Croatian Equivalents of of-Collocations

(A Mechanical Translation Approach)

As is known, mechanical translation calls for a specific approach in language research. This approach is dictated by the limited abilities of logical operation of computers, as opposed to the human brain. Since translating means supplying acceptable equivalents in another language, mechanical translation means supplying such equivalents in keeping with the analytical processes possible for and suitable to computers. All linguistic research for mechanical translation purposes must, consequently, be carried out with these conditions in mind.

Our particular object, as presented in this paper, was to examine the Croatian equivalents of the preposition of, as shown in a continuous sample of English text.

The text used was taken from the book Diseases Transmitted from Animals to Man, a collection of articles edited by T. G. Hill, and published in Springfield, Illinois, in 1953. A non-fictional work was chosen for obvious reasons: the primary object of mechanical translation is supplying prompt acceptable translations of technical texts. The choice of a veterinarian text, rather than a chemical or mathematical, for instance, was made because it ensured textual continuity (no formulas and the like).

The Croatian translation was supplied by the author, with the observance of the following principles: practical acceptability (i.e. no aesthetic considerations) and maximum word-for-word parallelism (to ensure programming simplicity). The published Croatian translation, Bolesti koje se od životinja prenose na čovjeka (Zoonoze), Zagreb, 1961, was consulted only when technical terms were involved.

The size of the text actually examined was 1,302 sentences, a sample securing a 95-per cent degree of reliability (tolerating a ± 2.5 deviation on either side) in making conclusions valid for the whole text, i.e. the whole book, or 7,383 sentences (esti-
mated). Expressed in number of words, the sample contained a total of 32,432 words.

IBM business machines were used in processing this material. The sample was encoded and transferred onto punched cards. A total of 1,528 cards was used, as 228 longer sentences required two cards each, and 4 exceptionally long sentences three cards each. Further stages in processing were interpreting, sorting and tabulating. In its final form, the sample was presented on 4,307 tabulator sheets, showing the distribution of each of the 161 syntactic elements examined (including 13 punctuation mark codes) at any (from 1st to 112th) position in the sentence. The distribution obtained was further subdivided by sorting according to the next left element. Our mechanical sorting did not go beyond this context (-1,0). Though a mechanical sorting, and subsequent tabulating, for the next right element as well (the context -1,0,+1) would have greatly facilitated our task, both inadequate funds and the prospect of additional piles of tabulator sheets together acted as a decisive deterrent.

The total frequency of of in the sample was 1,921, or 5.9 per cent of the text.

There was a zero equivalent, i.e. no translation necessary in Croatian, in 1,720 cases (or 89.5 per cent). In terms of basic linguistic information for mechanical translation programming, this would be No output for of unless . . .

The next step is outlining the other cases, falling under this unless, i.e. the remaining 201 (10.5 per cent) cases. Their Croatian equivalents were:

<table>
<thead>
<tr>
<th>Croatian</th>
<th>Frequency</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;od&quot;</td>
<td>146</td>
<td>idiom-integrated 8</td>
</tr>
<tr>
<td>&quot;za&quot;</td>
<td>15</td>
<td>&quot;na&quot; 7</td>
</tr>
<tr>
<td>&quot;o&quot;</td>
<td>11</td>
<td>&quot;iz&quot; 5</td>
</tr>
<tr>
<td>&quot;da&quot;</td>
<td>8</td>
<td>&quot;u&quot; 1</td>
</tr>
</tbody>
</table>

As these equivalents were dependent on their immediate contexts, here is a detailed survey of the of-collocations, requiring the above solutions:

1

In an initial position in the sentence, irrespective of the right-hand context, of has the Croatian equivalent "od".

6 cases

---

1 The formula used to determine sample size was \( n = \frac{t^2 R}{P \cdot I} \). Cf. V. Serdar, *Udžbenik statistike*, Zagreb, 1961, pp. 144–146.
There is a variety of Croatian equivalents for of, when the latter is a member of certain collocations, listed systematically below. It is interesting to note, from this survey, that it is the left-hand context which is decisive in the choice of the Croatian equivalent in a majority (177) of the cases. The right-hand context determines the of-equivalent in only 46 cases. Both contexts affect the selection of the equivalent in 60 cases.

a) Left-hand context of-collocations

1. Proper nouns\(^2\) + of
   Insufficient without the right-hand context. See under b)
   (collocations: of + proper nouns; of + geographical nouns).

2. Adjective\(^3\) + of\(^4\)
   No. of cases: 14
   Croatian equivalent: "za"
   9
   none
   6
   "od"
   1
   "na"
   1
   "da"

3. that + of
   A wider left-hand context required:

   3.1. Vb (exc. be for ms) +
   \[
   \begin{align*}
   \text{that} & + \text{of} \\
   \text{to the effect} + \text{that} + \text{of}
   \end{align*}
   \]
   that = preposition
   1 case
   Croatian equivalent: "od"

   3.2. preposition + that + of
   \[
   \begin{align*}
   \text{be forms} + \text{that} + \text{of}
   \end{align*}
   \]
   that = demonstrative
   4 cases
   Croatian equivalent: none

\(^2\) Excluding geographical names, which are treated as a separate class on account of the different prepositional collocations they can join.
\(^3\) Positive form only.
\(^4\) This is a more detailed survey of these lexical ("non-structural") collocations:

<table>
<thead>
<tr>
<th>&quot;za&quot;</th>
<th>no Croat. equivalent</th>
<th>&quot;od&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>capable + of</td>
<td>(7) many + of + the + Npl</td>
<td>(3) dead + of</td>
</tr>
<tr>
<td>indicative + of</td>
<td>(4) several + of + the + Npl</td>
<td>(1) free + of</td>
</tr>
<tr>
<td>confirmatory + of</td>
<td>(1) much + of + this</td>
<td>(1) many + of + the + Num</td>
</tr>
<tr>
<td>true + of</td>
<td>(1) worthy + of + Nsg</td>
<td>(2) few + of + many</td>
</tr>
<tr>
<td>suggestive + of</td>
<td>(1) most + Adj + of</td>
<td>(2)</td>
</tr>
<tr>
<td>&quot;na&quot;</td>
<td>&quot;da&quot;</td>
<td></td>
</tr>
<tr>
<td>irrespective + of</td>
<td>(1) skeptical + of + -ing</td>
<td>(1)</td>
</tr>
</tbody>
</table>

157
4. **others** + of  No. of cases: 1  Croatian equivalent “od”
5. **each** + of  
6. **either** + of  “od”
7. **all** + of  none
8. **some** + of  none⁵
9. **none** + of⁶  none

10. **cardinal numbers** + of

Not sufficient without the right-hand context:

10.1 **card. num.** + of + card. num.⁷  No. of cases: 17
Croatian equivalent: “od”
10.2 **card. num.** + of + the + card. num.
10.3 **card. num.** + of + a + total + of + card. num. 1  “od”
10.4 **card. num.** + of + whom⁸  1  “od”
10.5 **card. num.** + of + any other context  8  “od”¹⁹

11. **one** + of  29  “od”¹¹⁰
12. **ordinal number** + of  2  “od”¹¹¹
13. **(per) cent** + of

Right-hand context also required, the material here reacting in the same way as in the collocation: **card. num. + of**:

13.1 **cent** + of + card. num.  4  “od”
13.2 **cent** + of + any other context  24  none¹²

---

⁵ However, **of** = “od” if followed by: these, those, them, us.
⁶ All colocations of the type: indefinites + of require further research.
⁷ Note: dividing elements are possible between the left-hand **card. num.** and **of**, such as the comma (3 cases), with its own right-hand context (**card. num.** + **per** + **cent**..) and “opening” brackets (2 cases), with own right-hand context (**card. num.** + **per** + **cent**; card num + cc.).
⁸ Further research is necessary into the collocations: **card. num.** + **of** + relatives.
⁹ Exceptional cases: **card. num.** (= years, in dates) + **of** + proper nouns (excluding geographical), where **of** = “od” (because of the proper noun); and **card. num.** (= years) + **of** + Npl, where **of** = “od” (on account of the lexical element in the sixth position of the left-hand context).
¹⁰ Including three cases where **of** would be better rendered with no Croatian equivalent. Two of them have a removed lexical determiner (“case”), and in the third there is the anaphoric “one”. However, leaving **of** in these cases with no Croatian equivalent would require very complicated programming for mechanical translation. As the uniform solution: **of** = “od” would most probably not damage the meaning of the sentence, neglecting these cases is suggested as the best practical solution.
¹¹ One of these two cases can be solved either as **of** = “od”, or with no Croatian equivalent. However, the other example must have **of** = “od”. For the sake of uniformity, **of** = “od” was accepted. Undoubtedly, further research is required.
¹² All the solutions are lexical. It should be noted that practically all these cases (with some difficulty in only three of them) could also be resolved as **of** = “od” in our rough practical translation.
14. Vb + of\textsuperscript{13}  No, of cases: 12  Croatian equivalent: “od”
15. Vbs + of  4  none
16. Vbed + pret + of  3  “o”
17. Vbed pp + of  1  “da”
18. , + of\textsuperscript{14}  9  “od”
19. ) + of\textsuperscript{15}  1  none

b) Right-hand context of-collocations

1. of + proper noun\textsuperscript{16}  9  none
2. of + geographical name  25  none
   (In most of the following collocations, as will be seen, the
   left-hand context must also be considered)
3. proper noun + of + geog. n.  5  “iz”
4. capitals\textsuperscript{17} + of + geog. n.  2  “u”
5. of + whom (left-hand context decisive. Cf. a/ 18.)
6. of + which (left-hand context decisive. Cf. a/ 7, 9, 10, 11.)
7. Nsg/Npl + of + which  2  none
8. , + of + which  1  “od”

Note. In two cases of the collocation of + which, the right-hand
lexical context was decisive. The lexical determiners record (n) and capable were three and seven positions
away respectively, their Croatian equivalents being “o” and “za”.

9. of + cardinal numeral\textsuperscript{18}  Left-hand context decisive:

\textsuperscript{13} In one case there were dividing elements (its + Nsg) between
Vb and of.
\textsuperscript{14} “No equivalent for of” could have been possible in a few cases,
but considerations of uniformity prevailed. Further research needed.
\textsuperscript{15} That is, any “closing” bracket + of. Investigation of more ex-
tensive texts suggests itself.
\textsuperscript{16} Note: Dividing element (Sir) in one case. Also, there was one
occurrence of card. num. + of + proper noun, in which the right-hand
context was determining. Therefore, no equivalent was accepted as a so-
lution rather than of = “od” (which would be required by the collocation
card. num + of).
\textsuperscript{17} If the preceding word begins with a capital letter, and is itself a
geographical term (City, Island, Valley, etc.), the basic rule for the collo-
cation of + geog. name is applicable, i. e. there is no Croatian equivalent
for of. If, however, this word, though beginning with a capital, is not a
geographical term (Academy, Clinic, and the like), of = “u”.
\textsuperscript{18} Nineteen cases of this collocation have been omitted, as they have
already been treated under these left-hand collocations: card. num + of
(9 cases), per cent + of (3), that + of (2), (,) + of (2), one + of (1). Also,
in 2 cases, this collocation has previously been treated as of in the
initial position.
9.1. nouns of measure\textsuperscript{19} + of + card. num.
No. of cases: 21 Croatian equivalent: "od"

9.2 other nouns + of + card num.
\begin{tabular}{ll}
  3 & "od" \\
  14 & none \\
  10. \textit{are} + of + card. num. & 1 & "od" \\
  11. \textit{while} + of + card. num. & 1 & "od" \\
  12. \textit{a total} + of + card. num. & 10 & none \\
  13. of + -ing\textsuperscript{21} & 3 & "da"
\end{tabular}

c) Individual lexical solutions of of-collocations

No. of cases: 24\textsuperscript{22} Croatian equivalent: "od"
\begin{tabular}{ll}
  6 & "na" \\
  5 & "o" \\
  3 & "da" \\
  1 & "u"
\end{tabular}

d) Idiom-integrated of-collocations

8 cases\textsuperscript{23}

\textsuperscript{19} This wide common term, arbitrarily chosen, includes the following nouns: \textit{period} (6 cases), \textit{series} (2), \textit{fatality (rate)} (2), and one each of: \textit{incidence, figure, temperature, concentration, doses, strength (= concentration), consignment,} and \textit{herd. Injection} (1), when followed by terms of quantity (such as: \textit{cc}), may also be included here.

\textsuperscript{20} Concretely: \textit{man, loss, and outbreak}.

\textsuperscript{21} The three -ing forms encountered were: gerund, being, and having. Dividing elements are possible.

\textsuperscript{22} Out of which, 6 cases had a left-hand context determiner, and 17 a right-hand context determiner. In most cases, there are one or more dividing elements between \textit{of} and the determiner. In one case, superficially \textit{that + of}, the number of dividing elements is even 24. In one specific case (\textit{cc} + \textit{of} + \textit{a} + card. num.), the solution \textit{of} = "od" is required for the sake of clarity — to separate the abbreviation from the numeral.

\textsuperscript{23} They were: \textit{on account of}, \textit{in view of}, and one each of: \textit{as of}, \textit{as a matter of fact, of course, animals of prey}. The last example indicates the broad conception of the category "idioms" applied.