

Željko Bujas

A Manual Conversion Procedure in Bilingual (Croatian-English) Lexicography

1.0. A not so negligible irritant for the discriminative user of a bilingual dictionary — especially if he is a lexicographer himself — is the undue disparity in the inventory of the dictionary's two halves, foreign-native and native-foreign. The stress is on 'undue' since one does expect minor divergences, resulting from the inevitably different linguistic and cultural approaches adopted by the lexicographer when compiling each half, or volume, especially when different authors are involved.

1.1. Thus, for instance, while *sporrán* or *coleslaw* or *Christmas cracker* will probably be included in a medium-sized English-Croatian dictionary, these items are rather less likely to be encountered, glossing some Croatian entry, in the companion Croatian-English volume (or an independent Croatian-English dictionary). It is indeed unnecessary for these glosses to be contained there since they denote things typical of Anglo-Saxon civilization and, as such, are absent from the normal set of references of a speaker of Croatian. Conversely, *klijet*, *kulen*, *zubatac* and similar typically Croatian items will not normally turn up as the Croatian glosses for any English entry in an English-Croatian dictionary of any size, due to their absence from the normal Anglo-Saxon frame of reference.

1.2. What one cannot accept, however, is the absence of perfectly adequate (some indeed felicitous) English or Croatian glosses absent from the companion volume, or the other half — especially when the dictionaries in question are by the same author. The extent of this 'negative overlap' may range from minor flaws, through a nagging awareness of additional possible equivalents in the other language, to glaring, crippling omissions.

2.0. I will illustrate this by a few examples from the only existing parallel English and Croatian bilingual dictionaries (in the 30—50,000 headword range) by the same author. The dictionaries in question are by Milan Drvodelić (1883—1963), former lecturer of English at Zagreb University. His English-Croatian Dictionary (DEH henceforth) has so far gone through four editions: of 1954, 1962 (revised by Ž. Bujas), 1970 (second revision by Ž. B.) and 1973 (third revision by Ž. B.)¹ The Croatian-English companion volume (DHE for short) has appeared three times so far: in 1954, 1961 (minor revisions by L. Spalatin) and 1971 (new impression).

2.1. A bilingual dictionary should be visualized as made up of a left-side inventory of headwords and sub-entries, and a right-side inventory of glosses (translation equivalents). These two corpora may evidently switch sides, with the right-side items upgraded to headword status and the former headwords relegated to the subordinate standing of glosses. This should be obvious to every bilingual lexicographer, especially those producing parallel, ambidirectional (to coin a word) volumes. He may, therefore, be reasonably expected to have utilized the undoubtedly substantial lexicographic potential involved, by having converted each of his volumes in the manner suggested. Particularly the foreign-native volume, which is normally first to appear. That M. Drvodelić has failed to do so — has failed to convert his DEH (1954) material into a Croatian-English file — is readily demonstrable by almost any of his entries.

2.2. This is, for example, what *lot* (noun) looked like in the DEH 1954 — minus its phraseological section:²

lot [lot] *n* ždrijeb, ždrebanje: udio, sudjelovanje;
sudbina; parcela, gradilište; sila, mnoštvo;
količina istovrsne robe

Visualized as made up of implicit English-Croatian translation pairs (*lot* : *ždrijeb*, *lot* : *ždrebanje*, *lot* : *udio*, etc.), this entry converts easily to a new set of Croatian-English pairs:

¹ All published by Školska Knjiga, Zagreb, under slightly varying titles: *Englesko-hrvatski rječnik* (1954), *Englesko-hrvatskosrpski rječnik* (1962, 1970) and *Englesko-hrvatski ili srpski rječnik* (1973). These variations reflect the changing official attitudes, but have in no way affected the dictionary itself which represents the Croatian or Western Standard type of Serbo-Croatian, native both to the author (M. Drvodelić) and the revisor (Ž. Bujas). The same is true of the companion volume which has the same publisher and has appeared under the titles: *Hrvatsko-engleski rječnik* (1954) and *Hrvatskosrpsko-engleski rječnik* (1961, 1971).

² With a few spelling, punctuation and typographic improvements from later editions, included to avoid what would have otherwise been minor but needless attention-diverting flaws.

ždrijeb : lot
 ždrebanje : lot
 udio : lot
 sudjelovanje : lot
 sudbina : lot
 parcela : lot
 gradilište : lot
 sila : lot
 mnoštvo : lot
 količina [istovrsne robe] : lot
 [količina] istovrsne [robe] : lot
 [količina istovrsne] robe : lot

2.3. When checked against the companion-volume Croatian-English dictionary by the same author (DHE 1954), the 12 Croatian-English pairs thus obtained amply illustrate the hidden potential of bilingual-dictionary conversion as a compilation technique in lexicography. This new material reveals its applicability to expanding and improving the said Croatian-English dictionary in triple fashion:

a) Adding new English glosses to existing Croatian headwords.

E. g. ždrebanje : lot
 udio : lot
 gradilište : lot

We have omitted three pairs — *sila* [= *sva* ~] : (a) lot, *mnoštvo* : (a) lot, and *sudjelovanje* : lot — whose usefulness is only marginal, or in the nature of hints. For instance, *sudjelovanje* : lot might remind the looker-up of *throwing in one's lot with* as a possible translation for the Croatian *sudjelovanje u borbi*

b) Supplying hitherto absent Croatian headwords.

parcela : lot

c) Suggesting new Croatian collocations.

količina (~ *istovrsne robe*) : lot
istovrs|tan (*količina* ~*ne robe*) : lot
rob|a (*količina istovrsne* ~*e*) : lot

And all this resulting from only one converted entry — and only its first, non-phraseological, half!

2.4. Now, obviously, once an extensive project like the one discussed here is considered, there would not be much point in converting the earliest, unrevised and smaller in volume version

of the DEH (of 1954, with 946 pages), rather than the latest expanded edition available at the time of conversion (the 1163-page edition of 1970 in our case). The only gain would be the precisely established extent of the material left untransferred by the author from his English-Croatian volume — the extent of the 'negative overlap'. Using the revised version of the DEH, however, one obtains a larger and improved Croatian-English lexicographical file. This more than compensates for the fact that the presence of the revisor's material makes it now a little harder to decide which negative-overlap items are original (i. e. due to M. Drvodelić).

2.4.1. For example, the DEH entry just analyzed provides, in its 1970 expanded version, four more translation pairs (bringing the total to 16, or 25% more):

partija : lot
filmski (~ grad) : lot (sl)
grad (filmski ~) : lot (sl)
[filmski] studio : lot (sl)

2.5. So far, as pointed out, we have only discussed what happened when the nonphraseological section of our illustrative entry was converted. We will now consider its phraseological section, which looked like this in DEH 70:

..... | (coll) *it's a ~ mnogo je*; (coll) *a ~ of people* (ili ~s of people) *mnogo ljudi*; (coll) ~s and ~s *vrlo mnogo*; to cast (draw) ~s *vući ždrijeb, ždrebati*; the ~ *fell upon me ždrijeb je pao na mene*; to cast (throw) in one's ~ with *dijeliti sudbinu s nekim*; it fell to my ~ *zapalo me*; it is my ~ to na meni je da; my ~ in life *moja životna sudbina*; hard ~ *težak život* (sudbina); (coll) *he is a bad ~ on je ništarija*; a building ~ *građevna parcela*; these pictures form a single ~ *ove se slike prodaju zajedno* (ne pojedinačno); (coll) *take the (whole) ~ uzmi sve*; (iron) *a ~ you care briga tebe!*

2.6. Converting the phraseological sections — which normally account for at least one-half of the total volume of any

bilingual dictionary of comparable size — means involvement with differently structured material. As a consequence, this conversion will overwhelmingly produce pairs of non-one-for-one equivalents, that is of contrasted Croatian and English items at group, phrase, clause and sentence levels. Naturally, the two languages will now have changed sides. Also, in English equivalents the swung-dash, indicating the position of the original headword (*lot*), is now best replaced by the word itself:

- mnogo je* : (coll) *it's a lot*
mnogo ljudi : (coll) *a lot of people (ili lots of people)*
vrlo mnogo : (coll) *lots and lots*
vući ždrijeb, ždrebati : *to cast (draw) lots*
ždrijeb je pao na mene : *the lot fell upon me*
dijeliti sudbinu s nekim : *to cast (throw) in one's lot with*
zapalo me : *it fell to my lot*
na meni je da : *it is my lot to*
moja životna sudbina : *my lot in life*
težak život (sudbina) : *hard lot*
on je ništarija : (coll) *he is a bad lot*
građevna parcela : *building lot*
ove se slike prodaju zajedno : *these pictures form (ne pojedinačno) : a single lot*
uzmi sve : (coll) *take the (whole) lot*
briga tebe! : (iron) *a lot you care*

2.7. Scrutinizing this ragged-edged phrasal-pair list, pivoting as it were upon its colons, one quickly becomes aware of two things. First: the phraseological sections contain perhaps a little less, but still uniformly significant, potential for the expansion and improvement of the companion (Croatian-English) volume. Second: ordered like this, these sections bear a striking visual resemblance to segments of a KWIC (key-word-in-context) concordance.

Applying this particular concordancing technique to our phrasal-pair list, with each and every Croatian item functioning as alphabetized key-words, we will now produce another restructured version of the phraseological list under examination:

- | | | |
|-------|-----------------------------------|----------------------------------|
| | (iron) a lot you care | <i>briga tebe!</i> |
| | it is my lot to | <i>na meni je da;</i> |
| | to cast (throw) in one's lot with | <i>dijeliti sudbinu s nekim;</i> |
| [lot] | | <i>filmski grad, studio</i> |
| | <i>filmski</i> | <i>grad, studio</i> |

	<i>parcela,</i>	<i>gradilište;</i>
	a building lot	<i>građevna parcela;</i>
	<i>količina</i>	<i>istovrsne robe, partija;</i>
	(coll) it's a lot <i>mnogo</i>	<i>je;</i>
	the lot fell upon me <i>ždrijeb</i>	<i>je pao na mene;</i>
	it is my lot to <i>na meni</i>	<i>je da;</i>
	(coll) he is a bad lot on	<i>je ništarija;</i>
[lot]		<i>količina istovrsne robe,</i> <i>partija;</i>
(coll) a lot of people (<i>ili</i> lots of people) <i>mnogo</i>	<i>ljudi;</i>	
it fell to my lot <i>zapalo</i>	<i>me;</i>	
the lot fell upon me <i>ždrijeb je</i>		
<i>pao na</i>	<i>mene;</i>	
it is my lot to <i>na</i>	<i>meni je da;</i>	
(coll) it's a lot	<i>mnogo je;</i>	
a lot of people (<i>ili</i> lots of people	<i>mnogo ljudi;</i>	
(coll) lots and lots <i>vrlo</i>	<i>mnogo;</i>	
[lot]	<i>mnoštvo;</i>	
<i>sila,</i>	<i>moja životna sudbina;</i>	
my lot in life	<i>na mene;</i>	
the lot fell upon me <i>ždrijeb je pao</i>	<i>na meni je da;</i>	
it is my lot to		
these pictures form a single lot		
<i>ove se slike prodaju zajedno</i>	<i>(ne pojedinačno);</i>	
to cast (throw) in one's		
lot with <i>dijeliti usdbinu s</i>	<i>nekim;</i>	
(coll) he is a bad lot on <i>je</i>	<i>ništarija;</i>	
(coll) he is a bad lot	<i>on je ništarija;</i>	
these pictures form a single lot	<i>ove se slike prodaju za-</i> <i>jedno (ne pojedinačno);</i>	
	<i>pao na mene;</i>	
the lot fell upon me <i>ždrijeb je</i>	<i>parcela, gradilište;</i>	
[lot]	<i>parcela;</i>	
a building lot <i>građevna</i>	<i>ove se slike prodaju zaje-</i> <i>dno (ne pojedinačno);</i>	
these pictures form a single lot		
these pictures form a		
single lot <i>ove se slike</i>	<i>prodaju zajedno (ne</i> <i>pojedinačno);</i>	
[lot]	<i>količina istovrsne</i>	<i>robe, partija;</i>
to cast (throw) in one's		
lot with <i>dijeliti sudbinu</i>	<i>s nekim;</i>	

these pictures form a single lot	<i>ove se slike prodaju zajedno</i>	<i>se slike prodaju zajedno</i>	<i>(ne pojedinačno);</i>
[lot]		<i>sila, mnoštvo;</i>	
	<i>ove se</i>	<i>slike prodaju zajedno</i>	<i>(ne pojedinačno);</i>
[lot]	<i>filmski grad</i>	<i>studio</i>	
[lot]		<i>sudbina;</i>	
	<i>my lot in life</i>	<i>moja životna</i>	<i>sudbina;</i>
	<i>hard lot</i>	<i>težak život</i>	<i>(sudbina);</i>
to cast (throw) in one's	<i>lot with</i>	<i>dijeliti</i>	<i>sudbinu s nekim;</i>
		<i>udio,</i>	<i>sudjelovanje;</i>
(coll) take the (whole) lot		<i>uzmi</i>	<i>sve;</i>
(iron) a lot you care		<i>briga</i>	<i>tebe!</i>
	<i>hard lot</i>		<i>težak život (sudbina);</i>
[lot]		<i>udio, sudjelovanje;</i>	
	(coll) take the (whole) lot		<i>uzmi sve;</i>
	(coll) lots and lots		<i>vrlo mnogo;</i>
	to cast (draw) lots		<i>vući ždrijeb, ždrijebati;</i>
these pictures form a			
single lot	<i>ove se slike prodaju</i>	<i>zajedno (ne pojedinačno);</i>	
	<i>it fell to my lot</i>	<i>zapalo me;</i>	
[lot]	<i>ždrijeb,</i>	<i>ždrebanje;</i>	
to cast (draw) lots	<i>vući ždrijeb,</i>	<i>ždrebati;</i>	
	<i>lot</i>	<i>ždrijeb, ždrebanje;</i>	
to cast (draw) lots	<i>vući</i>	<i>ždrijeb, ždrijebati;</i>	
	<i>the lot fell upon me</i>	<i>ždrijeb je pao na mene;</i>	
	<i>hard lot</i>	<i>težak</i>	<i>život (sudbina);</i>
	<i>my lot in life</i>	<i>moja</i>	<i>životna sudbina;</i>

2.8. As can be seen, this mechanical approach yields an easy-to-consult listing of individual Croatian items from the whole entry. It also ensures that *not a single Croatian item* (as a potential headword of a Croatian-English translation pair) *is missed*. This, to be sure, means that not a single occurrence — among the hundreds or thousands — is missed of any Croatian 'functional word' (however empty of 'lexical meaning'). Barely hinted at in the above listing (4 *je* and 2 *na*), the profusion of such items, less than obviously and immediately useful in conventional dictionary writing, has nevertheless important lexicographic potential (cf. note 8).

3.1. Obviously, the most efficient way of converting mechanically (and unselectively) an English-Croatian dictionary into a Croatian-English concordance, to serve as a specific lexicographic file — would be by computer. In that case, the entire dictionary text (probably minus pronunciation) would have to be copied on to punched cards or paper tape. The beginning and end of English segments of the text would have to be marked by the use of any convenient sign. After corrections of copying errors, the final copy would be transferred on to magnetic tape and computer-processed (concordanced). The programs used would ensure that each and every item of the text (both English and Croatian) would appear as context, but only Croatian items would be used as key-words in the ultimate product of conversion: a Croatian-English Lexicographical Concordance (in the physical shape of a computer printout).³

3.2. Needless to say, few lexicographers could pay for such extensive concordancing, and equally few have free access to computers.⁴ On the other hand, most lexicographers are inured

³ To illustrate its size: Drvodelić's *English-Croatian Dictionary* would, computer-converted, yield a printout of some 4,000 pages (of 63 lines each) for its estimated 250,000-plus running words of Croatian text.

⁴ I am one of the lucky few who can afford to pay for large-scale computer concordancing. Funding by the Croatian Council for Research Work has made it possible for me to engage in a three-year project (1972—1975) to convert a desk-size English-Croatian dictionary into a Croatian-English Lexicographic Concordance. The dictionary in question (*Englesko-hrvatski rječnik*, R. Filipović, ed., Zora, Zagreb, 1971; 1,467 pages) is being copied on a Flexowriter machine at the Institut za lingvistiku, Filozofski fakultet, Zagreb. The resulting punched tape will be processed on an IBM 360/170 computer at the Zagreb Municipal Computer Center. The end product — estimated to contain some 900,000 Croatian items as key-words — will be a 14,000-page computer printout. Its purpose:

- a) To serve as a reverse Croatian-English lexicographical file in concordance form.
- b) To generate Croatian synonymy clusters by alternating English and Croatian glosses as headwords (cf. Ž. Bujas, "Kompjuterske sinonimije" [A Computational-Lexicographic Approach to Compiling Croatian Synonymy], in *Suvremena lingvistika*, Zagreb, 5—6, 1972, 19—26; with abstract in English).
- c) To compile exhaustive lists of bilingual word-formation pairs (by reverse concordancing).
- b) To establish various contrastive patterns in English and Croatian lexis (by normal concordancing).
- e) To compile English-Croatian and Croatian-English glossaries by vocabulary areas.
- f) To ensure uniformity in English-Croatian lexicographical material (both in terms of editorial intervention and a consistent application of established lexicographically productive patterns).
- g) To update English-Croatian lexicographical files.

to drudgery and can perform veritable feats of manual toil. So, three years ago, aided by my students, I launched a *Manual Conversion Lexicographical Project (English to Croatian)*. Its purpose was twofold. First: to devise an efficient and reliable procedure for a manual conversion of bilingual dictionaries. Second: to compile, in the process, a lexicographical file directly applicable to improving the companion volume (the DHE) and of general usefulness in Croatian-English lexicography.

3.3. My line of reasoning was as follows: Let us provide a number of copies of each dictionary (DEH) page. Let us then mark every Croatian item in the dictionary text⁵ (with, say, a slant line across it), but *so that each following Croatian item is marked always on the next copy of the same page*. Then, after reaching the last page of the batch, let us go back to the top page, repeating the process (coming full circle) a number of items, until the last Croatian word has been marked in first the left, then the right, column of the text on the page. So, in the end, we will have produced stacks of pages, with each page covered by a number of slants fairly evenly spaced over its face (staggered, as it were, by the number of copies used for each page). If we now cut each fully marked page along the center line, and then cut up each column half into as many slips as one gets by cutting between the slant marks — the result will be a number of slips roughly equal in size. On each of them will be one slant-marked Croatian item (the new headword), within several lines of context, and the former English headword in bold print (now relegated to the status of a gloss to the new headword).⁶

3.4. Some preliminary experimenting revealed that 15 copies of each DEH page would result in slant marks appearing about one to one-and-a-half inches apart on the face of each column (half-page). With the width of each DEH column being just under three inches, the average size of actual slips to be obtained after cutting up was, consequently, just under 3 in. long by 1—1½ in. wide. However, in view of the obvious fact that marking and cutting up the face of each page rendered its obverse side useless, 30 duplicates had to be secured for each original dictionary page. In other words, 30 copies of the dictionary had to be provided.

⁵ Which will be facilitated by their being in plain print, and thus visually distinct from the English items in italics or (when headwords) in bold print.

⁶ Every sixth slip or so (in cases of the long original English entries) will, after cutting up, be left without the former English headword. These will then have to be entered by hand on the slip margin — preferably before cutting up.

In the fall of 1970, the 30 copies were bought⁷ and taken apart by the members of my Third-Year Language Seminar until each leaf was loose. Each undergraduate was then assigned three dictionary pages (in 15 copies each, or a total of 45 pages) to slant-mark and cut up into slips. The resulting approximate 650 slips per Seminar member were then alphabetized by the marked Croatian item, to facilitate look-ups in the DHE 70.

3.5. The look-ups were carried out by following strictly regulated, unambiguous procedure steps. This approach was amply rewarded by the highly uniform distributions of look-up results (that is, of the categories obtained). In evaluating whether the new Croatian slant-marked entry, with its English equivalent, was a useful or potential addition to the DHE 70, each analyst tested all such entries, using the following six-step procedure:

If the Croatian (slant-marked) entry was:

a) **non-located** and **absent** from the DHE 70, it was assigned to C A T E G O R Y 1.

E. g. *arbalest* = *crossbow* was absent from the DHE 70, though present in the DEH of the same year (as *crossbow* = *arbalet*).

b) **non-located** and **present** in the DHE 70, and its **English equivalents** (the former headwords) were the **same**, it was assigned to C A T E G O R Y 2.

E. g. *dosadivati* = *importune*, because exactly this pair (as *importune* = *dosadivati*) was present in the DEH 70.

c) **non-located** and **present** in the DHE 70, but its **English equivalents** were **different**, it was assigned to C A T E G O R Y 3.

E. g. *dočekati* = *greet*. Although *greet* = *dočekati* is present in the DEH 70, the headword *dočekati* in the DHE 70 has only the following English equivalents: *receive, welcome; (go to) meet; see one's hope realized, live to see*.

d) **located** and **absent** from the DHE 70, it was assigned to C A T E G O R Y 4.

E. g. *to je za mene kineski* = *that's Greek to me*. Though included in the DEH 70, this phrase was absent from the DHE 70, where, under the headword

⁷ The funds (2,000 dinars) were provided by the Commission for Research Work of my school, the Filozofski Fakultet, Zagreb University.

kineski (adj), we find only two collocations: *kineski jezik* and *kineski narod*.

- e) **collocated** and **present** in the DHE 70, and its **English equivalents** were the **same**, it was assigned to **C A T E G O R Y 5**.

E. g. *grčki jezik* = *Greek*, because the same pair (naturally, in reverse: *Greek* = *grčki jezik*) was found in the DEH 70.

- f) **collocated** and **present** in the DHE 70, hut its **English equivalents** were **different**, it was assigned to **C A T E G O R Y 6**.

E. g. *nepovoljan vjetar* = *cross-wind*. Although *cross-wind* = *nepovoljan vjetar* is present in the DHE 70, *contrary (wind)* is the only English equivalent listed under either *nepovoljan* (~ *vjetar*) or *vjetar* (*nepovoljan* ~) in the DHE 70.

3.6. Obviously, in terms of their usefulness for expanding and improving the Croatian-English volume, categories 1, 3 and 6 are the most desirable. Category 4 contains a lot of potentially useful material. Categories 2 and 5, made up of translation pairs already included in the DHE 70 were of no use.

A high total share of the 'pay-dirt' correspondence categories (1, 3, 6 and, partially, 4) will have been anticipated, in view of the ample potential of only one random entry analyzed (lot¹ n; cf. 2.3.). We only have to remember that each page in the DEH 70 holds an average 20—30 entries, while there are 1,162 full pages in the dictionary.

4.1. A much clearer idea of the share of these categories will be obtained from the actual precise distribution data for each of the six categories per alphabet letter, as yielded by the 48 third-year language-seminar reports, submitted in the academic year 1972/73. As each report analyzed three dictionary pages, these then are the actual distribution figures for 144 pages in the DEH 70, or 12.4% of the dictionary's volume. The location of these 144 pages within the dictionary body may be represented as follows:

Table 1

Hundred pages	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	Total
Number of pages converted	1	—	12	17	43	29	34	—	—	—	8	—	144

In spite of its obvious concentration in the mid-alphabet range (D through N, to be precise), this is still undoubtedly a sufficiently dispersed sample to make the data derived from it highly reliable for the whole DEH mass, when completely converted into a Croatian-English slip file in the near future.

4.2. Here, then, are the (absolute and relative) distributions established for our 144-page sample, broken down per alphabet letter and each of the six correspondence categories:

Table 2 Absolute distribution of the six conversion correspondence categories in the sample analyzed

Croatian item beginning in	1	2	3	4	5	6	Total
A	33	7	20	136	3	3	202
B	60	149	168	753	32	45	1,207
C	18	19	12	114	6	7	176
Č	12	41	51	272	18	13	407
Ć	2	3	6	37	2	—	50
D	129	128	180	1,132	34	59	1,662
DŽ	16	2	—	9	—	—	27
Đ	1	7	2	8	1	—	19
E	61	8	8	67	—	—	144
F	25	13	12	65	2	2	119
G	91	113	133	461	27	27	852
H	56	50	46	130	6	5	293
I	167	209	208	1,049	33	41	1,707
J	35	66	54	527	11	6	699
K	93	178	140	1,266	50	42	1,769
L	35	42	53	299	10	10	449
LJ	7	14	23	54	1	—	99
M	96	181	197	773	31	76	1,354
N	330	486	604	1,945	31	59	3,455
NJ	8	12	8	41	—	1	70
O	117	252	379	1,199	48	36	2,031
P	232	442	795	2,414	70	91	4,044
R	57	114	205	588	24	31	1,019
S	169	270	426	2,487	76	90	3,518
Š	39	31	49	240	5	22	386
T	67	121	155	682	25	20	1,070
U	91	131	291	1,127	20	24	1,684
V	42	92	142	770	19	22	1,087
Z	67	140	289	936	32	34	1,498
Ž	24	24	46	162	5	6	267
Total	2,180	3,345	4,702	19,743	622	722	31,364

Table 3 Relative distribution of the six conversion correspondence categories in the sample analyzed

Croatian item beginning in	1	2	3	4	5	6	Share of total letter in sample
A	16.3	3.5	9.9	67.3	1.5	1.5	0.64
B	5.0	12.4	13.9	62.4	2.7	3.7	3.86
C	10.2	10.8	6.8	64.8	3.4	4.0	0.56
Č	3.0	10.1	12.5	66.8	4.4	3.2	1.30
Ć	4.0	6.0	12.0	74.0	4.0	—	0.16
D	7.8	7.7	10.8	68.1	2.1	3.6	5.31
DŽ	59.3	7.4	—	33.3	—	—	0.09
Đ	5.3	36.8	10.6	42.1	5.3	—	0.06
E	42.4	5.6	5.6	46.5	—	—	0.46
F	21.0	10.9	10.1	54.6	1.7	1.7	0.38
G	10.7	13.3	15.6	54.1	3.2	3.2	2.72
H	19.1	17.1	15.7	44.4	2.1	1.8	0.94
I	9.8	12.2	12.2	61.5	1.9	2.4	5.45
J	5.0	9.4	7.7	75.4	1.6	0.9	2.23
K	5.0	10.1	7.9	71.6	2.8	2.4	5.65
L	7.8	9.4	11.8	66.6	2.2	2.2	1.43
LJ	7.1	14.1	23.2	54.6	1.0	—	0.32
M	7.1	13.4	14.6	57.1	2.3	5.6	4.33
N	9.6	14.1	17.5	56.3	0.9	1.7	11.04
NJ	11.4	17.1	11.4	58.6	—	1.4	0.22
O	5.8	12.4	18.7	59.0	2.4	1.8	6.49
P	5.7	10.9	19.7	59.7	1.7	2.3	12.92
R	5.6	11.2	20.1	57.7	2.4	3.0	3.26
S	4.8	7.7	12.1	70.7	2.2	3.6	11.26
Š	10.1	8.0	12.7	62.2	1.3	5.7	1.23
T	6.3	11.3	14.5	63.7	2.3	1.9	3.42
U	5.4	7.8	17.3	66.9	1.2	1.4	5.38
V	3.9	8.5	13.1	70.8	1.8	2.0	3.47
Z	4.5	9.4	19.3	62.5	2.1	2.3	4.79
Ž	9.0	9.0	17.2	60.7	1.9	2.3	0.85
Total	7.0	10.7	15.0	63.0	2.0	2.5	100.00

4.3. Comment on Tables 2 and 3:

4.3.1. The most readily observable data are the total (read: average) shares of each conversion-correspondence category. Polarizing — in terms of direct lexicographical applicability — as useful (1, 3, 6), partially useful (4) and non-serviceable (2, 5), these categories reveal a promising combined share of 24.5% (+ 63.0%?) applicable items, as opposed to the aggregate 12.7% for the non-serviceable translation pairs.

4.3.1.1. Category 4, alone accounting for two-thirds of all items, obviously calls for a closer look. A pilot analysis (of page 644 of the DEH 70, intuitively selected as fairly representative of our 144-page sample) has shown that at least 70% of category

4 items (or some 44% of the sample total) are directly applicable to the improvement and expansion of the companion volume (DHE).⁸ Added to the aggregate share of categories 1, 3 and 6, they now bring the total of useful items in the sample up to (24.5 + 44 =) over 68%, leaving only slightly over 30% as the non-serviceable item total.

4.3.1.2. Converting all these relative distributions for our 144-page sample into absolute figures for the entire mass of the DEH 70 (with its, 1,163 pages), we may, with a high degree of reliability, predict the following absolute distributions after our manual conversion project has been completed:

12,700	27,100	38,000	159,200	5,000	6,300	253,300
(1)	(2)	(3)	(4)	(5)	(6)	Total ⁹

In other words — after splitting category four 70% : 30% (or 101,440 : 57,760) — we can expect to obtain a total of (17,700 + 38,000 + 101,440 + 6,300 =) 163,440 items of immediate lexicographical applicability, as opposed to (27,100 + 57,760 + 5,000 =) 89,860 non-usable, or at least not immediately applicable, items. However, even these latter contain, as already emphasized, a certain contrastive potential of undoubted lexicographical applicability.

4.3.2. Other observable patterning reflects deviations from the average distributional figures, due either to the specific character of the particular (alphabetic) portion of the Croatian vocabulary, or to specific combinations of category properties and Croatian vocabulary features.

4.3.2.1. Thus, for instance, the most conspicuous divergences from the average, i. e. total, values for category 1 (“Croatian item absent as entry”) are found in items beginning in A, DŽ, E, F and H (cf. Table 3). These are evidently the ‘foreign’ (non-Slavic) words — of Greek, Latin, French, English, Turkish origin etc. — so hard to accept for the old generation of diction-

⁸ Most items in the 30% portion of the non-serviceable material are illustrative prepositional phrases, explanations in parentheses, functional words in idioms, and the like. Part of the incidence of the reflexive pronoun *se*, and of the explanatory relative clauses may be added here — but by no means all. This is because they include certain contrastively functioning patterns that have unmistakable lexicographic potential. Discussing these, however, calls for a separate article.

⁹ Obtained by first computing the average total per page in the sample analyzed at (31,364:144=) 217.8 and multiplying this by 1,163. The result (253,301.4 precisely) then provided the absolute value for one per cent (2,533), and this was used to arrive at absolute figures for each category (rounded off to hundred before inclusion in these expected distributions).

ary writers (M. Drvodelić among them), with their prescriptive, late-19th-century puristic ideas about language.

4.3.2.2. Or, examining the category 4 column in Table 3 for distribution irregularities, we will observe a conspicuously above-average presence of category 4 items in letters Č, J, K, S and V — but also their under-representation in DŽ, Đ, E and H (to list only the lowest values). The explanation is simple in either case. The over-representation in Č, J, K, S and V is due to the presence of high-frequency functional words, such as *će, je, se, vi*, to name but a few, and of the equally frequent (in dictionary text) relative *koji* in its many oblique forms. As to the conspicuous under-representation DŽ, Đ, E and H, it should be pointed out that most Croatian words beginning in DŽ are of Turkish origin, that those in E and H (as already stated) have a high share of the foreign element, especially E. There are in fact no functional words beginning in any of these letters. As to Đ, its absolute incidence in the sample was so low (19) that all its distributions must be regarded as unreliable.

5.1. In conclusion, let me illustrate the high lexicographical value of the material compiled as a result of the procedures described in this report. Selecting at random only eleven DEH 70 pages (or just 1 per cent of the entire mass to be converted), and further limiting this sample to only the items beginning in BA-,¹⁰ we get the following twelve translation pairs, tabulated here in alphabetical order:

Table 4

Categ. ¹¹	Croatian entry	English equivalent	DEH 70 page ¹²
4	baciti se u trošak	go to the expense	310
4	baciti se na zemlju	hit the dirt (<i>sl</i>)	244
4	pipa baëve	key	532
4	bim-bam!	ding-dong!	242
3	banket	dinner-party	242
4	svršiti bar pola posla	be over the hump (<i>A sl</i>)	468
1	barhent (text)	dimity	242
3	barometar	glass	396
4	baš kao da	even as if	302
4	baš kad	even as	302
4	nada u baštinu	expectation	310
3	bazar	dime store (<i>A coll</i>)	242

¹⁰ An equally randomly picked combination of letters — though the high-frequency functional words were deliberately avoided.

¹¹ No category 6 items were found in this very restricted sample and, naturally, no category 2 and 5 items were listed.

¹² No useful (category 1, 3, 4 and 6) items were found on pages 488, 506, 528 and 536.

5.1.1. The tabulated items speak for themselves. If we had not chosen to limit ourselves to just eleven pages, the whole converted file would have yielded a minimum anticipated 160 useful items in BA- (estimated at about one-fifth of all the words beginning in B; and minus 30% of the category 4 share). If all useful items in B had been chosen, the list would have quintupled to an approximate 800. And so forth.

5.2. It calls for no special effort of imagination to visualize the stupendous wealth of useful lexicographical material over the full alphabet range, already (cf. 4.3.1.2.) estimated to contain a total of 163,440 items of immediate lexicographical applicability. Especially when this mass is added — to expand and improve it — to the companion-volume Croatian-English dictionary (DHE), with its estimated bulk of just under 200,000 (198,850 to be exact)¹³ Croatian-English translation pairs. Adding 163,440 items to a body of 198,850 similar units represents an increase of 81.1%. Now even one-tenth of that increase (16,000 items) would have made itself felt, as it would have affected every other main DHE entry (of an estimated rough total of 30,000). The full useful increase that can now be anticipated thanks to the analyses carried out in this paper (the said 163,440 items), will mean an average addition of five units per each main DHE entry — itself holding no more than six or seven glosses (i. e. right-hand halves of potential conversion pairs).

5.3. Part of this lexicographically productive mass is, as already stressed (cf. 2.4.) not due to M. Drvodelić's original DEH material but to the interventions by the revisor (Ž. Bujas). The very increase in size — 22.7% — between the DEH 54 (946 pages) and the DEH 70 (1,163 pages) testifies to the revisor's direct share. When deletions (often extensive) are added, the revisor's share rises to an estimated minimum of 30%. This still leaves 70% of the increase established above (i. e. of 81.1%) — or ca. 57% — as the invariably present, *inherent conversion potential* of any bilingual dictionary of the size analyzed (30—50,000 headwords), even when the same author is involved.

5.4. This share (57%) should stay roughly the same in larger dictionaries. The increasing number of entries specific for the Anglo-Saxon civilization and its natural-geographic environment — that is, of non-convertible items — will, in my view, be largely offset by the increasing number of items required by the

¹³ This figure was obtained by multiplying 913 — the total of DEH 70 page (cf. note 9), intuitively accepted as also approximately valid for the DHE.

more encyclopedic character of the dictionary (scientific, technological and similar entries). These latter are obviously convertible since a Croatian left side of any desk-size dictionary could not dispense with them.

However, as the dictionary size gets smaller, the said share (57%) is bound to fall off. A 5,000-entry dictionary, for instance, will contain a significantly higher percentage of the 1,000—2,000 most frequent words in English. These represent the real core of the dictionary and their Croatian glosses (very limited in selection) will tend to be highly predictable, and as such of low conversion potential.

5.6. Still, if we keep above the 2,000-entry level (below which one should in any case speak of vocabulary lists or glossaries rather than dictionaries), there is indeed no bilingual dictionary too small or unpretentious for its conversion not to yield valuable material. The value of this material is twofold. It will, of course, aid us practically — in dramatically improving the existing parallel bilingual dictionaries (or in compiling them where not available). Apart from this, however, it may also assist us — and this cannot be overemphasized — in offering new approaches and insights into contrastive patterning in lexis and syntax (to the degree that the latter operates in lexicography). Part of these will be fed back into lexicography, helping to make it a less intuitive, a more formalized, more truly linguistic discipline. The remaining patterns, however, joining the main body of contrastive observations, are bound to make not inconspicuous contribution to the formulation of contrastive statements on more than one level of linguistic analysis, thus stressing again the linguistic potential of dictionary text, so often insufficiently recognized.