

## **Concordancing As a Method in Contrastive Analysis**

**0.0.** In contrastive linguistic analysis, as suggested by the very term, one analyses language material by contrasting it. Thus, we are dealing here with three basic factors, each requiring specific methodology in approaching it:

- a) what language material to select and how to select it;
- b) how to contrast the selected language material;
- c) how to analyse the contrasted language material.

Answering any of these questions would probably result in separate papers on the subject. Instead, the present writer would like to emphasise one point common to each of these three factors — the large volume of the language material contrasted and analysed.

Contrasting two languages linguistically means contrasting them as two systems in order to establish similarities and divergences of patterns. For a pattern to emerge clearly and reliably, it must be based on a fairly large sample of language material. The sample itself has to be organised in an easy-to-survey manner, with items under research arranged preferably in alphabetic order.

**1.0.** The obvious method, ensuring all these requirements, is concordancing by data-processing equipment.

**1.1.** The case for data-processing equipment is clear since we have to do here with a large mass of text that we want manipulated speedily, with total permutational coverage, and as free from human error as possible. All these requirements are met satisfactorily only by data-processing equipment.

**1.2.** A concordance, as it is known, is a special arrangement of a text, presenting it (usually through alphabetising) in the form of lists built around an unbroken sequence of identical textual segments.

1.21. To illustrate this, here is a short section of a simulated computer-produced KWIC (key-word-in-context) concordance of the preposition *to* as it would occur in a sample of text from "The Listener":

220308670005 <sup>1</sup>	OST COUNTRIES IN THE WEST. TO US THE MARXIST FORECAST OF
220308670007	SION RELATED ONLY REMOTELY TO THE REALITIES OF OUR DAILY
220308670050	E ME RELUCTANTLY RECEPTIVE TO SOME OF THE REVOLUTIONARY
220308670052	• IT SEEMED TOO HARD A NUT TO CRACK, AND I WAS NOT REALL
220308670054	F A PHILOSOPHICAL APPROACH TO ART. I WAS THEREFORE INTER
220308670058	Y FROM DAS KAPITAL I TRIED TO GRASP THESE FROM MARX'S AN
220308670061	EORIES ALWAYS POINTED BACK TO THE SOCIO-ECONOMIC REALITI
220308670065	ING IN ONE OF HIS PREFACES TO DAS KAPITAL THAT SCIENCE K
220308670076	LECTICAL SUBTLETIES SEEMED TO ME A TRIFLE OVER-ELABORATE
220308670084	ERE. HIS EXPOSITION SEEMED TO ME TOO SLOW AND LEISURELY
220308670087	ONE LIKE MYSELF. IMPATIENT TO UNDERSTAND THE WORLD AND T
220308670088	O UNDERSTAND THE WORLD AND TO CHANGE IT QUICKLY. I WAS R
220308670089	IT QUICKLY. I WAS RELIEVED TO HEAR THAT IGNACY DASZYNSKI
220308670090	POSITION. I FELT THE NEED TO RE-EXAMINE MY OWN POLITICA
220308670122	ISM AND MARXISM. I DECIDED TO TAKE NOTHING FOR GRANTED.
220308670124	QUE OF CAPITALISM STOOD UP TO THE EVENTS OF OUR TIME? TH
220308670128	BLED ME. I MADE UP MY MIND TO PLOUGH THROUGH THE WHOLE O
220308670130	CTRINES. I WAS DETERMINED TO SCRUTINISE THIS WHOLE INTE
220308670135	ALLY, KEEPING MY EYES OPEN TO ITS POSSIBLE FLAWS AND CRA
220308670137	E VAST ECONOMIC LITERATURE TO WHICH MARX REFERRED; STUDI
220308670149	I RETURNED AGAIN AND AGAIN TO DAS KAPITAL AND WAS EVER A
220308670165	THAT MARX, I FELT, OPENED TO ME. NO OTHER WORK HAS EVER
220308670173	CING THAN I HAD THOUGHT IT TO BE. I SAW WHERE IN THE OPE
220308670181	AT THEORY, HOWEVER, FAILED TO SATISFY ME - I COULD NOT A
220308670184	CCEPT IT AS AN ALTERNATIVE TO MARX'S CONCEPTIONS OF VALU
220308670185	NG HIM ALL THE WAY THROUGH TO HIS CONCLUSIONS.
220308670189	F CONCENTRATION OF CAPITAL TO DESCRIBE THE "HISTORICAL T
220308670216	Y OF ACCUMULATION" LEADING TO THE EXPROPRIATION OF MANY
220308670217	N, HE DID IT ONLY IN ORDER TO PROVE THAT THAT COMPETITIO
220308670222	

1.22. The above section has a limited context of only 56 spaces (26 on either side, plus TO with its two blank spaces). The same material, however, could easily be printed out by a computer in a full-width display line (129 characters), or with, say, a three-line context, the middle line containing the item analysed (TO in our case). Here is a simulation of such a computer output for the first five occurrences of TO from the section already presented.<sup>3</sup>

<sup>1</sup> This location code should be read as follows: 22-nd sample; of August 3, 1967; line 5. The sample represents the larger part of the article "Discovering *Das Kapital*" by Isaac Deutscher (pp. 139—141).

<sup>2</sup> The indentation indicates the beginning of paragraph. Further on, in line 0189, we have another indentation, now signalling the end of paragraph.

<sup>3</sup> As it would be fairly difficult to print TO always in the same position (though this would facilitate consulting these lists), overprint techniques with a boldface visual effect may be used to make the item analysed stand out from the surrounding text.

- 220308670005      NT FROM THOSE PREVAILING IN MOST COUNTRIES IN THE WEST. TO US THE MARXIST FORECAST OF THE COLLAPSE OF CAPITALISM WAS NOT AN APOCALYPTIC VISION RELATED ONLY REMOTELY
- 220308670007      F THE COLLAPSE OF CAPITALISM WAS NOT AN APOCALYPTIC VISION RELATED ONLY REMOTELY TO THE REALITIES OF OUR DAILY LIFE. THE OLD SOCIAL ORDER WAS CRUMBLING BEFORE OUR VER
- 220308670050      WISH UPBRINGING INCLINED ME AGAINST IT, THE SHAKINESS OF OUR SOCIAL EXISTENCE MADE ME RELUCTANTLY RECEPTIVE TO SOME OF THE REVOLUTIONARY IDEAS IN THE AIR.
- 220308670052      RECEPTIVE TO SOME OF THE REVOLUTIONARY IDEAS IN THE AIR . I TRIED TO READ DAS KAPITAL IN LATE ADOLESCENCE, BUT I DID NOT PERSEVERE. IT SEEMED TOO HARD A NUT TO CRACK.
- 220308670054      LATE ADOLESCENCE, BUT I DID NOT PERSEVERE. IT SEEMED TOO HARD A NUT TO CRACK. AND I WAS NOT REALLY INTERESTED IN POLITICAL ECONOMY. I HAD PRECOCIOUSLY STARTED OUT A

1.23. Of course, computers can be programmed to print out lists on a variety of contexts, and the two short lists presented are simply intended as an illustration of the computer output directly useful in linguistic research, notably in contrastive analysis research. Obviously, in contrastive analyses we shall need computer lists (concordances), offering a direct, line-by-line comparison of corresponding texts (original and its translation) from two, or more, languages being contrasted. The following is an illustration of what a list might look like, contrasting an English original (first 10 sentences of the "Listener" sample) and its Croatian translation (supplied by this writer):<sup>4</sup>

---

<sup>4</sup> The list which follows probably shows an overly regular lineal ("in face") parallelism of the two texts. In contrasting limited-context segments of textual material from the target and the source language we are faced with the technical difficulty of getting the target-language equivalent into lineal correspondence with the item being analysed in the source language. More simply, the translation line that the computer prints out as spatially (textually) parallel to the line from the original, containing a certain linguistic item (TO in our case), need not contain the translation of the item at all (an "out of face" display), as this may have been realised in a removed word-order position — and would have been present in the preceeding, or the following, line. This danger of missing the translation equivalent altogether can be completely avoided (for practical purposes, that is) by contrasting items only within a full-sentence context. It is very rare, namely, that an item is translated beyond the sentence limit in the source-language text.

- 220308670005 OST COUNTRIES IN THE WEST. TO US THE MARXIST FORECAST OF U VECINI ZEMALJA NA ZAPADU. ZA NAS MARKSISTICKO PREDSKAZ
- 220308670007 SION RELATED ONLY REMOTELY TO THE REALITIES OF OUR DAILY VRLO INDIREKTNO POVEZANA SA STVARNOSCU NASEG SVAKODNEVN
- 220308670050 E ME RELUCTANTLY RECEPTIVE TO SOME OF THE REVOLUTIONARY ILA ME JE DONEKLE PRIJEMLJIVIM ZA NEKE OD REVOLUCIONARNI
- 220308670052 I TRIED TO READ DAS KAPITAL IN LATE A POKUSAO SAM CITATI DAS KAPITAL POTKRA
- 220308670054 . IT SEEMED TOO HARD A NUT TO CRACK, AND I WAS NOT REALL UCINIO MI SE DA JE PRETVRD ORAH, A POLITICKA EKONOMIJA N
- 220308670058 F A PHILOSOPHICAL APPROACH TO ART. I WAS THEREFORE INTER AZIO SAM FILOZOFSKI PRISTUP UMJETNOSTI. ZATO ME JE PRVEN
- 220308670061 Y FROM DAS KAPITAL I TRIED TO GRASP THESE FROM MARX'S AN VSI DAS KAPITAL, POKUSAO SAM IZVUCI TE CRTE IZ MANJE VAZ
- 220308670065 EORIES ALWAYS POINTED BACK TO THE SOCIO-ECONOMIC REALITI VIJEK SU UPUČIVALE NATRAG NA DRUSTVENO-EKONOMSKU STVARNO
- 220308670076 ING IN ONE OF HIS PREFACES TO DAS KAPITAL THAT SCIENCE K DNOM OD NJEGOVIH PREGGOVORA KAPITALU DA ZA NAUKU NE POST
- 220308670084 LECTICAL SUBTLETIES SEEMED TO ME A TRIFLE OVER-ELABORATE GOVE DIJALEKTICKE FINESE CINILE POMALO PRETJERANE I STAR

2.0. And now, after this introductory digression, let us try to formulate the object of this paper. It is to attempt to list and analyse several possible modes in which the concordancing method can be applied to contrastive analysis. We must, however, point out immediately that, no matter how many modes we may list, the survey will probably not be complete, because it is only under conditions of actual research, responding to the challenge of ever-changing requirements, that we devise new applications of the method.

3.0. If we visualise a sample of, say, 10,000 English sentences (normally sufficient for syntactic research) we shall, in contrastive analysis, be faced by 20,000 sentences, or a total of 40,000 words of running text.<sup>5</sup> Transferred onto punched cards — even without spacing the words at regular intervals to facilitate sorting — this would result in a superficially unwieldy mass of 50,000 punched cards.<sup>6</sup> “Superficially unwieldy” has been

<sup>5</sup> The approximate average of 20 words per sentence is based on the writer's personal counts of five 1000-word samples of factual (medical) text.

<sup>6</sup> This total was arrived at by estimating a card's capacity to be eight words (seven columns as the average length of a word, plus two columns for its code, plus five columns for the card, or sentence, code). If, however, we were to space the words, giving 13 columns each (the length of which would cover over 90 per cent of all English words), plus two columns for the code, we should have a capacity of only five words per card. This would, in its turn, mean a total of 80,000 punched cards.

used, because any language material transferred onto punched cards, and from them to the magnetic tape, becomes in fact a very flexible mass, lending itself superbly to listing and tabular presentation by computer.

**3.1.** Leaving aside tabular presentation, more useful for statistical purposes, simple alphabetic listing (*ordinary*, or *forward*, concordances, the commonest form of concordancing) will prove very valuable in the contrastive analysis of our sample, or of any other large sample.

**3.11.** This is obvious with function words and determiners which can be singled out and listed by nothing other than simple alphabetic sorting, with no codes or identification procedures for formal signals necessary.

**3.12.** For instance, the English half of our sample can be expected to yield a total of about 3,750 occurrences of the preposition *to*.<sup>7</sup> This is more than enough for the principal patterns to emerge clearly. If this is done by hand (sorting slips) the researcher might be strongly tempted to stop at, say, 500 (or 750, or 1,000) occurrences.

These 3,750 occurrences would be normally presented in the form of an unbroken computer-output list of as many lines, similar to the first text illustrated under 1.2. Although we can easily observe the difference in context even in such long lists, the volume of this material (75 large printout sheets) promptly calls for further immediate-context sorting, resulting in sublists, representing collocations (patterns), such as:

	<i>to</i> + Vb	(... <i>to copy</i> ...)
right-hand	<i>to</i> + <i>have</i> + Vbed	(... <i>to have gone</i> ...)
context only	<i>to</i> + <i>some</i> + N	(... <i>to some extent</i> ...)
	etc.	
	Vbs + <i>to</i>	(... <i>tries to</i> ...)
left-hand	<i>in</i> + N + <i>to</i>	(... <i>in order to</i> ...)
context only	<i>the</i> + Adj + Ns + <i>to</i>	(... <i>the frantic attempts</i>
	etc.	<i>to</i> ...)
	Vbs + <i>to</i> + Vb	(... <i>tries to copy</i> ...)
bilateral	<i>in</i> + N + <i>to</i> + Vb	(... <i>in order to copy</i> ...)
context	<i>the</i> + Adj + Ns + <i>to</i> + V	(... <i>the frantic attempts</i>
	etc.	<i>to copy</i> ...)

---

<sup>7</sup> This estimate is based on the writer's analysis of a 32,000-word sample of a textbook in veterinary surgery.

As can be seen, we are dealing here with syntactic patterning systematically expandable at will, until the whole range of linguistically possible permutations has been exhausted, i. e. checked for the occurrence in, or the absence from, our sample.

**3.13.** Let us, for the sake of illustration, take the first list from 1.2 and break it into sublists that would result if this subsorting were done according to the immediate right-hand context:

<i>us</i>	(0005)
<i>the</i>	(0007, 0065, 0128, 0217)
<i>some</i>	(0050)
Vb	(0052, 0054, 0061, 0088, 0089, 0090, 0122, 0124, 0130, 0135, 0184, 0216, 0222)
N	(0058)
non-English	(0076, 0165)
<i>me</i>	(0084, 0087, 0173)
<i>its</i>	(0137)
<i>which</i>	(0149)
<i>be</i>	(0181)
N's	(0185)
<i>his</i>	(0189)

The sublists which we get even in such a restricted material are already indicative of what we should have to expect from the whole *to*-sample of 3,750 lines.

**3.14.** The clearly emerging high-frequency pattern *to* + Vb (13 occurrences in 30 lines) should obviously be the first to be tested for possible contrastive-analysis value. Before contrasting it, however, it would be wise<sup>8</sup> to break it down further, by sorting it according to immediate left-hand context as well. What we now obtain is:

Vbed + <i>to</i> + Vb	(0052, 0061, 0090, 0124, 0135, 0184)
N + <i>to</i> + Vb	(0054, 0122, 0130, 0216, 0222)
Adj + <i>to</i> + Vb	(0088)
<i>and</i> + Vb	(0089)

---

<sup>8</sup> One should always keep in mind that, dealing with the entire sample of 3,750 estimated occurrences, we should at this stage of subsorting have to manipulate not tens but hundreds of examples.

3.15. The two outstanding patterns now are: Vbed + to + Vb and N + to + Vb. When these are contrasted with their Croatian equivalents they establish the following correspondence:

English	Croatian
Vbed + to + Vb	$\left\{ \begin{array}{l} \text{Perfect + Infinitive / } da + \text{ Present}^9 \\ \text{(0052, 0061, 0124, 0135)} \\ \text{Perfect + } kada + \text{ Perfect (0090)} \\ \text{Negative + Perfect (0184)} \end{array} \right.$
N + to + Vb	$\left\{ \begin{array}{l} \text{idiom-integrated (0054)} \\ \text{N + } da + \text{ Present (0122, 0216)} \\ \text{idiom-integrated + } da + \text{ Present} \\ \text{(0130, 0222)} \end{array} \right.$

We proceed in this manner until we have established and analysed the full range of English-Croatian pattern correspondence contained in the sample researched.

4.0. Another form of alphabetic concordancing — the *reverse*, or *backward*, concordance — can be a very useful method in contrastive analysis.

4.1. It is easy to envisage linguistic advantages offered by this approach. If word identification proceeds from right to left morphemes are taken in before anything else. In a highly inflected language like Croatian this is particularly important. Let us now consider a specific morphological area in the contrastive analysis of English and Croatian, and visualise the steps in concordancing which it would be useful to take towards solving specific problems in the area.<sup>10</sup>

---

<sup>9</sup> Now, of course, we are faced by the dilemma of whether to list all Croatian alternatives (as was easily possible here since the alternative operates for all the four occurrences), or whether to settle for the most frequent equivalent in the cases where the alternative is not general — ignoring the possible translation with the Present Conditional in 0216 and 0222, and with the Infinitive in 0130.

<sup>10</sup> It is very important here — as indeed it is for this paper as a whole — to stress the formulation “towards solving”. A full and definitive contrastive description of the manner in which English and Croatian each express a specific language relationship or function is the task of man, the linguist analyser. In producing concordances machines merely aid (“man-machine partnership”) the human researcher by organising, presenting and handling the vast volume of linguistic data that would otherwise be very much beyond his capacity.

4.2. One of the principal linguistic tasks here would be establishing the patterns used in English to convey the language relationships that are expressed in Croatian by the case system. We shall, for practical reasons, narrow down the problem to a very specific contrast: the Croatian Instrumental case and the corresponding patterns in English.

4.3. We now, obviously, start from Croatian, and our first step is to compile a list of the Croatian Instrumental morphemes, covering the singular, plural and dual, taking care of all the three genders, and encompassing the nominal, adjectival and pronominal declensions.

4.31. Our taped sample, already described in 3.0. (or any other language material we have selected) is sorted backwards. The resulting reverse KWIC concordance, grouping words that end in the same letter, or combination of letters, is a very convenient linguistic tool which makes it possible for us to spot quickly and effortlessly all possible Croatian Instrumentals within our sample.

4.32. Naturally, these lists would have to be sight-checked for all superficially identical non-Instrumental morphemes (or any accidental combinations of letters), such as, for instance, the plural (masculine and neuter) Datives and Locatives in *-ima*, as opposed to the plural (masculine and neuter) Instrumentals also in *-ima*; or words like *svima*, *plima*, etc.

This may seem like an excessive amount of work which can be avoided by having the cards coded before everything else. Coding, however, is a slow process, and we might have to start analysing before it is completed. We may safely assume that sight-checking and eliminating (by crossing through) from the concordance-lists takes about two or three times less than coding.

Even if we are lucky enough to have punched cards with a two-column alphanumerical code with its capacity of recording 169 different elements, we might still find it inadequate to cover the full inflexional inventory of Croatian, and we might have to do quite extensive additional sight-checking.

4.4. At any rate, the Instrumental concordances that we should get might look something like these 16 occurrences obtained within the first 80 lines of our sample from *The Listener*:

220308670007 VRLO INDIRECTNO POVEZANA SA STVARNOSCU NASEG SVAKODNEVN  
SION RELATED ONLY REMOTELY TO THE REALITIES OF OUR DAILY

220308670009 OREDAK SE RUSIO PRED NASIM OCIMA. TO JE BILA NAJVAZNIJA  
AS CRUMBLING BEFORE OUR VERY EYES. THIS WAS THE OVERWHEL

220308670011 EPRESTANO POTRESALA MOJIM VLASTITIM DJETINJSTVOM I PUBER  
Y OWN CHILDHOOD AND ADOLESCENCE WAS SHAKEN BY IT AGAIN A

220308670012 MOJIM VLASTITIM DJETINJSTVOM I PUBERTETOM. ODRASTAO SAM  
ADOLESCENCE WAS SHAKEN BY IT AGAIN AND AGAIN. I GREW UP

220308670027 PRENIJELA SE NA MENE NAPETOST KOJOM SU STARIJI PRATILI V  
BSORBED FROM THE ADULTS THE TENSE MOOD IN WHICH THEY WAT

220308670029 ENE ARMIIJE NA VARSAVU. GODINAMA SMO GOTOVO NEPRESTANO ZI  
ADVANCE ON WARSAW. FOR YEARS WE LIVED ALMOST CONSTANTLY

220308670031 GRADJANSKOG RATA, MEDJU GALOPIRAJUCOM INFLACIJOM, MASOV  
BRINK OF CIVIL WAR, AMID GALLOPING INFLATION, MASS UNEMP

220308670031 LOPIRAJUCOM INFLACIJOM, MASOVNOM NEZAPOSLENOSCU, POGROMI  
LLOPING INFLATION, MASS UNEMPLOYMENT, POGROMS, ABORTIVE

220308670031 ASOVNOM NEZAPOSLENOSCU, POGROMIMA, NEUSPJELIM REVOLUCIJA  
NEMPLOYMENT, POGROMS, ABORTIVE REVOLUTIONS AND FUTILE CO

220308670032 OSCU, POGROMIMA, NEUSPJELIM REVOLUCIJAMA I JALOVIM KONTR  
MS, ABORTIVE REVOLUTIONS AND FUTILE COUNTER-REVOLUTIONS.

220308670032 JELIM REVOLUCIJAMA I JALOVIM KONTRAREVOLUCIJAMA. ALI I P  
ONS AND FUTILE COUNTER-REVOLUTIONS. BUT EVEN BEFORE THES

220308670039 NISTA JOS UVIJEK SMATRALI KAPITAL "BIBLIJOM RADNICKE KLA  
MUNISTS STILL CONSIDERED DAS KAPITAL AS THE "BIBLE OF TH

220308670075 TICKU BORBU. CAK SAM S PRIMJESOM ZLOVOLJE PROCITAO MARKS  
OLITICAL STRUGGLE. IT WAS EVEN WITH A HINT OF IRRITATION

220308670080 UKLANJATI TEGOBAMA USPONA NJENIN STRMIM STAZAMA". PITAO  
LL NOT DODGE THE TOIL OF CLIMBING UP ITS STEEP PATHWAYS"

4.5. The patterns which emerge from the above list, after we break it down through further sorting according to the equivalent English patterns, are:

English	Croatian
1. Preposition	Preposition + noun in the Instrumental case
<i>to</i>	<i>sa</i> (0007)
<i>before</i>	<i>pred</i> (0009)
<i>amid</i>	<i>među</i> (0031, 0031, 0031, 0032, 0032) <sup>12</sup>
<i>with</i> <sup>13</sup>	<i>s</i> (0075)
2. Preposition	No preposition + noun etc. in the Instrumental case
<i>in (which)</i>	<i>kojom</i> (0026)
<i>for (years)</i>	<i>godinama</i> (0029)
<i>up (... pathways)</i>	<i>stazama</i> (0080)
3. Verb	Verb + noun etc. in the Instrumental case
<i>was shaken</i>	<i>potresala ... djetinjstvom</i> (0011)
<i>(was shaken)</i>	<i>(potresala) ... pubertetom</i> (0012)
<i>consider (as)</i>	<i>smatrali ... biblijom</i> (0039)
<i>made</i>	<i>učinila ... prijemljivim</i> (0050)
<i>found</i>	<i>činile ... uvjerljivim</i> (0071)

4.6. Subsequent sorting routines proceed much along the same lines, with modifications in them introduced as required by the linguistic situation presented by each new breakdown of the language material.

5.0. After having, hopefully, illustrated some modes in which concordancing can be a useful aid in two important language areas — syntax (in 3. and 4.) and morphology (in 4.) — let us round off this survey with an attempt at describing some possibilities of concordancing in contrastive analysis in the areas of word-formation and lexis.

5.1. The lexicographical possibilities of simple alphabetic concordances, lineally contrasting two texts (the source and the target language; cf. 1.23.), should by now be self-evident.

<sup>12</sup> The last four of these examples are in fact elliptical, the preposition being suppressed to avoid repetition.

<sup>13</sup> It is interesting to note that this sublist (admittedly very limited — covering only 80 short lines of the sample text) contains just one example of *with* = (*sa* +) Instrumental, though one would intuitively anticipate the preponderance of this particular correspondence. Even this single example has an alternative adverbial translation (... *pomalo zlovoljno* ...), preferred by grammarians.

5.11. An average 300-page novel, containing some 130,000 running words, can be expected to produce some 25—30,000 graphically different forms.<sup>14</sup> Now if a contrastive alphabetic concordance is made of this text, with each graphically different textual segment between two blank spaces, or a punctuation mark, placed within a one-line (or one-sentence) context, we shall be provided with an extensive body of text eminently suitable for contrastive lexical research.

5.2. The most obvious use of such a concordance would be dictionary expansion. A concordance sublist of some twenty or thirty occurrences of words like *understatement*, *offset*, *challenging*, *wistful*, etc. (to mention only a few irritants well-known to every translator from English to Croatian) would be sure to produce several valuable equivalents still unrecorded in the existing English-Croatian dictionaries.

5.21. Even in our very brief sample from *The Listener* the simplest contrastive alphabetic concordance would reveal, within its first 15 lines, the following two Croatian equivalents thus far absent from all existing English-Croatian dictionaries:

*crumble* = *rušiti se* (0008; *The old order was crumbling before our very eyes.*)

*adolescence* = *pubertet* (0012; *My own childhood and adolescence was shaken by it again and again.*)

5.22. The same two lexical pairs of equivalence would equally be missing from the existing Croatian-English dictionaries. A contrastive alphabetic concordance, starting from Croatian, would produce an even higher number of acceptable (often, indeed, very felicitous) equivalents not recorded in any Croatian-English dictionary. This concordance would also make it possible to spot a number of highly acceptable English equivalents of more than one word (notably in the area of abstract notions), which would for formal reasons be excluded in a concordancing English-to-Croatian approach. Let us quote just one of these equivalent pairs (within the first 30 lines of our very restricted sample):

*napetost* = *tense mood* (0027; *At 13 I absorbed from the adults the tense mood in which they watched the news of the Red Army's advance on Warsaw.*  
... *prenijela se na mene napetost  
kojom su stariji pratili . . .*)

---

<sup>14</sup> Which is not the number of different "words" (lexical units, dictionary entries) since oblique forms are included, and homographs are not differentiated.

**5.23.** Contrastive alphabetic concordances can reveal a number of other similarly useful lexicographic patterns. Of these, we should perhaps single out those based on microcontexts and producing:

a) idioms and phrases based on keywords;

b) set terms

(e. g. *trade union local* = *sindikalna podružnica* [0043]);

c) useful collocations and clichés (e. g. *galloping inflation* [0031], *mass unemployment* [0031])  
etc. etc.

**5.24.** Finally, marginal lexicographic forms, such as frequency dictionaries, dictionaries of synonyms and antonyms, rhyming dictionaries, etc, also offer a wide scope for the application of alphabetic concordancing methods, some of which are directly relevant to contrastive studies.

**6.0.** And thus we reach the last linguistic area in this survey of concordancing methods in contrastive analysis — that of word-formation.

**6.1.** Dealing, like lexicography, with words at the lexical level, word-formation, whose very *raison d'être* is semantic patterning in the composition of words, lends itself equally readily to contrastive analysis. Contrasting English and Croatian in the word-formation area means, then, establishing divergences and parallels among the discernible patterns of word-building in the two languages.

**6.2.** A very simple illustration:

A punched-card file is compiled (using existing Croatian-English dictionaries) of all words beginning in *ne-* with their English equivalents (i. e. with the full English side of the dictionary entry, minus the phraseological section). Next step in sorting now proceeds from English, and the cards are subdivided into groups according to the English equivalents. This results in simple alphabetic lists which are, however, quite adequate for providing us with visually ascertainable typical alphabetical word beginnings in English (in this case *un-*, *in-*, *non-*, etc.), corresponding to the Croatian *ne-*.

**6.21.** Of course, lists like these, obtained from dictionary material are based on specially (lexicographically) organised, non-natural material, and their quantitative relationships would have no bearing whatsoever on the actual distribution of these prefixes in English.

To obtain this, i. e. to establish which are the truly commonest negative prefixes in English (and as such the closest contrastive equivalents to the undoubtedly predominant Croatian *ne-*), we should have to take recourse to alphabetic concordancing of large samples of text. Subsequent sorting would produce readily (visually) comparable sublists of occurrences of each prefix in question.

6.3. What is more, these sublists from alphabetic concordances would enable us to expand even the prefix inventory obtained from the Croatian-English dictionaries. Since dictionaries tend to be insufficiently based on the exploration of actual texts, they would be sure to miss some of the very valuable solutions (i. e. contrastive equivalents) produced by translators who, by the nature of their work, are in some respects among the truest contrastive analysts. However, translators, in their turn, respond to isolated translation (i. e. contrastive) challenges as they occur, and tend not to be aware of any patterning in their responses, or, indeed, of the general applicability of some of what they would probably consider mere ad hoc solutions.

6.31. An example springs to mind: The normal gerund ending (thus, a very frequent one) in Croatian is *-nje*. Coupled with the already discussed prefix *ne-* and with a verbal base (in this order: *ne* + Vb base + *nje*), it represents a word-formation pattern of wide distribution in Croatian (e. g. *nepolaženje*, *nepavanje*, *nevršenje*, *nepokoravanje*, etc.). If, however, a Croatian-English dictionary is consulted it reveals no significant patterning among the English equivalents, and is, in addition, curiously incomplete and burdened by a great deal of lexical deadwood in the form of improbable or clumsy English equivalents.

But if we consult an alphabetic contrastive concordance, sorting first for initial *ne-* and, as the next step, for final *-nje* (and its oblique forms), we should be sure to come across repeated cases of the corresponding pattern in Croatian. Its formula (though at the level of phrase) is: *failure* + Vb in the Infinitive.

Thus:

Croatian	English
<i>nepriznavanje</i>	<i>failure to admit</i>
<i>neprestajanje</i>	<i>failure to stop</i>
<i>neširenje</i>	<i>failure to spread (expand, disseminate)</i>
<i>nedjelovanje</i>	<i>failure to act</i>
<i>nekomuniciranje</i>	<i>failure to communicate</i>
<i>neprisustvovanje</i>	<i>failure to attend</i>
etc.	etc.

**6.32.** Obviously, there are limits to this pattern. We cannot ignore that: 1) there are quite a few Latin-base words (of the formula *non...al*, *non...ent*, *in...ent*, etc.; or *non* + noun / verb [e.g. *non-transfer*]) which operate opposite to the *ne* + Vb base + *nje* formula in Croatian; 2) the pattern is not always possible in English as corresponding to the Croatian formula (e.g. *nepostojanje* is not *failure to exist* but *non-existence* or *absence*); 3) there is overlap within English itself with the formula *absence of* + noun (describing verbal action), e.g. *nespavanje* = *absence of sleep*; 4) semantic (and perhaps the transitive/intransitive) subdivision of the Croatian words based on the formula should be taken into account.

Nevertheless, a pattern has been established as operating in clear contrastive correspondence between the English and the Croatian word-formation systems; and concordancing as a method has been instrumental here. In each of the necessary additional analyses, tentatively surveyed in the four stages outlined in the preceding chapter, alphabetic listing, reverse alphabetising and similar concordancing procedures are of self-evident utility. Describing them, however, would take us beyond the scope of this paper.