

The Shape of the Dictionary
For Mechanical Translation Purposes

by
L. Zgusta

Linguistics Research Center
The University of Texas at Austin
and
Department of Linguistics
The University of Illinois

THE SHAPE OF THE DICTIONARY
FOR MECHANICAL TRANSLATION PURPOSES.¹

L. Zgusta

A dictionary of the type we have in mind here should contain the lexical units of the source language, selected according to the needs of the type of texts to be translated. Lexical equivalents of the target language should be coordinated with these lexical units in such a way that the choice is as precise and as automatic as possible. Great difficulties are caused in this task not only by the polysemy and homonymy of the lexical units of the source language, but also by the fact that the equivalents usually cannot be coordinated in a one-to-one way. We call "lexical equivalent" a lexical unit of the target language which has the same lexical meaning as the respective lexical unit of the source language; that means the equivalent should have the same polysemy, the same stylistic value, etc., as the lexical unit of the source language. However, this is seldom the case, and, consequently, more than one equivalent is often needed to cover the lexical meaning of the source word. We should, then, make the distinction between absolute equivalents, which comply with the definitional requirement of a one-to-one correlation in lexical meaning, and partial equivalents; but general usage allows us to

¹ Work on this article was performed at the Linguistics Research Center, University of Texas at Austin, Texas, under Rome Air Development Center contract. The paper profited from a discussion of its contents with M. Kay and R. Stachowitz.

speak about equivalents when it is usually the partial ones we have in mind.

The present article is not primarily concerned with the problem of (partial) equivalents, their choice, their mutual disambiguation and the delimitation of their applicability in an entry.² This article concentrates on problems of choice from among more than one (partial) equivalent within the entry of a lexical unit of the source language. The point of view taken here is that, on the one hand, the more we can rely on simple formal indications of the source language the better, but that, on the other hand, such simple formal indications do not always exist; and that one of the cardinal difficulties with which we have to cope is that the selection of a suitable (partial) equivalent is to be made by an agent which is by far less imaginative than the human mind.

Semantic difference in the source language (and, therefore, the necessity of a certain selection among the partial equivalents) is frequently indicated by some difference in form. The situation is rather simple if the difference of form is easy to detect. It is easy to change an entry of the type

German *in* → English *in, into*

into the following shape

German *in* + Dat. → English *in*

in + Acc. → English *into*

² See "Equivalents and Explanations in Bilingual Dictionaries", to be published in

Since we envisage, for the moment, only basic translational needs, this form of the entry should suffice to guarantee a good selection of the equivalent in sentences like German *-in dem Wald gehen* - English *to walk in the wood*, and German *in den Wald gehen* - English *to walk into the wood*, that is, given the ability to recognize which German substantive is governed by *in* and whether it is dative or accusative.

The example just discussed is one of the simplest ones. It can be said that the recognition of semantic difference and the choice of the equivalent entailed by it are not difficult if the semantic difference is indicated by a clear morphological difference.

The formal distinction, however, does not necessarily have to be a morphological one; the main thing is that the distinction should be clear in itself and non-ambiguous. For instance, it should be easy to discern the polysemy of German *handgreiflich*, because in one of its senses, it is used exclusively with the forms of *werden*: *handgreiflich werden* "to use physical force". In its other sense, it is used with *machen*, *sein* and a few other verbs: *handgreiflich machen* "to make available", *handgreiflich sein* "to be available".

Perhaps more complicated is the following type of case. If we simplify the multiple meaning of German *ableiten*, we can construct an entry of the following type:

German *ableiten etwas* → English *to lead away*

ableiten etwas von etwas → *to derive*

(As in German *den Strom ableiten* - English *to lead the current away*, German *die Adjektive aus den Substantiven ableiten*, English *to derive adjectives from substantives*.)

It would seem that it should not be too difficult to distinguish the two types of reactions quite automatically, and make the choice accordingly. The next example will, however, be more complicated.

The simplest way to construct the entry of German beraten seems to be

German <i>beraten jemanden</i>	→	Eng. <i>to advise</i>
<i>beraten etwas</i>	→	<i>to deliberate (upon)</i>
<i>beraten ueber etwas</i>	→	<i>to deliberate (upon)</i>

The last two German reactions are different in their grammatical form, but there is no semantic difference. On the other hand, there is no grammatical difference between the two first reactions, but there is a semantic difference which entails a different choice of English equivalent. The abstract expression of the two reactions in the lexicographic entry (*jemanden* : *etwas*) is rather simple, and no human user of a dictionary could have difficulty with it. Still, for the purpose of automatic recognition and choice, the presence of this entry in the dictionary entails the necessity of indicating in the lemma of each substantive whether it belongs to the category "*jemand*" or "*etwas*". This should not be too difficult a task; let us, however, discuss yet another type of situation.

The entry of German *abhalten* can be constructed in the following way:

German *abhalten* (1) *jemanden von sich* → *to hold off*

(2) *jemanden von etwas* → *to hinder,*
prevent

(3) *etwas* → (a) *to keep out*
→ (b) *to hold*

We see that within one reaction, (3), there are two choices (a), (b) which are semantically governed: (a) is chosen if the object (represented by *etwas*) is, e.g., *Wasser, Naesse, Zugluft, Regen*; (b) is chosen if the object is, e.g., *Sitzung, Wahlen, Gericht*, etc.

Another example of this type is German *auslegen*. One of its reactions (the most frequent one) is *auslegen etwas*. The respective part of the entry would have to have a form similar to the following one:

auslegen (1) *etwas* (a) [*im Ladenfenster*] → *to display*
(b) [*Geld*] → *to pay provisionally*
(c) [*Texte*] → *to interpret*

In a case like this, the really important indication is the one contained in brackets. And as every lexicographer knows, to construct these restrictive (or semantic) glosses (as they are frequently called in lexicographic theory) belongs among the most difficult tasks because it is hard to find the real limits of the restriction. This is, however, a purely lexicographic task which every good lexicographer is accustomed to coping with. It is not without significance that in the compilation of a dictionary of a living language, it is nearly always native informants who are used for this task. But in the

situation envisaged in this article we try to count with an automatic choice from among the equivalents, and this causes much trouble. The reason is that every human user of a dictionary will immediately understand that an indication like [*im Ladenenster*] is simply an example since goods can be displayed also on stands within a shop or in the market, and so on. Not only that; the human user will also understand that the restrictive gloss [*im Ladenfenster*] is, at the same time, the representative of a certain type of situation, since one can speak about somebody displaying his goods without mentioning where and how, and choice (a) is then entailed. Therefore, this part of the entry could also have the following form:

auslegen (1) *etwas* (a) [*Waren*] → *to display*

This restrictive gloss would have other difficulties of its own. We mention it to show that restrictive glosses have to be chosen from among various possibilities inherent in the facts of language.

In the same way, [*Geld*] is both an example and the representative of a class of synonyms, near-synonyms, and semantically related words (*eine Summe auslegen*). In (c) , [*Texte*] would seem to be simply the hierarchically higher notion (*Oberbegriff*) comprising singularia like Bible, Homer, 6th Amendment to the Constitution, etc., or any text(s); but in reality, it must be understood as a representative of other expressions, too. There is no need to go as far as poetical language to have a sentence like *Falls die Datenverarbeitungsmaschine den gestrigen Verkauf von Papieren auf der New Yorker Boerse falsch auslegt, dann . . .*

The difficulty of this problem is obvious. One of the easiest answers would be that we should increase the number of concrete examples quoted in the restrictive glosses. For instance, one could imagine the following form for the entry quoted above:

abhaltun (3) *etwas* (a) [*Wasser, Naesse, Fluessigkeiten, Regen, Hagel, Wind, Zugluft*] →
to keep out

The increase in the number of concrete examples in the restrictive glosses would be an enormous gain; but we should count with dozens and perhaps hundreds of them in one gloss. It does not seem to me, however, that the more or less exhaustive enumeration of examples could be a real solution. Let us discuss the following example. That part of the entry of German *verjuengen* which is concerned with technical terminology could have the following form:

verjuengen (1) *etwas* (a) [*Maßstab*] → *to reduce*
(b) [*in biology*]³ *to rejuvenate*

The restrictive gloss pertinent to (a) could be expanded by an enumeration of examples. I cannot, however, see that choice (b) could be governed by the indication of concrete examples. First, because the area of objects of rejuvenation, attempted or real, is rather vast; still, one can imagine a restrictive gloss with perhaps hundreds of examples, e.g. [*Gewebe, ...Knochen, ...Zellen, ... Greise, ... Reflexfaehigkeit, ... Regenerationsfaehigkeit, ...etc.*]. But the second difficulty seems to be more grave. The area of objects of rejuvenation is not only vast; it is always getting more vast, and the very

3 We do not take into consideration that the theory of lexicography usually distinguishes indications of this type from the restrictive (or semantic) glosses. Indications of this type are usually called labels.

purpose of science is to render it more vast. Consequently, one must take into consideration that after we have established our set of examples in the restrictive gloss, there will be biological texts reporting new investigations, discoveries, etc., concerned with new objects not stated among our examples; which would make a correct choice of the equivalent impossible. And since the main purpose of machine translation is to translate recent reports on new discoveries, etc. quickly, we can conclude that the choice of the equivalent cannot be based on an exhaustive enumeration of contextual examples (understood as key words), lest we block our way to the very goal we are trying to reach.

It seems that what is needed is a classification of all entry-words selected for the future dictionary into classes constituted by the restrictive glosses and the semantic criteria contained in them. For instance, since the correct choice of an equivalent in some entries depends on whether the object is a person or not, this category should be indicated in the lemma of each substantival entry-word; since a correct choice in another entry depends on whether the context is a biological one or not, the pertinent indication should be a part of the lemma of the respective entry-words. This should be done with all the restrictive glosses involved in the corpus of entries. It would require further researches, but it seems that the number of different restrictive glosses could be slightly reduced if they had, when possible, the form of hierarchically higher notions (*Oberbegriffe*), or if they indicated terminological areas (such as "biology", "chemistry", etc.). In this way, though the automatic procedure would certainly not command an abstractive ability of its own, it would possess a rich repertory of coherently constructed criteria for the necessary choices,

applicable not to a broader semantic range of texts but at least to a much larger corpus of them than that on which the original investigations were based.

What has been discussed up to now is certainly no panacea. There will be bases which will resist a generalization. For instance, another rection of German *auslegen* (not mentioned above) is *auslegen* (2) *etwas mit etwas*. In German, contexts characterized by this formal feature are not only clearly differentiated from the contexts of the type *auslegen* (1) *etwas*, but they also form a unified group, with a unified if general meaning. But there is no general equivalent in English, and the choice of the partial ones is governed by the object of the action. Consequently, we have to imagine that this part of the entry could have a form similar to the following one:

auslegen (2) *etwas mit etwas* (a) [*Teppiche*] → *to cover*
with (carpets),
to carpet
(b)[*Zement*] → *to line with*
(cement)
(c)[*Elfenbein*] → *to inlay,*
encrust with
(ivory)

On the other hand, the semantic classification is necessary even in cases where one would not immediately suspect it. We

4 The method proposed here has some similarity to the method using so-called "semantic parameters". (Cf., e.g. Ubin, expression of the parameter Magn in Russian, in: *Mašinnyj perevod i prikladnaja lingvistika*, 11, 1969, p. 60 ff.; Šaljapina, Ways of expressing semantic parameters in English, *ibid.* p. 106 ff.) The difference, however, is that whereas the search for semantic parameters leads to establishing more or less purely notional frameworks and constructions, the present approach tries to remain as close to really occurring contexts as possible.

have stated above that it is relatively easy to find a solution for those cases in which a difference in meaning is indicated by a difference in form, preferably in morphology. The dictionary can make use of such differences. Sometimes, the morphological distinction alone is sufficient to indicate the difference in meaning. For instance, the series of German forms *die, der, der, die Diaet, die, der, den, die Diaeten* can be seen as a normal paradigm of a feminine substantive. There is, however, the semantic difference that the forms of the singular require the English equivalent "diet, regimen", whereas those of the plural require the equivalent "daily allowances". Such a situation is easy to solve. Probably every lexicographer will take *Diaet* "diet" as *singulare tantum*, and *Diaeten* "daily allowances" as another word, a *plurale tantum*; and such a solution is undoubtedly even more practical for an automatic procedure.

But not all cases are as beautifully clear-cut as this. A morphological difference is sometimes of only partial value. For instance, if we try to find an English equivalent for German *Ort*, in its application as a technical term, we can arrive at the following result:

Ort (1) [in geography] → *place*
(2) [in geometry] → *locus*

It is usually maintained that the two are sufficiently differentiated by the fact that *Ort* (1) [geogr.] has the plural *Orte*, whereas *Ort* (2) [geom.] has the plural *Oerter*. This morphological distinction is fully sufficient for the plural; if we had to deal with *pluralia tantum*, this part of the reduced entry could have the form:

Orte → *places*
Oerter → *loci*

Since the singular is not morphologically differentiated, a semantic (that is, contextual) differentiation is necessary.

The same situation can be observed in *die Mutter*, plural *Muetter* "mother"; *die Mutter*, plural *Muttern* "nut": the singular is not morphologically differentiated. On the contrary *das Erkenntnis* "decision, judgment, sentence" and *die Erkenntnis* "comprehension, perception, cognition" are well differentiated in the singular, but since they have identical forms in the plural, *die Erkenntnisse*, they should be semantically differentiated as a juristic, and a psychological and philosophical term, respectively.

Cases like those just discussed are particularly disagreeable if there is a semantic difference only in a small part of a paradigm. Let us discuss an example. We can imagine a strongly reduced form of the entry of German *erledigen* as follows:

erledigen (1) *etwas* → *to finish, arrange, settle*
 (2) *jemanden* → *to dispose of*

This German verb has the normal participle *erledigt* which has the same polysemy: *Das ist erledigt* "That's settled", *Durch die naechste Saeuberung wird er erledigt (werden)* "He will be disposed of by the next purge". This form, however, has other senses of its own, so that an eventual entry could have the following form:

erledigen (1) *etwas* → *to finish, arrange, settle*
 (2) *jemanden* → *to dispose of*
erledigt (1) participle to *erledigen* (1), (2)
 (2) [Person] → *done for, finished*
 (3) [Stelle, Posten] → *vacant*⁶

5 As in: *Nach diesen Strapazen bin ich erledigt* "I'm finished after these labors".

6 As in: *Jede erledigte Stelle* "Each vacant situation".

Cases like this are rather treacherous. Dictionaries are normally built on the principle that the form of the source language in which the entry-word is indicated and to which the equivalent is coordinated (the so-called canonical form) is a representative of the whole paradigm of the entry-word, that is, if the source language happens to be a language with paradigms. Therefore, before the inclusion of a word, with its equivalent(s), into the dictionary, its whole paradigm should be checked, and the more important semantic peculiarities of its single forms should be duly noted.

If polysemy needs semantic differentiation by the context, we can expect that the same will be true of homonymy (overlapping as the two notions are). The situation is basically identical, so there is no need to discuss special examples. There is, however, a special type of situation, in which a homonymous pair or polysemous meanings are differentiated by the form. German *Abrede* generally has the meaning of "understanding, agreement"; but the set expression *in Abrede stellen* means "to disavow, to dispute". This expression being rather frequent, the reduced entry could have a form like:

Abrede (1) → *understanding, agreement*
(2) *in Abrede stellen* → *to disavow, to dispute*⁷

This brings us to a topic which I shall mention only briefly, namely the fact that there are combinations of words which are set, which have a unified meaning, and which even

⁷ This type overlaps with the type of *handgreiflich werden* as discussed above.

function as a lexical unit of a language. There are many various types of them. A dictionary of the type under consideration here, prepared for coping with texts of a limited range only, will hardly select many colorful idioms such as *Das Hasenpanier ergreifen* "To fly away". But it will have to list frequently occurring set expressions like *in Abrede stellen*, particularly when their meaning is not predictable from that of their individual parts. Also, a dictionary of our type will probably select many technical terms which consist of more than one word. The technical terminology of any science gives many examples of the type *leichte Infanterie, schwere Infanterie*, etc. The situation in German is particularly easy, because a large number of such terminological coinages have the form of compound word, cf. *Dampflokomotive* "steam engine". Still, there is no predictable regularity in this, cf.

Sauerkraut "pickled cabbage", but
saure Gurken "pickled cucumber",

so that the lexicographer has to check the whole semantic area carefully. It will also be necessary to have the productive parts of compound words listed in the dictionary as entries of their own if they have a regular effect on the meaning of the whole compound. With real compound words, this is not too frequently the case, but affixes and elements which approach the status of affixes can be treated this way. For instance, German *ur-* → "proto-"; *pseudo* → English "pseudo-", etc. Such an indication has the big advantage that it is so to say productive: it can take care of newly coined expressions (assuming they are coined regularly), unknown at the moment of the compila-

8 On these, see my "Multiword Lexical Units", linguistic studies presented to A. Martinet I, p. 578 ff.

tion of the dictionary.

There are some points which may deserve to be mentioned. Many a dictionary tends to forget that we find multiword lexical units not only among the denotative words. But the inclusion into the dictionary of expressions like German *ab und zu* "from time to time"; or German *auf und ab* "up and down" is useful. And again, we will have to put into the dictionary indications of how to discern polysemy. Consider the difference between German *von* (*heute, nun, jetzt, gestern, etc.*) *ab*, English "from (today, now, yesterday, etc.) onwards", and German *vom Bahnhof ab* (*geht die Strasse, bergab*) "the street begins to go downhill at the station". Therefore, a strongly reduced part of the entry should have the form:

von ab (1) [*Zeitangaben*] → *from . . . onwards*
(2) [*Ortsangaben*] → *at, beyond*

A particularly obnoxious type of set expressions are those which allow a certain variation. For instance, German *es* (*tut, schadet, macht*) *nichts* has a good English equivalent in "it does not matter". It would seem that there is no complication in this. Let us, however, consider the following sentences: *Es tut nichts*. "It does not matter". -- *Er tut nichts*. "He is doing nothing". This shows us that a set expression may have parts which allow some variation, but again, it has parts that do not. Therefore, a good dictionary will have to contain indications of the following type:

es (*tut, schadet, macht*) (*nichts, wenig*) →
→ <*it does not matter.*

It can be said that the most difficult problem will be how to guarantee that an automatic device will make the correct choice from among the partial equivalents of the target language. This task is so difficult in itself that we should not make it even more difficult by indicating too many (partial) equivalents of the target language. Let us consider some entries discussed above. An entry of the type

auslegen (2) *etwas mit etwas*

(a) [*Teppiche*] → "to cover with (carpets) ,
to carpet"

(c) [*Elfenbein*] → "to inlay, encrust with
(ivory)"

does not strike us as unusual. The verbs *to inlay* and *to encrust* are synonymous for all practical purposes. Every human user of a dictionary is accustomed to understanding an indication like this, so that he is free to use either one or another synonym.

On the other hand, if we take a part of the entry of *erledigen* discussed above

erledigen (1) *etwas* → *to finish, arrange, settle*

we see that it has the same form, but the difference is in the fact that the English verbs are rather mere near-synonyms than full synonyms. Again, a human user is accustomed to seeing entries of this type in any dictionary. Some dictionaries try to distinguish the synonyms from the near-synonyms by using commas in the first and semicolons in the second case: *to inlay, encrust with, but to finish; (to) arrange; (to) settle*. It is, however, extremely difficult to make the distinction in a systematic way, there being

probably more borderline cases than clear-cut ones. And then again, a human user does not need a typographical indication of the distinction so badly: if he is a native speaker of the target language, he knows the distinction anyhow; if he is a speaker of the source language, he may make an error in his choice, but an error which will not be too grave, and with growing knowledge of the target language, he will also acquire the "feeling" for when to use one or another of the near-synonyms.

This is how bilingual dictionaries, particularly the smaller ones, operate: they rely on the abilities and knowledge of the human user. The indications of such dictionaries very frequently have the main purpose of triggering in the human user thinking and imaginative processes which make him recollect words and expressions not immediately indicated in the dictionary.⁹ We cannot rely on all these abilities when we construct a dictionary for mechanical use. Therefore, the rule should be that there should be no unspecified indication of synonyms as partial equivalents: if there *is* more than one partial equivalent, they should be accompanied by the necessary restrictive glosses which will show which to choose. If both equivalents can really be used unrestrictedly, i.e. if they are fully synonymous, it is possible to indicate only one of them (preferably the more frequently occurring one) or to indicate the possibility of free variation, e.g. for stylistic purposes.

⁹ This statement is focussed particularly on bilingual dictionaries of living languages for general use. Large philological dictionaries of dead languages are of a different type: they frequently contain an enormous mass of quoted contexts with concrete translations and thus make the information given quite factual and concrete. The human user, however, will still tend to go beyond the indications of this dictionary, since, after all, the indications of a dictionary and an adroit translation of a text are always two things.

To prepare a dictionary which will reach this degree of explicitness and accuracy is an extremely difficult task. Moreover, I am afraid that even when all this is done, there will still occur situations in which the automatic device will not be able to make a choice. This may occur, for instance, in any text where the relevant context is not close to the passage which needs disambiguation. It would seem that in such a situation no random choice should be made but both (or all) possible equivalents should be printed in the output with a sign showing their mutual complementarity.

A similar but much worse situation will occur when the automatic device is faced with a neologism, i.e. with a genuinely new expression or with an "old" expression used with a new sense. To discuss this difficulty, however, is quite a different task, because an attempt at the solution of this problem would require an investigation of the regularity of new coinages. For instance, new terminological coinages tend to have a high degree of regularity. In any case, a discussion of these problems must be reserved for another occasion.

BIBLIOGRAPHY

ARAPOV, M. V.

- 1967 *Sintaksicheskaya model' yazykov* (A syntactic model of languages) [with V. B. Borshchev]. Moscow: Nauka.

Avtomatizatsiya perevoda tekstov (Automatization of the translation of texts). Editorial introduction to a partial translation of the ALPAC report *Language and Machines, Computers - Translation and Linguistics*. NTI, ser. 2, No. 8.

BACH, EMMON

- 1971 "Syntax since *Aspects*", in *Report of the 22nd Annual Round Table Meeting*. Washington, D.C.: Georgetown University Press.

BAR-HILLEL, YEHOShUA

- 1964 *Language and Information*. Reading, Mass.: Addison-Wesley Publishing Co.

1967a "Dictionaries and meaning rules". *Foundations of Language* 3:409-14.

1967b "Measures of syntactic complexity" (with A. Kasher and E. Shamir), 29-50 in *Machine Translation*, A.D. Booth, ed. Amsterdam: North-Holland. New York: John Wiley S Sons.

1967c "Review of *The Structure of Language: Readings in the Philosophy of Language* (ed. by J. A. Fodor and J.J. Katz)". *Language* 43:526-50.

1968 "Universal semantics and philosophy of language: Quandaries and prospects", in *Substance and Structure of Language*, J. Puhvel, ed.. Berkeley and Los Angeles, California: University of California Press.

1969 "Formal logic and natural languages: A symposium". *Foundations of Language* 5:256-84.

1970 "Some reflections on the present outlook for high-quality machine translation (Position paper on machine translation in 1970)". Austin, Texas: Linguistics Research Center, The University of Texas at Austin (appended).

BORSHCHEV, V. B.

- 1967 *Dispozitsii, algoritmi i porozhdayushchiye procedury* (Dispositions, algorithms, and generative procedures) [with Yu. A. Shreider]. Moscow: Nauka.

FILLMORE, CHARLES

- 1968a "The case for case", 1-88 in *Universals in Linguistic Theory*, E. Bach and R. Harms, eds.. New York: Holt, Rinehart & Winston.
- 1968b "Lexical entries for verbs". *Foundations of Language* 4:373-93.
- 1968c "Types of lexical information", 65-103 in *Working Papers in Linguistics No. 2*. Columbus, Ohio: Ohio State University. [Also to be in *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, Anthropology, and Psychology*, D. Steinberg and L. Jakobovits, eds. Cambridge, Mass.: Cambridge University Press; and in *Proceedings of the Balatonszabadi Conference on Mathematical Linguistics*, F. Kiefer, ed. Dordrecht, Holland: D. Reidel.]
- 1969 "Verbs of judging: An exercise in semantic description", 91-117 in *Papers in Linguistics* 1:i. Tallahassee, Florida: Florida State University.
- 1970: "On a fully developed system of linguistic description". Austin, Texas: Linguistics Research Center, The University of Texas at Austin.

FRASER, J. BRUCE

- 1965 *An Examination of the Verb-Particle Construction in English*. Ph.D. dissertation. Cambridge, Mass.: M.I.T.
- 1966a "Some remarks on the verb particle combination in English", 45-61 in *Report of the 17th Annual Round Table Meeting*, C.I.J.M. Stuart, ed. Washington, D.C.: Georgetown University Press.
- 1966b "Survey of automated language processing 1966" [with D.G. Bobrow and M.R. Quillian]. Cambridge, England: Bolt, Beranek & Newman.
- 1969 "An augmented state transition network analysis procedure", 557-67 in *Proceedings of the International Joint Conference on Artificial Intelligence*. Bedford, Mass.: The MITRE Corporation.

FRASER, cont'd

- 1970 "Idioms within a transformational grammar".
Foundations of Language 6:22-k2.

GARVIN, PAUL L.

- 1966a *Predication Typing--A Pilot Study in Semantic Analysis* [with J. Brewer and M. Mathiot].
Canoga Park, California: Bunker-Ramo Corporation.
- 1966b "Some comments on algorithm and grammar in the automatic parsing of natural languages".
Mechanical Translation 9:2-3.
- 1967 "Machine translation - fact or fancy?" *Datamation*,
April .
- 1970 "Operational problems of machine translation: a position paper". Austin, Texas: Linguistics Research Center, The University of Texas at Austin (appended) .

JOSSELYN, HARRY H.

- 1967 "Lexicography and the computer", 1046-1059 in
To Honor Roman Jakobson. The Hague, Holland:
Mouton.
- 1968a "The lexicon: a matrix of lexemes and their properties", in *Proceedings of the Balatonszabadi Conference on Mathematical Linguistics*, F. Kiefer, ed. Dordrecht, Holland: D. Reidel.
- 1968b *Research in MT: Russian to English. Ten Year Summary Report*. Detroit, Michigan: Wayne State University.
- 1969a "The lexicon: a system of matrices of lexical units and their properties", paper presented at the International Conference on Computational Linguistics. Stockholm, Sweden: KVAL.
- 1969b "A linguistic interpretation of machine translation in the sixties", 1/1-72 in *Research in Computer-Aided Translation, Russian → English, 11th Annual Progress Report*. Detroit, Michigan: Wayne State University.
- 1970 "The matrix as a concept for structuring the lexicon". Detroit, Michigan: Wayne State University, Xerox.

KARTTUNEN, LAURI

- 1967 "The identity of noun phrases". Santa Monica, California: The RAND Corporation.
- 1968a "What do referential indices refer to?" Santa Monica, California: The RAND Corporation.
- 1968b "What makes definite noun phrases definite?" Santa Monica, California: The RAND Corporation.
- 1969a "Discourse referents", Preprint No. 70, International Conference on Computational Linguistics. Stockholm, Sweden: KVAL
- 1969b "Problems of reference in syntax". Austin, Texas: The University of Texas at Austin, mimeograph,
- 1970a "The logic of English predicate complement constructions". Austin, Texas: Linguistics Research Center, The University of Texas at Austin (appended)
- 1970b "On the semantics of complement sentences", 328-39 in *Papers from the 6th Regional Meeting*. Chicago, Illinois: Chicago Linguistic Society.
- to appear "Implicative verbs". *Language*.

KAY, MARTIN

- 1967 "Experiments with a powerful parser", paper No. 10 in *Proc. Deuxième Conférence Internationale sur le Traitement Automatique des Langues*, Grenoble, France. Santa Monica, California: The RAND Corporation .
- 1969 "Computational competence and linguistic performance". Santa Monica, California: The RAND Corporation.
- 1970a "The MIND system: The morphological-analysis program" [with G. R. Martins]. Santa Monica, California: The RAND Corporation.
- 1970b "The MIND system: The Structure of the semantic file" [with S. Y. W. Su]. Santa Monica, California: The RAND Corporation.
- 1970c "Performance grammars". Santa Monica, California: The RAND Corporation.

KULAGINA, O. S.

- 1969 "Eshche raz k voprosu o realizatsii avtomaticheskogo perevoda (Once more on the problem of the realization of automatic translation)" [with I. A. Mel'chuk and V. Yu. Rozentsveyg]. NTI, ser. 2, No. 11.

LAKOFF, GEORGE

- 1965 *On the Nature of Syntactic Irregularity*, Ph.D. dissertation. Bloomington, Indiana: Indiana University. *Mathematical Linguistics and Automatic Translation Report No. NSF-16*. Cambridge, Massachusetts: Harvard Computation Laboratory. [Published as *Irregularity in Syntax*. New York: Holt, Rinehart & Winston, 1970.]
- 1968a "Deep and Surface grammar". Bloomington, Indiana: Indiana University Linguistics Club, mimeograph.
- 1968b "Instrumental adverbs and the concept of deep structure". *Foundations of Language* 4: 4-29.
- 1968c "Is deep structure necessary?" [with J. R. Ross]. Bloomington, Indiana: Indiana University Linguistics Club, mimeograph .
- 1968d "Pronouns and reference". Bloomington, Indiana: Indiana University Linguistics Club, mimeograph.
- 1969 "On derivational constraints", 117-39 in *Papers from the 5th Regional Meeting, Chicago Linguistic Society*. Chicago, Illinois: Department of Linguistics, University of Chicago.
- 1970 "Natural logic and lexical decomposition", 340-62 in *Papers from the 6th Regional Meeting, Chicago Linguistic Society*. Chicago, Illinois: Chicago Linguistic Society.
- to appear "On generative semantics", in *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, Anthropology, and Psychology*, D. Steinberg and L. Jakobovits, eds.. Cambridge, England: Cambridge University Press.

LAKOFF, ROBIN

- 1969 "Some reason why there can't be any some-any rule." *Language* 45:608-15.

LAKOFF ROBIN, cont'd

1970 "Tense and its relation to participants".
Language 46:838-49.

to appear "Questionable answers and answerable questions",
in *Papers in Linguistics in Honor of Henry and
Renee Kahane*, B. Kachru, et al., eds .

LEHMANN, WINFRED P.

1969 *Research in Russian-English Machine Translation
on Syntactic Level*, vols, 1 & 2 [with L. W. Tosh,
R. R. Macdonald, and M. Zarechnak]. Austin, Texas:
Linguistics Research Center, The University of
Texas at Austin.

1970 *Research in German-English Machine Translation on
Syntactic Level*, vols. 1 & 2 [with R. Stachowitz
and L. W. Tosh]. Austin, Texas: Linguistics
Research Center, The University of Texas at Austin.

LYONS, JOHN

1967 "A note on possessive, existential, and locative
sentences". *Foundations of Language* 3:390-96.

1970 "The feasibility of high-quality machine trans-
lation". Austin, Texas: Linguistics Research Center,
The University of Texas at Austin (appended).

McCAWLEY, JAMES D.

1968a "Concerning the base component of a transformational
grammar". *Foundations of Language* 4:243-69.

1968b "Lexical insertion in a transformational grammar
without deep structure", 71-80 in *Papers from the
4th Regional Meeting, Chicago Linguistic Society*.
Chicago, Illinois: Department of Linguistics,
University of Chicago.

1968c "The role of semantics in a grammar", 124-69 in
Universals in Linguistic Theory, E. Bach and
R. Harms, eds.. New York: Holt, Rinehart & Winston.

1970a "Semantic representation", 227-48 in *Cognition:
A Multiple View*, P. Garvin, ed.. New York and
Washington, D.C.: Spartan.

1970b "Where do noun phrases come from?" in *Readings in
English Transformational Grammar*, R. Jacobs and
P. Rosenbaum, eds.. Waltham, Massachusetts:
Ginn-Blaisdell.

MEY, JACOB

- 1969 "Syntax or semantics: Some controversial issues in computational linguistics". *Norsk Tidsskrift for Sprogvidenskap* 23:59-70.
- 1970 "Toward a theory of computational linguistics", paper presented at the Annual Meeting of the Association for Computational Linguistics. Austin, Texas: Linguistics Research Center, The University of Texas at Austin.

NATIONAL ACADEMY OF SCIENCES - NATIONAL RESEARCH COUNCIL

- 1966 *Language and Machines: Computers in Translation and Linguistics*, a report by the Automatic Language Processing Advisory Committee Division of Behavioral Sciences, Publication 1416, Washington, D.C.

NIDA, EUGENE A.

- 1969 *The Theory and Practice of Translation* [with Charles R. Taber]. Leiden: Brill.

PETERS, P. STANLEY, Jr.

- 1969a "Ambiguity, completeness and restriction problems in the syntax-based approach to computational linguistics" [with R. Tabory]. *Linguistics* 46:54-76.
- 1969b "A note on the universal base hypothesis" [with R. W. Ritchie]. *Journal of Linguistics* 5:151-52.

PETRICK, STANLEY R.

- 1965a "On the relative efficiencies of context-free grammar recognition" [with T. V. Griffiths]. Bedford, Massachusetts: USAF Cambridge Research Laboratories .
- 1965b *A Recognition Procedure for Transformational Grammars*. Ph.D. dissertation. Cambridge, Massachusetts : M. I. T.
- 1966 "A program for transformational syntactic analysis". Bedford, Massachusetts: USAF Cambridge Research Laboratories.
- 1967 "Syntactic analysis" [with S. J. Keyser]. Bedford, Massachusetts: USAF Cambridge Research Laboratories.

PETRICK, cont'd

- 1969 "On coordination reduction and sentence analysis" [with P. M. Postal and P. S. Rosenbaum]. *Communications of the ACM* **12:223-33**.
- 1971a "On the use of syntax-based translators for symbolic and algebraic manipulation", 224-37 in *Proceedings of the Second Symposium on Symbolic and Algebraic Manipulation*. New York: Association for Computing Machinery.
- 1971b "Syntactic analysis for transformational grammars". Austin, Texas: Linguistics Research Center, The University of Texas at Austin.
- 1971c "Syntactic analysis requirements of machine translation". Austin, Texas: Linguistics Research Center, The University of Texas at Austin (appended).

PUMPYANSKIY, A. L.

- 1966 *Upryazhneniya po perevodu nauchnoy i tekhnicheskoy literatury* (Studies on the translation of scientific and technical literature). Moscow: NAUKA.

ROSS, JOHN ROBERT

- 1965 "Underlying structures in discourse" [with T. G. Bever], VIII/1-12 in *Proceedings of the Conference on Computer-Related Semantic Analysis*. Detroit, Michigan: Wayne State University.
- 1967 *Constraints on Variables in Syntax*. Ph.D.dissertation. Cambridge, Massachusetts: M.I.T. Bloomington, Indiana: Linguistics Club, mimeograph, 1968.
- 1970a "On declarative sentences", in *Readings in English Transformational Grammar*, R. Jacobs and P. Rosenbaum, eds. Waltham, Massachusetts: Ginn-Blaisdell.
- 1970b "Gapping and the order of constituents", in *Proceedings of the 10th International Congress of Linguistics*, Bucharest.

SIMMONS, ROBERT F.

- 1966 "Storage and retrieval of aspects of meaning in directed graph structures". *Communications of the ACM* **9:211-15**.

SIMMONS, cont'd

- 1967 "Answering English questions by computer", 253-89 in *Automated Language Processing*, H. Borko, ed., New York: John Wiley & Sons. [Also in *Communications of the ACM* 8 (1965):53-70.]
- 1970a "Generating English discourse from semantic networks" [with J. Slocum]. Austin, Texas: Computer Assisted Instruction Laboratory, The University of Texas at Austin.
- 1970b "Natural language question-answering system". *Communications of the ACM* 13:15-31.

STACHOWITZ, ROLF

- 1969 *On the definition of the term "discontinuous constituents" in context-free phrase structure grammar*. Ph.D. dissertation. Austin, Texas: The University of Texas at Austin.
- 1970a "The construction of a computerized dictionary", paper presented at the Modern Language Association Lexicographical Conference, Columbus, Ohio. Austin, Texas: Linguistics Research Center, The University of Texas at Austin.
- 1970b "A model for the recognition and production of synonymous expressions with different deep structures", paper presented at the Conference on Linguistics at the University of Iowa, Iowa City. Austin, Texas: Linguistics Research Center, The University of Texas at Austin.
- 1971a "Lexical features in translation and paraphrasing: an experiment". Austin, Texas: Linguistics Research Center, The University of Texas at Austin.
- 1971b "Requirements for Machine Translation: Problems, Solutions, Prospects". Austin, Texas: The University of Texas at Austin (appended).
- to appear *Normalization of Natural Language for Information Retrieval* [with W. P. Lehmann]. Austin, Texas: Linguistics Research Center, The University of Texas at Austin.
- Development of German-English Machine Translation System* [with W. P. Lehmann]. Austin, Texas: The University of Texas at Austin.

SWANSON, ROWENA

- 1967 *MOVE THE INFORMATION... A Kind of Missionary Spirit*.
Arlington, Virginia: USAF Office of Aerospace Research.
- 1970 "Trend in information handling in the United States".
Arlington, Virginia: USAF Office of Scientific
Research.

WALKER, DONALD E.

- 1965a "English preprocessor manual". Bedford, Massachusetts:
The MITRE Corporation.
- 1965b "The MITRE syntactic analysis procedure for trans-
formational grammars" [with A. M Zwicky, J. Friedman,
and B. C. Hall], 317-26 in *Proceedings of the 1965
Fall Joint Computer Conference*. New York and
Washington, D.C.: Spartan.
- 1966 "Recent developments in the MITRE syntactic analysis
procedure" [with P. G. Chapin, M. L. Geis, and
L. N. Gross]. Bedford, Massachusetts: The MITRE
Corp.
- 1967a "On-line text processing: Introduction and overview".
Bedford, Massachusetts: The MITRE Corporation.
- 1967b "SAFARI, an on-line text-processing system",
Vol. IV, 144-47 in *Proceedings of the American
Documentation Institute*. London, England:
Academic Press.
- 1969a "Computational linguistic techniques in an on-line
system for textual analysis". Bedford, Massachusetts:
The MITRE Corporation.
- 1969b "on-line computer aids for research in linguistics"
[with L. N. Gross], 1531-36 in *Information Pro-
cessing 68*. Amsterdam: North-Holland.
- 1970 "The current status of computer hardware and
software as it affects the development of high
quality machine translation". Austin, Texas:
Linguistics Research Center, The University of
Texas at Austin (appended).

WINOGRAD, TERRY

- 1970 *Procedures as a Representation for Data in a
Computer Program for Understanding Natural
Language*. Ph.D. dissertation. Cambridge,
Massachusetts: M. I. T. [Revised version published
1971, Cambridge, Massachusetts: M. I. T., Project
MAC.]

ZGUSTA, LADISLAV

1970a "Equivalentents and explanations in bilingual dictionaries", paper presented at the Modern Language Association Lexicographical Conference, Columbus, Ohio. Austin, Texas: Linguistics Research Center, The University of Texas at Austin (appended).

1970b "The shape of the dictionary for mechanical translation purposes". Austin, Texas: Linguistics Research Center, The University of Texas at Austin (appended) .

to appear *Manual of Lexicography*. Prague, Czechoslovakia: Academy of Sciences.