modifications, changes (in normal life); especially if (a state of) movement disturbs someone’s peace and rest];
- rädsla* – “förhållandet. egenskapen att vara *rädd* (för ngn l. ngt), *fruktan*; klenmod, försagdhet; *ängslan*, *bävan*; äv.: *förskräckelse*, *skräck*”; [ratio. property to be *rädd* (because of somebody or something) *fruktan*; timidity *ängslan*, *bävan*; or *förskräckelse*, *skräck*];
- förfäran* – “starkt, *skräckblandad obehag*” [a strong *fear* mixed with discomfort];

15 The definitions are quoted after *SAOB*, *SOB* and *LEXIN*.

- fasa* – “stark ihållande *förfäran*” [a strong, persistent *förfäran*];
- ångest* – “känsla av stark *oro* eller *fruktan*” [*Lexin*]; [a feeling of strong *oro* or *fruktan*];
- ängslan* – “obehaglig känsla att vara utsatt för fara” [*Lexin*]; [an uncomfortable feeling of being in danger]

The first of the analysed lexical units, i.e. *skräck*, is also one of the two most frequent items among the selected emotions (in Swedish texts in general, as well as in the mini-corpus; another one is *oro*) and is part of numerous phraseological constructions. It is also considered to be the main lexeme from the discussed semantic field and a point of reference for the other items. It is defined as “djupgående *ångest*” [profound *ångest*] and that is why it can be considered as stronger than *ångest*. According to the dictionary definition *rädsla* is very similar to *skräck*. *Oro* seems to be the weakest of all the analysed items. On the other hand, *panik*, *fasa* and *förfäran* are stronger than *skräck*. The first one is defined as very strong, and *fasa* is described as stronger than *förfäran*.



Similarly to Polish, in the above definitions the quality of being strong or weak is one of the main differentiating features. We have tried to order the analysed nouns according to the ‘strong’/‘weak’ parameter. Their place on the scale is depicted in Figure 3, which is also based on Fries’s (1992) diagram<sup>16</sup>.

The next step in our analysis involved examining how individual emotions from the semantic field of ‘fear’ were translated from Polish to Swedish and vice versa and how the translation equivalents in both languages were distributed along the strong/weak scale. We analysed 97 pairs of sentences retrieved from the Swedish-Polish and Polish-Swedish parallel mini-corpus containing the analysed words. Tables 2 and 3 present the examined nouns in the two languages together with their translations.

Figure 3. An approximate ordering of the lexical units from the semantic field of ‘skräck’ according to the ‘strong’/‘weak’ parameter

<sup>16</sup> The differences in distance between the words depicted in Figure 3 are not proportional to the differences in strength between them.

Swedish source texts	Polish translated texts	Number of hits
oro	niepokój	18
skräck	strach	9
skräck	przerażenie	4
ångest	lęk	8
ängslan	niepokój	7
panic	panika	6
0	[addition] przerażenie	5
förfäran	przerażenie	1
0	[addition] popłoch	1

Table 2. Polish equivalents of the Swedish nouns from the semantic field of 'skräck' in the mini-corpus

Polish source texts	Swedish target texts	Number of hits
niepokój	oro	12
strach	skräck	8
	rädsla	2
przerażenie	skräck	5
przerażenie	förfäran	3
popłoch	panik	3
trwoga	oro	2
lęk	oro	1
	ångest	2
	skräck	2

Table 3. Swedish equivalents of the Polish nouns from the semantic field of 'strach' in the mini-corpus

Figures 4 and 5 present the relative positions of the analysed nouns and their translations on the strong/weak scale in the two languages. If the hypothesis about the differences between Polish and Swedish emotionality is correct these differences should be reflected in discrepancies between the strength of the translation equivalents.

Both scales – the one presenting the ordering of the linguistic representations of emotions in Polish according to the 'strong'/'weak' parameter, and the other presenting the Swedish expressions ordered according to the same criterion – can be assumed to be comparable. In both of them the central position is occupied by one lexeme, and all the remaining nouns are situated closer or farther from the centre in the direction of stronger or weaker emotions (as stipulated by the prototype theory, Rosch, 1973). In Polish the central lexeme is *strach* and in Swedish it is *skräck*. The graphical positioning of the centres of both graphs on the same level makes it possible to juxtapose the scales and compare them with each other. However, it should be noted that the distances between individual expressions of emotions on the scales are symbolic and have

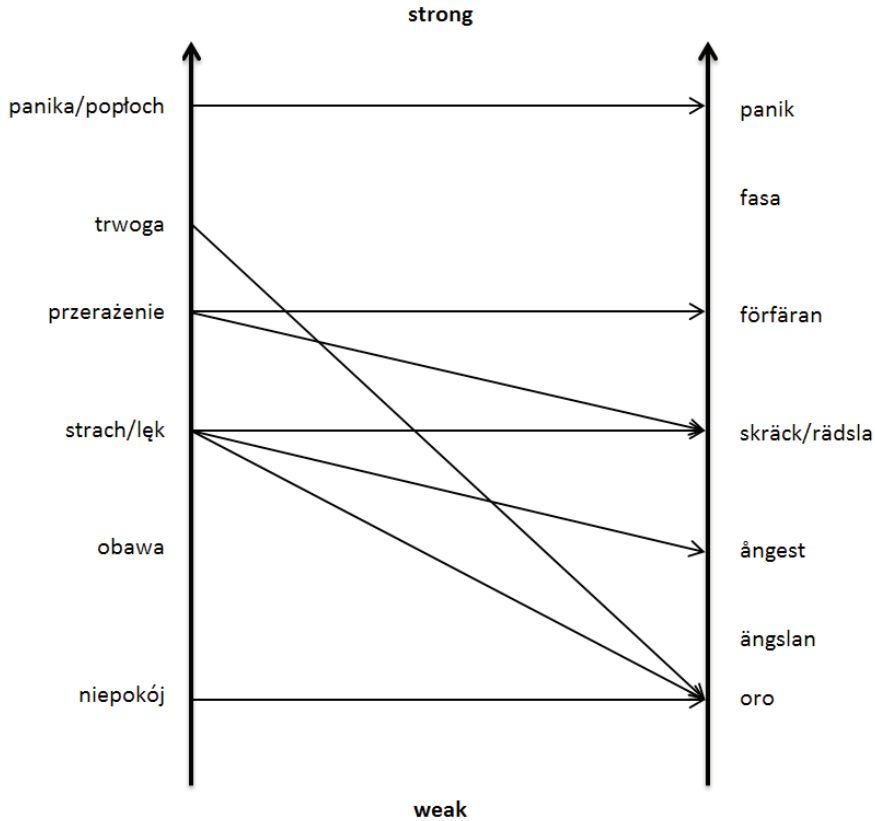


Figure 4. The Swedish equivalents of the Polish nouns from the semantic field of 'strach'

no influence on the overall picture of the observed tendencies. What is important is not the distances but the ordering, which was determined by the definitional properties of the individual items.

As Figure 4 demonstrates, Swedish translators often rendered Polish nouns denoting emotions related to fear with Swedish lexical items expressing weaker feelings. At the same time, Figure 5 indicates that Polish translators behaved in the opposite way: they preferred stronger Polish items or even additions as equivalents to Swedish nouns expressing the feelings from this semantic field. Figures 4 and 5 show that the translation equivalents of the items situated in the extreme positions on the strong/weak axis, i.e. Polish *niepokój* and *popłoch* and Swedish *oro* and *panic* are rendered by nouns in the other language which are identical (or only slightly different) in terms of their strength. On the other hand, the equivalents of the items situated in the middle of the scale are more varied and show a tendency to be weaker (in the case of Swedish translations) or stronger (in the case of Polish renderings). In other words, translators

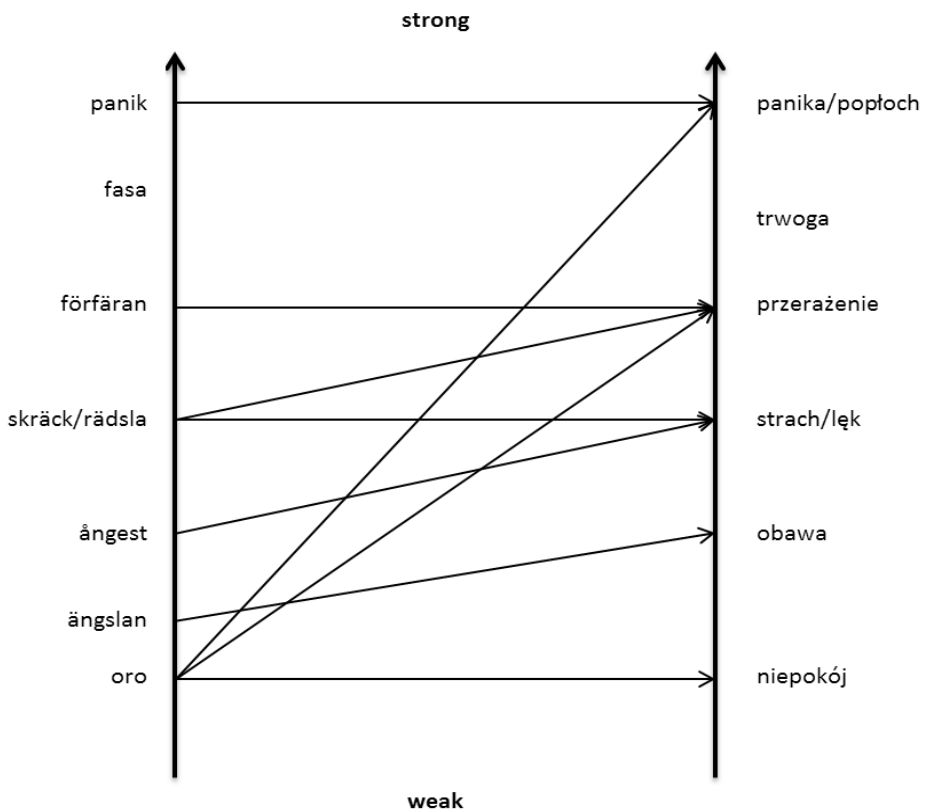


Figure 5. The Polish equivalents to the Swedish nouns from the semantic field of 'skräck'

tend to downgrade strong emotions when translating from Polish to Swedish, and upgrade them, conversely, when translating in the opposite direction. Such results suggest that the Swedish culture is characterised by a weaker/more subdued expression of emotionality in comparison with the Polish culture which, in turn, tends to express feelings by giving them a more intense undertone. This conclusion demonstrates that Hofstede's (2001) observation was not fully complete. The Polish and Swedish cultures differ significantly from each other not only in terms of three main dimensions: power distance, uncertainty avoidance and masculinity but also in terms of emotionality. However, as the size of our parallel mini-corpus is still very limited this conclusion should be treated with caution and must be confirmed in wider-scale research.

## 5. Conclusions

This article has introduced a new project on contemporary Polish-Swedish language contacts through translation which has recently been launched at the

University of Warsaw's Scandinavian Research Centre. An important part of this project is a compilation of a large, balanced and representative Swedish-Polish and Polish-Swedish parallel corpus of literary texts. The paper has described the mini-corpus which has been created in the pilot phase of the project. It has also presented the results of a small-scale study into translations of emotion terms related to 'fear' between the two languages, which was based on the data retrieved from the mini-corpus. The outcomes of the project's pilot phase have confirmed its feasibility. They have also proven that the planned Swedish-Polish and Polish-Swedish parallel corpus will be a valuable source of data for the kinds of analysis envisaged within the project.

## References

- BAKER, Mona (1995): Corpora in Translation Studies. An Overview and Suggestions for Future Research. *Target* 7(2), 223–243.
- BASSNETT-MC GUIRE, Susan (1980): *Translation Studies*. London: Methuen & Co. Ltd.
- BOWKER, Lynne (1998): Using specialized monolingual native-language corpora as a translation resource: A pilot study. *Meta* 43(4): 631–651.
- BRANTS, Thorsten. (2000): TnT – A Statistical Part-of-Speech Tagger. In: *Proceedings of the 6th Applied Natural Language Processing Conference, Seattle, Washington, USA*. <http://www.coli.uni-saarland.de/~thorsten/publications/Brants-ANLP00.pdf>, (17 October 2015).
- CLORE, Gerald, ORTONY, Andrew, FOSS, Mark A. (1987): The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology* 53, 751–766.
- EKMAN, Paul (1973): Cross Cultural studies of facial expressions. In: Paul EKMAN (ed.): *Darwin and Facial Expression: a Century of Research in Review*. New York: Annals of the New York Academy of Sciences, 169–229.
- EKMAN, Paul (1989): The argument and evidence about universals in facial expressions of emotions. In: Hugh WAGNER, Antony S.R. MANSTEAD (eds.): *Handbook of Social Psychophysiology*. Chichester: Wiley, 143–164.
- EKMAN, Paul (1992): An argument for basic emotions. *Cognition and Emotion* 6(3/4). Special Issue on Basic Emotions, 169–200.
- EVERT, Stefan, HARDIE, Andrew (2011): Twenty-first century Corpus Workbench: Updating a query architecture for the new millennium. In: *Proceedings of the Corpus Linguistics 2011 conference, University of Birmingham, UK*. <http://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2011/Paper-153.pdf>, (17 October 2015).



- FEHR, Beverley, RUSSEL Fehrand (1984): Concept of emotion viewed from a prototype perspective. *Journal of Personality and Social Psychology* 113, 464–486.
- FRIJDA, Nico H. (1986): *The Emotions*. Cambridge: Cambridge University Press.
- FRIES, Norbert (1992): Emocje. Aspekty eksperymentalne i lingwistyczne. In: Gabriel FALKENBERG, Norbert FRIES, Jadwiga PUZYNNINA (eds.): *Wartościowanie w języku i tekście*. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego, 105–135.
- GRUSZCZYŃSKA, Ewa (2001): *Linguistic Images of Emotions in Translation from Polish into Swedish. Henryk Sienkiewicz as a Case in Point*. Studia Slavica Upsaliensa 42. Uppsala: Acta Universitatis Upsaliensis.
- GRUSZCZYŃSKA, Ewa, SWOROWSKA, Anna (2015): Współczesna literatura szwedzka w polskim przekładzie. – [in:] Ewa GRUSZCZYŃSKA: *Spotkania językowe szwedzko-polskie za pośrednictwem przekładu*. Warszawa: Oficyna Wydawnicza ASPRA-JR, 31–75.
- HALACSY, Peter, KORNAI, Andras, ORAVECZ, Csaba. (2007): Hunpos – an open source trigram tagger. In: *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Prague, Czech Republic. Companion Volume: Proceedings of the Demo and Poster Sessions*. Association for Computational Linguistics, 209–212. <http://www.kornai.com/Papers/acl07poster.pdf>, (17 October 2015).
- HOFSTEDE, Geert (2001) *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. Second Edition. Thousand Oaks, CA: Sage Publications.
- CHOJNACKI, Hieronim (2003): *Szwedzka literatura piękna w Polsce 1939-1996*. Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego.
- JOHNSON-LAIRD Philip, OATLEY Keith (1989): The language of emotions: an analysis of semantic field. *Cognition and Emotion* 3, 81–123.
- KENNY, Dorothy (1998): Corpora in Translation Studies. *Routledge Encyclopedia of Translation Studies*, 50–53.
- MEGYESI, Beata (2001): Comparing Data-driven Learning Algorithms for PoS Tagging of Swedish. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2001), Carnegie Mellon University, Pittsburgh, PA, USA*, 151–158. <https://www.aclweb.org/anthology/W/W01/W01-0519.pdf>, (17 October 2015).
- MEGYESI, Beata (2008): The Open Source Tagger HunPoS for Swedish. Report, September. Department of Linguistics and Philology, Uppsala University. <http://stp.lingfil.uu.se/~bea/publ/megyesi-hunpos.pdf>, (17 October 2015).

- MEYER, Roland, VON WALDENFELS, Ruprecht, WOŹNIAK, Michał, ZEMAN, Andreas (2006-2015): *PARAVoz – a simple web interface for querying parallel corpora*. Second Version. Bern, Regensburg, Berlin, Krakow. <https://bitbucket.org/rvwfels/paravoz>, (17 October 2015).
- ORTONY Andrew, CLORE, Gerald (1989): Emotions, moods, and conscious awareness. *Cognition and Emotion* 3(2), 125–137.
- ORTONY, Andrew, CLORE, Gerald, COLLINS, Allan (1988): *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press.
- ÖSTLING, Robert (2013): Stagger: an Open-Source Part of Speech Tagger for Swedish. *Northern European Journal of Language Technology*, 3, 1–18.
- PEARSON, Jennifer (2003): Using parallel texts in the translator training environment. In: Federico ZANETTIN, Silvia BERNARDINI, Dominic STEWART (eds.): *Corpora in Translator Education*, Manchester: St Jerome, 15–24.
- PLUTCHIK, Robert (1994): *The Psychology and Biology of Emotions*. New York: Harper Collins College Publishers.
- ROSCHE, Eleanor (1973): *Natural categories*. – *Cognitive Psychology*, 4 (3), 328–50.
- RUSSEL James A., BULLOCK Marry (1986): Fuzzy concepts and the perception of emotion in facial expressions. *Social Cognition* 4, 309–341.
- SCHMID, Helmut (1995): Improvements in Part-of-Speech Tagging with an Application to German. In: *Proceedings of the ACL SIGDAT Workshop. Dublin, Ireland*. <http://www.cis.uni-muenchen.de/~schmid/tools/Tree-Tagger/data/tree-tagger2.pdf>, (17 October 2015).
- STEINBERGER Ralf, POULIQUEN, Bruno, WIDIGER, Anna, IGNAT, Camelia, ERJAVEC, Tomaž, TUFİŞ, Dan, VARGA Dániel (2006): The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages. In: *Proceedings of the 5<sup>th</sup> International Conference on Language Resources and Evaluation (LREC'2006), Genoa, Italy, 24-26 May 2006*, 2142–2147. <http://www.lrec-conf.org/proceedings/lrec2006/>, (1 March 2016).
- TIEDEMANN, Jörg (2012): Parallel data, tools and interfaces in OPUS. In: Nicoletta CALZOLARI, Khalid CHOUKRI, Thierry DECLERCK, Mehmet Uğur DOĞAN, Bente MAEGAARD, Joseph MARIANI, Asuncion MORENO, Jan ODIJK, Stelios PIPERIDIS (eds.): *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012)*. Istanbul: European Language Resources Association (ELRA), 2214–2218.
- TOMCZAK, Katarzyna (1997): Wyrażenia z leksemami “strach” i “bać się” we współczesnej polszczyźnie. In: Renata GRZEGORCZYKOWA, Zofia ZARON (eds.): *Semantyczna struktura słownictwa i wypowiedzi*. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego, 174–198.

- VON WALDENFELS, Ruprecht (2011): Recent developments in ParaSol: Breadth for depth and XSLT-based web concordancing with CWB. In: Daniela MAJCHRÁKOVÁ, Radovan GARABÍK (eds.): *Natural Language Processing, Multilinguality. Proceedings of Slovko 2011, Modra, Slovakia, 20–21 October 2011*. Bratislava: Tribun EU, 156–162.
- WIERZBICKA, Anna (1990): The semantics of emotions: fear and its relatives. *Australian Journal of Linguistics*, 10 (2), 133–138.
- WIERZBICKA, Anna (1991): *Cross-cultural Pragmatics. The semantics of Human Interaction*. Berlin/New York: Mouton de Gruyter.
- WIERZBICKA, Anna (1992a): *Semantics, Culture and Cognition. Universal Human Concepts in Culture-specific Configurations*. Oxford/New York: Oxford University Press.
- WIERZBICKA, Anna (1992b): Talking about emotions: Semantics culture and Cognition. *Cognition and Emotion* 6 (3/4), 285–319.
- WIERZBICKA, Anna (1994): Emotion, language and “cultural scripts”. In: Shinobu KITAYAMA, Hazel Rose MARKUS (eds.): *Emotion and Culture: Empirical studies of mutual influence*. Washington, DC: American Psychological Association, 130–198.
- WIERZBICKA, Anna (1999a): *Emotions across Languages and Cultures: Diversity and Universals*. Cambridge: Cambridge University Press.
- WIERZBICKA, Anna (1999b): *Język, umysł, kultura*. Warszawa: Wydawnictwo Naukowe PWN.
- ZANETTIN, Federico (1998): Bilingual comparable corpora and the training of translators. *Meta* 43(4), 616–630.

## Dictionaries

- SKORUPKA, Stanisław (1974): *Słownik frazeologiczny języka polskiego*. Warszawa: Wiedza Powszechna.
- SOB *Svensk Ordbok*, (1990): Esselte Ordbok.
- SAOB Svenska Akademiens Ordbok (Internet version) (1997). Lund Göteborg <http://g3.spraakdata.gu.se/saob/>.
- KURCZ, Ida et al. (1990): *Słownik frekwencyjny polszczyzny współczesnej*. Kraków/Warszawa: Instytut Języka Polskiego, Polska Akademia Nauk.
- SZYMCZAK, Mieczysław (ed.) (1992): *Słownik języka polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- LEXIN: <http://lexin.nada.kth.se/lexin/#>

