A Corpus Based Investigation of Collocations in Hungarian

Júlia PAJZS
Research Institute for Linguistics, Hungarian Academy of Sciences
H-1068 Budapest Benczúr u. 33 Hungary
pajzs@nytud.hu

Abstract
The paper presents a corpus based pilot study of Hungarian collocations. The toolset to acquire the frequent word combinations is described briefly along with the analysed Historical Corpus of Hungarian. The collocations of the words ad 'give' – kap 'get' – kér 'ask for' – vesz 'buy' will be examined in detail, to model the study of semantic relationships through the analysis of collocates.

1 Definitions of Collocations
The past decades have witnessed the recognition of the significance of investigating corpus-based collocations. Since Sinclair's essential book [1991] many attempts have been devoted to the automatic collocation extraction from corpora. While Sinclair defined collocations simply as frequently co-occurring word sequences, most authors make an effort to define them more precisely, partly by a classification of phrase types. These are often categorized in the following major classes: "FREE COMBINATIONS are: Not preconstructed, and semantically literal (i.e. the words have retained their conventional literal meanings) COLLOCATIONS are: Preconstructed, and semantically literal (i.e. the words have retained their conventional literal meanings) IDIOMS are: Preconstructed, not semantically literal (i.e. the words, or at least one of them, have not retained their conventional literal meanings, or at least cannot be analysed as such)." [Van Der Meer, 1998:314]. Although collocation has been a hot topic in corpus linguistics for the past decade, there have hardly been any attempts to investigate them in Hungarian corpora so far. There are some undergoing projects with more or less similar aims: a dictionary of Hungarian collocations is under compilation, using the Historical corpus among its resources; two French-Hungarian dictionaries are also under compilation, one specialising in phrases while the other one in idioms. Still, there has not been an organised project for joining forces. As the revised plan for compiling the Historical Dictionary of Hungarian is aiming to focus more on collocations, there is a growing need for research in this field.

2 Collocation Investigation in the Hungarian Corpus

2.1 The Historical Corpus of Hungarian
For the project of the Historical Dictionary of Hungarian a representative corpus was collected (25 million running words). Most of the texts are different kinds of prosaic texts (prosaic fiction: 31%, other kinds of prose: 51%, poetry: 8,5%, drama: 5,7%), chosen from three centuries (1772-2000). A morphological analyser programme was successfully run on
the modern texts, but the analysis of the earlier texts was problematic. We have designed and applied a special programme to improve the recognition of historical texts. The method to convert and analyse archaic texts without losing the original word forms was described in Kiss & Pajzs [2001] and Kiss et al. [2001]. The result is an analysed corpus containing the original running word in the first field and the recognised, or presumed, analysis of the word in another field. If a running word has more than one analysis, each one is kept in the analysed field, separated by vertical bars. From this file the lemmas, the codes given by the analyser and the corresponding running words can be retrieved. For the compilers of the dictionary a retrieval interface was made to support the search in the concordances of the words in the order of their date of writing. (Currently two different text retrieval tools are being used for the same corpus: the search engine of the Open Text program is still in use, and we are testing the Corpus WorkBench corpus retrieval software, developed at the University of Stuttgart. Both are available through user-friendly interfaces. (The English version interface is available at the web site: www.nytud.hu/hhce.) While this retrieval mode is sufficient for words with a relatively low frequency (below 100 occurrences), it does not support the recognition of recurrent word phrases, and it is hardly usable in the case of very frequent words (over 1000 occurrences).

2.2 Current Research Purpose

The aims of this research are manifold, including:

- the development of a specialised software toolset to facilitate the recognition of recurrent word phrases,
- the investigation of different kinds of frequent word phrases in Hungarian, based on corpus research and other resources (dictionaries, traditional archives, etc.),
- the preparation of a Hungarian phrase database, retrievable by any of its key elements and properly classified according to different aspects: whether it is just a frequent free combination, or a semi-preconstructed and partially literal collocation, or a completely preconstructed and semantically non-literal idiom.

The database should also contain the definitions of each phrase, and additional fields where the compiler provides his/her suggestion whether the entry should be included in monolingual or bilingual dictionaries, also indicating the type (language, size etc.).

2.3 The Collocations and Other Frequent Word Phrases in the Historical Dictionary of Hungarian

The project for the Historical Dictionary of Hungarian has been going on for several years now. At the end of 2000 a committee, set up to revise the original concept of the dictionary, decided to redesign the whole project to make it more realistic. The updated concept outlines the compilation of an eight-volume dictionary, to cover the vocabulary of Hungarian from 1772 up to 2000. The number of main entries will be around 100,000 (the largest available dictionary on Hungarian contains 52,000 entries). Each defined sense will be illustrated by at least one citation from the corpus, supplied with exact bibliographic reference. The first occurrence of each sense in the given period will be illustrated either with a citation or with the date of its writing and the bibliographic reference. In the electronic version more citations are to be included, with some additional possibilities in this version as well, e.g. direct retrieval of synonyms, hyponyms, the availability of the full bibliographic database, frequency information etc.
One of the major novelties of this work compared to the earlier Hungarian monolingual dictionaries is the special treatment of collocations and frequent word phrases. While idioms and proverbs will be excluded, this will be the very first Hungarian dictionary to handle collocations in a systematic way: a special section of the entries will contain collocations that are, at least partially, preconstructed, their meaning being not completely transparent from the meaning of the elements. The arrangement of collocations is strictly defined:

- the ADJECTIVE+NOUN phrases are included in the entry of the adjectival element,
- NOUN+VERB phrases are included in the entry of the nominal element. If there are more than one noun in the phrase it is either included in each place, or a cross reference is given,
- PRONOUN+VERB phrases are included in the entry of the verbal element,
- NOUN+NOUN or NOUN+PRONOUN phrases are included in the entry of the nominal element [Ittzés 2002].

Apart from collocations it is possible and advised to give frequently occurring free word combinations in the citation section as additional examples, supplied with bibliographic reference.

3. Collection of collocations from the corpus

3.1 Tools for Extracting the Frequent Word Phrases

In the first step the full concordance of the keyword is retrieved by the Open Text programme (see Figure 1 for an example). This format is unreadable for human users, because it contains both the running words and their analysed versions and the necessary XML tags between them. In Open Text the length of the concordance can only be specified by the number of characters to be outputted to the right and left of the keyword.

```
1772..mert[MN]mer{IGE}[Me3]<a /></w>
<w><t>akkor</t><a>az[NM][TEM]</a></w>
<w><t>leg</t/><a><t>leg[NE]</t></a></w>
<w><t>ko2mo2sbben</t>/<a>NIN</a></w>
<w><o>faggattya</o><t>faggat[IGE][Te3]</t></w>
<w><t>o2tet</t>/<a>o2tet[NE]</a></w>
<w><t>Elejbe</t><a>elejbe[HA]</a></w>
<w><o>adgya</o><t>ad[IGE][TPe3][ad][IGE][Te3]</t></w>
<b>Figure 1: An analysed concordance line after Open Text retrieval</b>
```

A PERL programme then converts the concordance to a more transparent format for further processing. The output contains three words before and after the keyword, each of the fields are kept and tabs, as well as the XML tags, placed between the different fields of the original lemmatised file, separate them. This converted file can then be imported to any database or spreadsheet programme to be further processed with that. The XML tags are kept to produce
an output in XML format when required. The result is imported into ACCESS. From the table
several different kind of retrievals can be generated. The table can be easily sorted by any of
the fields or their combinations (e.g., the first preceding or following word/suffix to the
keyword). In each case it is possible to display the running words only, while sorts and
selections can be made on the hidden fields as well.

<table>
<thead>
<tr>
<th>Wdate</th>
<th>tok-3</th>
<th>tok-2</th>
<th>tok-1</th>
<th>tok</th>
<th>tok+1</th>
<th>tok+2</th>
<th>tok+3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1936</td>
<td>pamlagot vagy ágyat adtak telőm most nem</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1935</td>
<td>imádságos magyar áhitattal adhatunk Áldja meg az</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1882</td>
<td>szemeit nagy áttatosan Adj el azonnal száz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1923</td>
<td>Vilharodnak palástját áldva add rám S villámodat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1900</td>
<td>Kar Adjon áldást adjon áldást Szeressétek holtig</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1869</td>
<td>kogyelet adóját áldást adva a szent hantokra</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1846</td>
<td>hogy az áldást adják gyermekőkre És ha</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1857</td>
<td>Végre befejezésül áldást ad a bémáltakra mondvan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1880</td>
<td>megint csak áldásunkat adjuk a frigyhez Ha</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1866</td>
<td>Az ég áldását adja rád Bármit keres</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1989</td>
<td>tolmácsoljuk és áldásunkat adjuk szervezetüknek és az</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: A sample from the concordance of the word ad ‘give’ sorted on the preceding
lemma.

While trying to study the most frequent collocations in the tables described above, I realised
that many frequent collocates can occur in either of the right or the left hand side of the
keyword, and in many cases one or two words can be inserted in between them (articles,
adjectives, modifiers, conjunctions etc.). Therefore, I collected all the words occurring in the
three-word proximity of the keyword into an array, and made a frequency list of them sorted
by their frequency. When trying to collect the collocation candidates, we can first examine
this sorted frequency list and select the first guesses for the possible collocations. The actual
collocation list can be compiled using the sorted tables exemplified in Figure 2.

When one further combines the ordering possibilities, the typical syntactic constructions
along with the most frequent collocates and their semantic content will become transparent
simultaneously. Figure 3 exhibits several typical examples of the phrase kifejezést ad vminek
‘express sg’, arrived at when sorting the concordance by the suffix of the first word to the
right, then the suffix of the first word to the left of the keyword.
### Figure 3: Concordance sorted by the suffix of the following and the preceding word

<table>
<thead>
<tr>
<th>Date</th>
<th>Word (Suffix)</th>
<th>Collocate</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989</td>
<td>elkeseredésüknek</td>
<td>Megkönnyödésüknek</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1978</td>
<td>s</td>
<td>Reményüknek</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1983</td>
<td>iránti</td>
<td>Aggodalmuknak</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1982</td>
<td>azon</td>
<td>Véleményüknek</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1980</td>
<td>kivánságuknak</td>
<td>is</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1985</td>
<td>azon</td>
<td>Hítemnek</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1914</td>
<td>a</td>
<td>Reményének</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1872</td>
<td>azon</td>
<td>Meggyőződésnek</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1895</td>
<td>sokszor</td>
<td>Sokképpen</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1872</td>
<td>azon</td>
<td>Ohajtásnak</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1900</td>
<td>szép</td>
<td>Türelmességnek</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1905</td>
<td>hogy</td>
<td>kézzorítással</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1883</td>
<td>újév</td>
<td>napján</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1891</td>
<td>hogy</td>
<td>tiszteletemnek</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1882</td>
<td>ténynek</td>
<td>akart</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1887</td>
<td>nézetüknek</td>
<td>korántsem</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1881</td>
<td>kézpelődésünk</td>
<td>regékben</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1889</td>
<td>azon</td>
<td>hozzáadással</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1889</td>
<td>ama</td>
<td>reményüknek</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1898</td>
<td>iránti</td>
<td>méltánylátszának</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1880</td>
<td>azon</td>
<td>nézetének</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1880</td>
<td>azon</td>
<td>reményének</td>
<td>Kifejezést</td>
</tr>
<tr>
<td>1883</td>
<td>mély</td>
<td>fájdalmának</td>
<td>Kifejezést</td>
</tr>
</tbody>
</table>

### 3.2 Investigation of the Collocations of Four Semantically Related Words: *ad* 'Give' - *kap* ' – kér 'Ask for' – *vesz* 'Buy'.

Each of the four words were examined focusing on their collocates (see Figure 4). Among their several different meanings, just like in English, some are in antonym relation to one or more meanings of the other words. As all of these words are much too frequent for a traditional entry compilation (their absolute frequencies: *ad* 25,221, *kap* 9,491, *kér* 12,482, *vesz* 23,919, they are all within the range of the top most frequent 300 words), it would be a hopeless task to read every citation simply in the order of their date of writing, and then decide on the semantic groups to be identified and exemplified. I have tried to collect the most recurrent collocations for each of the words, and then compared the results. The first questions included: which collocate occurs with each of those words (except for the articles, conjunctions etc., of course), and among them, in which cases they produce collocations with the corresponding meaning. For example: *kéz* 'hand' is the topmost frequent collocate of *ad* 'give', partially because there are several collocations and idiomatic expressions where both *ad* and *kéz* occurs: *kezet ad* 'shake hands with sb'; *kezet adja* 'woman agrees to marry'; *kézre ad* 'give up sb to the police or jury'; *kézről kézre ad* 'pass sg from hand to hand', *sorsát kezedbe adja* 'put his/her fate into your hands' *szabad kezet ad* 'give sb full power'. Let me emphasize that none of this phrases has a word-for-word translation in English, so their inclusion is crucial in bilingual dictionaries. Among the existing Hungarian-
English/English-Hungarian dictionaries the still very best unabridged bilingual dictionary [Országh et al., 1998], unfortunately, hardly supports the search for these collocations, since the entries *ad* and *kéz* are, of course, rather long and complicated. Nearly each of these collocations are included, some of them in both entries (in these cases with different equivalents). Traditional Hungarian lexicography follows the routine of sorting the phrases by the order of the suffix of the collocate. Under each sense, the first ones are always the nominal collocates, then follows everything in the accusative, then the dative, etc. Inside this grammatical ordering there is also an alphabetic ordering. Although this thoroughly methodological way of presenting the collocates can be especially useful for the expert user, a more novice user is not likely to realise the logic of the ordering at all without a special instruction for use (and the lexicographers are all aware that everyday users never read the preface of any dictionary). One of the collocations, *sorsát kezébe adja* was absent from the entries. This experience further emphasizes the necessity of a complete and well-organized collocation database in Hungarian, which could then serve as the basis for the new or thoroughly revised dictionaries.

The word *kéz* 'hand' is among the most frequent collocates of each of the other three words *kap-kér-vesz* (in the range of the first sixty), but in none of the collocations can each word be replaced with the other one with a corresponding meaning. The results are presented in Figure 4. For example: the phrase *kezét adja* 'she agrees to marry' has a corresponding collocate with the opposite meaning *kéri a kezét* 'asks her to marry', but this is more often used by an additional verbal prefix *meg-* (*megkéri a kezét*), while *kezét adja* tends to be prefixed by a different verbal prefix *oda-* (*odaadja a kezét*). The word *kap* 'get' is also supplied with a verbal prefix in the corresponding collocation *meg-* (*megkapja a kezét* 'get her and her parents' consent'). With the word *vesz* a completely different expression is used: *elveszi*, or *elveszi feleségül 'marry her*. The latter is one of the rare cases, when all of the four alternative collocations exist with a corresponding meaning: *feleségül kér* 'asks her to marry him'; *feleségül ad* 'gives his/her consent to his/her daughter's marriage'; *feleségül kap* 'may marry her'; *feleségül vesz* 'marry her'. The other collocation in which both *kéz* and *ad* occurs is *kézbe ad* 'give sg to his/her hand'; there exists *kézbe kap*, 'gets sg into his/her hand' *kézbe vesz* 'takes it into his/her hand'; but there is no *kézbe kér*. The most frequent actor of the verb *ad* is *Isten* 'God' (or its synonyms), while the most frequent recipient of the word *kap* is *ember* 'man'. Naturally enough, all concrete objects can be a collocate of each of the examined verbs. In the case of *kap* the topmost frequent collocate is *levelet* 'letter' (the 12th on the sorted list, and the very first content word). In the case of *ad* 'give' the sense of *ad* would mainly fall into the one defined as first sense in LDOCE3: 'PROVIDE/SUPPLY', while *kap* means 'RECEIVE', *kér* means 'ASK FOR HELP ETC', *vesz* means 'TAKE OR BUY'. Among their collocates the topmost frequent one is *pénz* 'money', which occurs with the highest rank next to *ad* and with the lowest next to *vesz*. 
In Figure 5, the comparison of the relative frequency of the frequent collocates of the antonym verbs *ad* — *kap* is summarised. When comparing the relative frequency of the collocate of *ad* (0.0146) — *kap* (0.0158), we may realise that money is about as often given as it is actually received, which is not the case for most of the collocates of any of these word pairs. There are many things that you can give and receive theoretically, but the corpus suggests that you usually rather receive them (for example *ősztöndíj* 'scholarship' is only received and nobody gives them, you get *segéfy* 'help', *munka* 'work', *pofon* 'slap', and plenty of other things that are rarely given by somebody else). On the other hand *példa* 'good example' is much more often given than received. In the figure bold italicised fonts mean which element of the pair has a higher relative frequency.

### Figure 4: Frequent collocations of *kér-kap-ad-vesz*

<table>
<thead>
<tr>
<th>KÉR</th>
<th>KAP</th>
<th>AD</th>
<th>VESZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>kéri a kezét</td>
<td>megkapja a kezét</td>
<td>(oda)adj a kezét</td>
<td>kézbe vesz</td>
</tr>
<tr>
<td>&quot;ask her to marry sb&quot;</td>
<td>'get her and her parents consent'</td>
<td>'she agrees to marry'</td>
<td></td>
</tr>
<tr>
<td>kézbe kap</td>
<td>kézbe ad</td>
<td>(oda)adj a kezét</td>
<td>kézbe vesz</td>
</tr>
<tr>
<td>'get sg into his hand'</td>
<td>'give sg to sb's hand'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kézbe ad</td>
<td>kézbe vesz</td>
<td>'take sg into sb's hand'</td>
<td></td>
</tr>
<tr>
<td>ismét adj</td>
<td>Istentől kap</td>
<td>Istentől vette</td>
<td>'taken from God'</td>
</tr>
<tr>
<td>'God gives'</td>
<td>Istentől kap</td>
<td>Istentől vette</td>
<td></td>
</tr>
<tr>
<td>határozott</td>
<td>Istentől kap</td>
<td>Istentől vette</td>
<td></td>
</tr>
<tr>
<td>határozottat</td>
<td>Istentől kap</td>
<td>Istentől vette</td>
<td></td>
</tr>
<tr>
<td>&quot;ask God&quot;</td>
<td>Istentől kap</td>
<td>Istentől vette</td>
<td></td>
</tr>
<tr>
<td>&quot;ask God&quot;</td>
<td>Istentől kap</td>
<td>Istentől vette</td>
<td></td>
</tr>
<tr>
<td>Pénzt kér</td>
<td>pénzt kap</td>
<td>pénzt ad</td>
<td>pénzt ad</td>
</tr>
<tr>
<td>'ask for money'</td>
<td>'receive money'</td>
<td>'give money'</td>
<td>'buy sg for money'</td>
</tr>
<tr>
<td>Felségét kér</td>
<td>felségét kap</td>
<td>felségét ad</td>
<td>felségét vesz</td>
</tr>
<tr>
<td>'ask her to marry sb'</td>
<td>'may marry her'</td>
<td>'give his consent to his daughter's marriage'</td>
<td></td>
</tr>
<tr>
<td>Felségét kér</td>
<td>felségét kap</td>
<td>felségét ad</td>
<td>felségét vesz</td>
</tr>
<tr>
<td>'ask her to marry sb'</td>
<td>'may marry her'</td>
<td>'give his consent to his daughter's marriage'</td>
<td></td>
</tr>
<tr>
<td>Erőt kap</td>
<td>erőt ad</td>
<td>erőt ad</td>
<td>erőt vesz (magán)</td>
</tr>
<tr>
<td>'regain strength'</td>
<td>'strengthen, fortify'</td>
<td>'keep one's temper'</td>
<td></td>
</tr>
<tr>
<td>Szót kér</td>
<td>szót kap</td>
<td>szót ad</td>
<td>szót ad</td>
</tr>
<tr>
<td>'demand the floor'</td>
<td>'get the floor, is allowed to speak'</td>
<td>'allow to speak'</td>
<td></td>
</tr>
<tr>
<td>Szót kér</td>
<td>szót kap</td>
<td>szót ad</td>
<td>szót ad</td>
</tr>
<tr>
<td>'demand the floor'</td>
<td>'get the floor, is allowed to speak'</td>
<td>'allow to speak'</td>
<td></td>
</tr>
<tr>
<td>Sámon kér</td>
<td>sámon ad</td>
<td>sámon ad</td>
<td>sámon ad</td>
</tr>
<tr>
<td>'call sb to account for'</td>
<td>'give account of sg'</td>
<td>'take sg into account'</td>
<td></td>
</tr>
<tr>
<td>Kolcson kér</td>
<td>kolcson kap</td>
<td>kolcson ad</td>
<td>kolcson vesz</td>
</tr>
<tr>
<td>'ask for a loan/borrow'</td>
<td>'sg is lent to sb'</td>
<td>'loan'</td>
<td>'borrow sg from sb'</td>
</tr>
<tr>
<td>Kegyelmet kér</td>
<td>kegyelmet kap</td>
<td>kegyelmet ad</td>
<td>kegyelmet ad</td>
</tr>
<tr>
<td>'apply for mercy/pardon'</td>
<td>'be pardoned'</td>
<td>'give pardon'</td>
<td></td>
</tr>
<tr>
<td>Kihallgatást kér</td>
<td>kihallgatást kap</td>
<td>kihallgatást ad</td>
<td>'give sb an audience'</td>
</tr>
<tr>
<td>'received by sb in audience'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Felmentést kér</td>
<td>felmentést kap</td>
<td>felmentést ad</td>
<td>felmentést ad</td>
</tr>
<tr>
<td>'asked to be exempted'</td>
<td>'be granted an exemption'</td>
<td>'grant an exemption'</td>
<td></td>
</tr>
<tr>
<td>Segélyt kér</td>
<td>segély kap</td>
<td>segélyt ad</td>
<td>segélyt ad</td>
</tr>
<tr>
<td>'ask for aid'</td>
<td>'get aid'</td>
<td>'give aid'</td>
<td>'give aid'</td>
</tr>
<tr>
<td>Ségéseg kér</td>
<td>ségéseg kap</td>
<td>ségéseg ad</td>
<td>ségéseg ad</td>
</tr>
<tr>
<td>'ask for help'</td>
<td>'get help'</td>
<td>'give help'</td>
<td></td>
</tr>
<tr>
<td>Tanácsot kér</td>
<td>tanácsot kap</td>
<td>tanácsot ad</td>
<td>tanácsot ad</td>
</tr>
<tr>
<td>'ask for advice'</td>
<td>'receive advice'</td>
<td>'give advice'</td>
<td></td>
</tr>
<tr>
<td>Találkozót kér</td>
<td>találkozót kap</td>
<td>találkozót ad</td>
<td>találkozót ad</td>
</tr>
<tr>
<td>'ask for an appointment'</td>
<td></td>
<td></td>
<td>'make an appointment'</td>
</tr>
<tr>
<td>Szakvéleményt kér</td>
<td>szakvéleményt kap</td>
<td>szakvéleményt ad</td>
<td>'give an expert appraisal'</td>
</tr>
<tr>
<td>'ask for professional advice'</td>
<td>'take professional advice'</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Figure 5. the comparison of the relative frequency of the frequent collocates of the antonym verbs *ad* — *kap* is summarised. When comparing the relative frequency of the collocate of *ad* (0,0146) — *kap* (0,0158), we may realise that money is about as often given as it is actually received, which is not the case for most of the collocates of any of these word pairs. There are many things that you can give and receive theoretically, but the corpus suggests that you usually rather receive them (for example *ősztöndíj* 'scholarship' is only received and nobody gives them, you get *segéfy* 'help', *munka* 'work', *pofon* 'slap', and plenty of other things that are rarely given by somebody else). On the other hand *példa* 'good example' is much more often given than received. In the figure bold italicised fonts mean which element of the pair has a higher relative frequency.
The comparison of the frequencies of semantically related, commonly occurring words' collocates is an interesting topic to study in itself. One of the envisaged tasks is to test distribution differences among these collocates.

The szó 'word' is another collocate that occurs frequently next to each of the keywords. The collocation szavát adja 'give one's word for it' has the corresponding versions kéri a szavát 'asks for his/her promise'; (meg)kapja a szavát; 'gets his/her promise' szavát veszi 'take his/her word'. Among these (meg)kapja a szavát hardly ever occurs in the corpus, and szavát veszi has a completely different meaning as well: 'make sb speak'. Another collocation containing both szó and ad is szót ad 'allow to speak', which has corresponding collocations with kap and kér. There are several collocates which are shared by these three words ad-kap-kér. Many of them belong to the category described in LDOCE3 as: 'TELL •• •••/PROVIDE INFORMATION or DO STH'. Similarly to English, advice, information, help, explanation, etc. can be given, received, or taken.

**Conclusion**

When examining semantically related words (synonyms, antonyms, hyponyms) studying their common and uncommon collocates can help the identification of the different semantic features of each word.

The aim of this pilot study was to draw the attention to the importance of the collection and investigation of collocations in Hungarian. Only a handful directions have been touched upon, further research will include:

- a further improvement of the collocation collecting program. More sophisticated statistical methods are to be tested and built in later on. The final version should have a user friendly interface, which should be a module of the corpus retrieval program being used.
- the preparation of a Hungarian phrase database containing not only the collocations but the frequent word combinations, idioms, proverbs as well, together with a specific encoding and supplied with a flexible retrieval interface. The database is to be continuously updated both with data coming from the corpus of current Hungarian, and with multilingual equivalents as well.
Acknowledgements

The project for the Historical Dictionary of Hungarian is supported by the following grants:

<table>
<thead>
<tr>
<th>Grant</th>
<th>Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Science Foundation</td>
<td>NKA, 058438-02/1997; 1998; OKTK; A.1619/VIII.a; 1999; OKTK; A.1619/VIII.a; 2000; OTKA; T30297; 1999–2002;</td>
</tr>
</tbody>
</table>

In the editing of the English version of this paper, the work of my colleague Andrea Ágnes Reményi meant an invaluable help.

Endnotes

1. The number within the bracket is the relative frequency of the collocate vs. base.

References


