

ON THE USE OF TRANSITIONAL PROBABILITY ESTIMATES IN PROGRAMMING FOR MECHANICAL TRANSLATION

Bjarne Ulvestad

Paper read by Bjarne Ulvestad, University of Bergen,
to The Scandinavian Symposium on statistical Lin-
guistics at Sundbyholm Manor in November 1960.¹

Scientific-theoretical methodology is fundamentally concerned with the general rather than with the accidental, however interesting the latter may be. The accidental or deviational is of little or no importance in comparison with the general picture, and this picture can often be obtained through standard quantification procedures, although it must be admitted that such procedures are not always rigorously stated in terms of rules. Most of our scientific laws are essentially statements of high statistical validity, but the nonabsolute nature of the laws is usually not explicitly referred to in such documents as e. g. introductory textbooks for high-school and college.

Also the syntax of a natural historical language, say English or German, may be regarded as consisting of a long series of rules or statements most of which are true only in terms of relatively high statistical validity. The basically nonabsolute validity of such statements will on closer inspection be found to obtain even for many rules which are usually considered to permit of no deviations and which are categorically stated in our grammars and syntaxes. Actually a great many syntactical rules are true only on the whole, most of the time, and the majority of these can be made more precise through subdivision of the material covered and the concomitant elaboration of subrules. As a simple illustrative example one may mention the rules of agreement in number between subject and predicate in the current German standard language. The fundamental statement ordinarily runs as follows, to quote a recent

¹) This work was supported in part by the U. S. Army (Signal Corps), the U. S. Air Force (Office of Scientific Research, Air Research and Development Command), and the U. S. Navy (Office of Naval Research).

work, Jude's excellent *Deutsche Grammatik*:¹ 'Das Verb im Prädikat muss in der Regel mit dem Subjekt (im) Numerus überstimmen' (p. 212). Then seven admirably neat and concise subrules are added under this one, including the following: 'Haben die Subjekte *teils Singular, teils Pluralform*, so kann das Verb im Singular stehen, wenn es unmittelbar bei einer solchen Form steht und kein Plural vorausgeht' (p. 213, italics mine). As examples illustrating this rule one ordinarily cites, as does Jude (ibid.), sentences with junctional subject, such as: *da kommt die Mutter und die Kinder*; and *ihm gehört das Feld und die Wälder*. The first reservation that one may bring forth here pertains to the fact that the reference to plural form on the part of one of the subjects, or rather one of the elements in the subject nounphrase is not necessary in this particular subrule; also singular 'subjects' connected by *und* may appear, thus: *da kommt die Mutter und das Kind* (but perhaps more frequently: *da kommen die Mutter und das Kind*). Observe that Jude states that the verb *may* be in the singular (*kann . . . im Singular stehen*), which indicates by implication that it is normally in the plural.

In a corpus of twenty-four novels picked at random, eighteen sentences of the following basic type were found: Ssing Psing *und* Ssing (S = subject, P = predicate). Three of these sentences are: *und alle Wesen atmeten sie . . . selig in sich ein.*) *Victoria atmete sie und der Kerkermeister* (Niebelschütz, p. 19); *Vater war tot und Mutter* (Gaiser, p. 59); *Nichts als schlechte Laune war in seinem Blick und ein dünner, manichäischer Hohn* (Holthusen, p. 227). All the eighteen sentences of this type would seem to be taken care of by Jude's subrule, as stated, except that it would in this case be factually and logically wrong (from the point of view of rule-making), to say that the verb *may* be in the singular. It is preferable to state that the predicate verb regularly occurs in the singular number in such sentences as these, but that it *may* be in the plural. In the corpus examined I have found only one sentence of this form with a plural verb: *Leslie aber lebten und er* (Müller, p. 8). Consequently, we may establish two improved subrules instead of one faulty one (optional, but less frequent elements are in brackets; the symbol / means 'or'):

1. X + Ppl (sing) + Ssing + *und* + Ssing/pl
2. Y + Ssing + Psing (pl) + *und* + Ssing/pl

¹) See the bibliographical references at the end of the article.

The symbols X and Y may represent adverbs, adverbial clauses, coordinating conjunctions, and other sentence introductory elements. Formally, the only absolute difference between the data subsumed under the two rules would seem to lie in the 'normal' word order in sentences covered by rule 1, the inverted word order in sentences covered by rule 2.

It is true that the rules 1 and 2 may be regarded as trivial in the sense that they describe utterance types of rather low probability of occurrence in texts, but this short initial discussion of a simple problem will perhaps facilitate the readers' understanding of the main argument of this paper. The important thing to keep in mind is that two subrules rather than one were found to be preferable because the deviation patterns (from the point of view of the rule that was inspected) turned out to require definitional differentiation, *in a statistical sense, not in an absolute one*. Sentences generated by rules 1 and 2 clearly contain the same basic elements, but in one sequence a singular predicate is regularly found, in the other sequence, a plural predicate most frequently occurs.

The purpose of this short study is to demonstrate how cautious application of some of the salient statistical methods employed in the field of information theory can be useful also for the new discipline of mechanical translation, especially with regard to the codification of the output language. That the result is of interest also in the field of syntax in general, needs no elaborate justification. We shall start out with a set of particular mechanical translation problems, and in the course of the subsequent exposition it will be found that they are indeed solvable, and within a much more general formal-syntactical framework than the initial statement of the problems would seem to suggest.

These particular problems may be succinctly stated as follows: How should one translate these eight English word sequences into 'maximally normal' or 'good' German as the output language:

- (1) 'he believed (that) he had seen him'
- (2) 'he believes (that) he has seen him'
- (3) 'I believed (that) I had seen him'
- (4) 'I believe (that) I have seen him'
- (5) 'he knew (that) he had seen him'
- (6) 'he knows (that) he has seen him'
- (7) 'I knew (that) I had seen him'
- (8) 'I know (that) I have seen him'

The relative complexity of each of these eight transfer 'problems' may be seen from e. g. English input sentence (1) 'he believed (that) he had seen him', which can be adequately translated into a number of substantially isosemantic, (at least in a cognitive sense) German correspondents:

- (1.1) er glaubte, er habe ihn gesehen
- (1.2) _____, — hätte _____
- (1.3) _____, — hatte _____
- (1.4) _____, dass er ihn gesehen habe
- (1.5) _____, _____ hätte
- (1.6) _____, _____ hatte
- (1.8) _____, _____ Ø (zero)
- (1.8) _____, ihn gesehen zu haben

One observes that there are eight different German translations possible, but the mechanical translation apparatus can presumably pick only one for the output language, or simply choose randomly, unless some kind of formally identifiable differentiating signals be found in the English input language. Such signals are however not present in the case under consideration here. It would of course be tempting to resort to an arbitrary decision rule stating that if the conjunction *that* is found in English, the conjunction *dass* should be chosen in the German output language, but such a rule would seem too often to run counter to the normal German sentence patterns to judge from the following succinct, but probably greatly oversimplified, 'description' by Fowler (T. = 'that'):

- (1) T. is usual with *agree, assert, assume, aver, calculate, conceive, hold, learn, maintain, reckon, state, suggest*;
- (2) T. is unusual with *believe, presume, suppose, think*;
- (3) T. is used or omitted with *be told, confess, consider, declare, grant, hear, know, perceive, propose, say, see, understand*. The verbs with which the question may arise are many more than these few, which may however be enough to assist observation (p. 633).

To take an example from Fowler's group 1, the conjunction 'that' in 'he asserted that he was sick' probabilistically corresponds to conjunction

zero (0) in the German correspondent sentence *er behauptete, er sei krank*, as will be seen later, and the probability distribution for 'that' and *dass* in this particular example could accordingly be 'tabulated' in this manner ('low' = 'low probability'; 'high' = 'high probability'):

	Ø	'that'/dass
English: 'he asserted'	low	high
German: <i>er behauptete</i>	high	low

The eight German sentences (1.1) ... (1.8) listed above are all instances of what traditional grammarians ordinarily term indirect discourse in the wider sense, cf. Erdmann's often-cited definition: 'Indirekte Rede im weitesten Sinne nenne ich alle Sätze, deren Inhalt der Sprechende (A) als Gegenstand der Kenntnis, der Vorstellung oder Rede einer anderen Person (B) angibt. Da man auch über seine eigenen Kenntnisse, Vorstellungen, Reden berichten kann, so kann B mit A identisch sein' (p. 168). Having thus identified the category in which the output sentences (1.1) ... (1.8) belong, we must postpone the discussion regarding the preferable German translations of the input sentences (1) ... (8) and return to it only toward the end of the paper. There are two important reasons for doing this: 1. It is scientifically preferable to try to make as general statements as possible about typical German indirect discourse sentences, in such a way that a large number or particular problems can ultimately be solved deductively, preferably of course without recalcitrant residue. 2. There are, as will be seen below, several instances of English indirect discourse sentences that do not permit of the whole differentiation range of German correspondent sentences that for instance sentence (1) displays, and the German indirect discourse sentence types that have maximum, and possibly even universal, transfer applicability, must be given preference over those that do not, insofar as no appreciable distortion of sentence meaning takes place.

In this attempt to arrive at generally valid rules for the output language, we shall consider a number of possible German sentences of indirect discourse, whose governing verbs (which were selected on the basis of frequency estimates) are: 1. *ahnen*, 2. *antworten*, 3. *begreifen*, 4. *behaupten*, 5. *berichten*, 6. *bitten*, 7. *denken*, 8. *einsehen*, 9. *erkennen*, 10. *erklären*, 11. *erwidern*, 12. *erzählen*, 13. *fühlen*, 14. *fürchten*, 15.

gestehen, 16. *glauben*, 17. *hoffen*, 18. *meinen*, 19. *merken*, 20. *mitteilen*, 21. *sagen*, 22. *schreiben*, 23. *versichern*, 24. *verstehen*, 25. *wissen*, 26. *wünschen*.

On examining a large number of sentences with these verbs in the governing clauses, one arrives at what appears to be cogent reasons for excluding the German sentence types (1.7) and (1.8) from the series of correspondences to (1) that were listed above: 1. Omission of auxiliaries (type 1.7) is a highly restricted phenomenon; it is seldom permitted for auxiliary verbs other than *haben* and *sein* and is of course only found in compound verb phrases (*gewesen sein, gesehen haben* etc.); even here it is actually extremely rare (cf. Jude p. 140). 2. Type (1.8) is possible only if the logical subjects of the governing clause and the dependent clause are identical, and even so it has a rather limited range. Thus it may be found with *meinen, behaupten, glauben*, but hardly with e. g. *sagen* in nonreflexive usage, *verstehen*, and *wissen*.

We are consequently left with the six German sentence types (1.1) ... (1.6). These can also be reduced in number, since apparently usage of past versus present tense forms of the subjunctive is largely a matter of individual taste, though to some extent considerations of formal ambiguity enter in. Thus the standard traditional grammars inculcate the use of present tense forms except in cases where the subjunctive is not clearly marked (e. g., *ich hätte* rather than *ich habe*). Judged from the point of view of current German prose, the rule that past tense forms should be employed if otherwise modal ambiguity would arise, is however far too absolute in its usual formulation. Careful inspection of practically any modern novel will yield numerous unequivocal instances of nonconformity with this rule, and it may be further noted that the rule is of only limited usefulness to avoid ambiguity. Thus no manipulation of the simple tense forms will resolve the modal (indicative/subjunctive) neutrality in e. g. *ich höre|hörte*, and the somewhat puzzling suggestion, most recently made by Hoppe (p. 140), to use compounds with *werden* in such cases (thus *ich würde hören* instead of *ich hörte|höre*), does not seem to be adhered to at all by modern German authors. The problem as to subjunctive tense form usage in indirect discourse is actually a minor one from the point of view of mechanical translation and we shall not consider it further in this paper. It is perhaps to be regarded as a problem in morphology rather than one in syntax.

We have now reduced our output sequence inventory to four slightly different types of indirect discourse sentences, and these are evidently of universal applicability, by which I mean that they can occur with any governing verb. They may be represented diagrammatically as follows (s = subjunctive, i = indicative), considering the output sentence generation to run from left to right:

S — GV	∅	s
	d	i
		s
		i

The elements symbolized here may, at least for our particular purposes, be regarded as the nuclear or basic constituents of German indirect discourse sentences, and as a typical series of minimal sentences one may list the following.

- | | | | |
|----|--|----------------------------|--|
| 1. | <u>Ich</u> (S) <u>sagte</u> (GV) <u>dass</u> (d) | <u>ich</u> <u>ging</u> (i) | |
| 2. | _____ ∅ | _____ | |
| 3. | _____ <u>dass</u> | <u>ginge</u> (s) | |
| 4. | _____ ∅ | _____ | |

Note that the subject of the dependent clause may be regarded as invariant for our purpose (i. e. no specification in terms of noun, pronoun, number etc. needed). It has apparently no influence whatsoever on the basic form of the sentence in which it occurs, and it is the factors influencing transitional probabilities that we are interested in here. Also, we need not be concerned at this point with the fact that the finite verbs in *dass*-clauses, in contradistinction to those of zero-clauses, always occur in what is often loosely called clause final position. This is in modern German prose an obligatory syntactical relationship, and the difference in word order is accordingly predictable, given the conjunctive (∅ or *d*) that introduces the dependent clause.

Now we can raise our particular problems stated above to a more general level and ask questions pertaining to more inclusive categories, rather than to individual members of these: 1. What are the normal choices of conjunction (∅ or *d*) given the listed GVs (1—26) in the present or past, with 1 sg or 3 sg subject in the governing clause?

2. Given the preceding items, including the connective, what are the normal choices of mode in the finite verbs of the dependent clauses? ¹

We shall endeavour to answer these questions on the basis of modern German prose literature. Our corpus consists of two parts. The first one was collected from more or less randomly but impartially selected works by thirty-five contemporary German authors, ² the second one was collected from works by thirty-four German novelists born in 1906 or later and listed in F. Lennartz, *Dichter und Schriftsteller unserer Zeit*, 7th ed. (Stuttgart, 1957). The author names and book titles for the second part of the corpus will be given in a forthcoming monograph on the structure of German indirect discourse sentences and need not be listed here (all the writers quoted in this paper are of course represented in the corpus, but citing of titles of books and page numbers has been dispensed with here to economize with space). In collecting our material, we attempted not to overlook relevant data, but a few oversights would presumably not change the general picture as represented below.

On the basis of our corpus, we shall attempt to elaborate in terms of transitional probabilities the over-all selective habits of the sixty-nine German writers with reference to indirect discourse sentences with the nuclear constituent that were set down above. The word sequences with the lowest 'entropies', i. e. the highest transitional probabilities of occurrence, may be considered normal or regular, and these may be registered as the only ones that the output code of German needs to operate with. What we in effect propose to do is to reduce all low figures to zero and raise the remaining (high) figures to 1. After the registration of such normal translations, the problem of potential transfer distortion, which is admittedly a very complex one, will be discussed shortly.

¹) Question 2. has been discussed, on the basis of more limited material, in my article, "An Approach to Describing Usage of Language Variants", *IUPAL, Memoir* 12 (1956), pp. 37—59. The methodology employed in the present study is however entirely different, and so is the purpose of this paper.

²) B. Ulvestad, *Indirect Discourse in Modern German*, diss. (Madison, 1954, pp. 11—14.)

We shall not work out problems in terms of explicitly stated information content or entropy here, it is not even certain that it can be done, but we shall merely set down the transitional probabilities and then broadly refer to sequences of high transitional probabilities as low-entropy sequences.

It is important to note that the point where we begin to ask for the probability of the choice of the first sentence element following the given sequence was not established arbitrarily: It was chosen exclusively from the point of view of mechanical translation. We start at the end of the German sequence that has a unique English counterpart, for instance after *er sagte*, since this sequence is translationally equivalent to *he said*. In other words, we are not interested in the answers to largely trivial questions such as: given the subject *ich*, which are the probabilities for the choices of the various GVs (1—26), or: given *ich* and a given GV, what is the probability of its being in the past tense. At least from a mechanical translation point of view, such questions are of very minor interest.

Tables 1 and 2, which are given only for the sake of illustration, show how one can divide GVs 1—26 into two major groups, 1. GVs which normally take dependent clauses with subjunctive finite verbs; we shall call these A-verbs; 2. GVs which normally take dependent clauses with indicative finite verbs; we shall call them B-verbs. U1 is the first part of the corpus, U2 the second. L represents the corpus collected by Läftmann in the second decade of this century and is given only to demonstrate the rather great homogeneity which exists in the use of indirect discourse within the frames specified; it will be seen that the figures for L are, in the aggregate, substantially similar to those for U1 and U2. (Study Tables 1 and 2 at the opposite side)

It turns out that *wissen*, *erkennen*, *merken*, *fühlen*, *begreifen*, *einsehen*, *ahnen*, and *verstehen* are B-verbs, the rest of the twenty-six GVs in our list are A-verbs. Apart from the obvious differences in choice of mode on the part of the A- and B-verbs, one notes another interesting fact: The dependent clauses governed by A-verbs most frequently found are \emptyset -clauses (79.04% in U1, 65.45% in L, 80.24% in U2), whereas \emptyset -clauses are relatively rare after B-verbs (7.71% in U1, 8.57% in L, 9.66% in U2).

TABLE 1

GVs in past tense, all persons (i = number of 'indicative dependent clauses', s = number of 'subjunctive dependent clauses').

	U 1		L		U 2	
	\emptyset , s:i	d, s:i	\emptyset , s:i	d, s:i	\emptyset , s:i	d, s:i
bitten	21:0	1:0	7:0	1:0	27:0	2:0
behaupten	22:0	2:0	22:0	2:0	28:0	1:0
antworten	26:0	2:0	3:0	3:0	28:0	2:0
meinen	99:1	12:1	38:2	5:0	112:1	12:2
glauben	71:2	12:2	14:1	5:2	90:2	16:3
schreiben	52:1	6:1	4:0	4:0	66:1	7:2
erklären	74:2	15:1	26:0	13:0	84:0	14:1
sagen	223:5	52:17	60:2	16:5	249:5	70:12
denken	101:10	20:4	25:4	5:2	118:0	21:3
erzählen	37:0	17:2	11:0	21:6	49:0	17:3
erwidern	6:0	2:0	5:0	2:0	7:0	1:0
fürchten	11:0	4:1	6:0	2:0	18:0	3:1
versichern	9:0	6:1	10:0	11:0	15:0	5:1
hoffen	9:0	10:0	3:0	0:0	10:0	7:0
wünschen	7:0	5:0	0:0	6:0	11:0	6:0
berichten	8:0	7:0	4:0	5:0	13:0	7:1
mitteilen	8:0	4:1	0:0	5:1	7:0	3:1
gestehen	2:0	6:1	3:0	9:1	5:0	8:1
Sums	786:21	183:31	241:9	115:17	937:9	202:31

TABLE 2

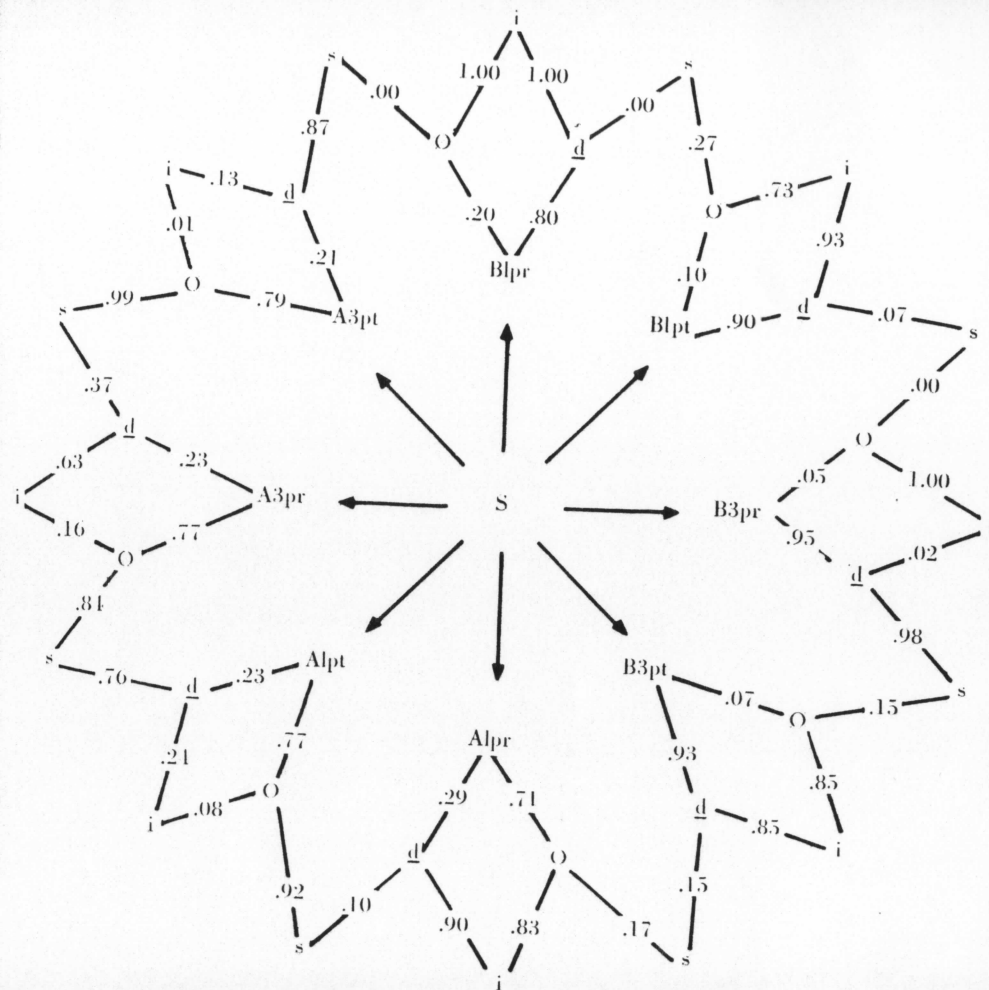
	U 1		L		U 2	
	\emptyset , s:i	d, s:i	\emptyset , s:i	d, s:i	\emptyset , s:i	d, s:i
wissen	3:19	18:125	1:3	6:14	4:28	11:149
erkennen	1:1	4:32	2:0	3:9	0:1	5:46
merken	2:3	7:75	0:0	5:25	1:4	6:80
fühlen	0:3	8:59	0:0	8:23	0:5	7:68
begreifen	0:1	9:26	0:0	1:5	0:3	9:34
einsehen	0:0	3:6	1:1	3:5	0:2	2:11
ahnen	0:0	3:10	0:1	4:6	0:1	3:11
verstehen	0:0	0:11	0:3	1:10	0:0	0:16
Sums	6:27	49:344	4:8	31:97	5:44	43:415

To return to our general problem, we shall regard the following syntactical entities as given: GV type (A or B), tense (pr or pt) person (1st or 3rd sing). From a series of tabulations substantially similar to Tables 1 and 2, we calculated the transitional probabilities given in the circular matrix, which is in effect composed of a number of twostep binary-choice Markoff processes proceeding from the source S toward the periphery.

A few words on methodology must be added here, to indicate how our approach differs from information-theoretical work so far. Weaver writes in his paper, "Recent Contributions to the Mathematical Theory of Communication":

At this point an important consideration . . . comes to the front for major attention. Namely, the role which probability plays in the generation of the message. For as the successive symbols are chosen, these choices are, at least from the point of view of the communication system, governed by probabilities; and in fact by probabilities which are not independent, but which, at any stage of the process, depend upon the preceding choices. Thus, if we are concerned with English speech, and the last symbol chosen is "the", then the probability that the next word be an article, or a verb form other than a verbal, is very small. This probabilistic influence stretches over more than two words, in fact. After the three words "in the event" the probability of "that" as the next word is fairly high, and for "elephant" as the next word is very low. (Shannon-Weaver, pp. 101—2).

While current information theory as applied to linguistic data is concerned with a large number of choices at each stage of calculation, so that the sets for selections may be the whole alphabet, the lexicon of a language etc., we have reduced our choices to binary ones only. For instance, with reference to Weaver's example cited above, we would also, like Weaver, be interested in the probability of the next word after 'in the event', but only within defined syntactical categories, here what may be termed 'in the event'-clauses, which may in turn belong to a larger category (including e. g. 'in case'-clauses). One particular dichotomous choice in Weaver's example would be between 'that' and zero, e. g. in the sentence, 'in the event that "elephant" is an English word' versus 'in the event "elephant" is an English word', not between 'that' and 'elephant', 'that' and 'man' etc.



MATRIX

LEGEND

- S = subject of governing clause
- A = governing verb of type A
- B = governing verb of type B
- 1 = first person singular
- 3 = third person singular
- pr = present tense
- pt = past tense
- Ø = connective zero
- d = connective dass
- s = subjunctive
- i = indicative

The probability of a sequence is the product of the probabilities of the transitions between the signs of which it is composed. Multiplying together the transitional probabilities in the matrix chains, e. g. probability (*d*) x probability *d* (*s*), we obtain the results stated below (the highest probability figures are underlined; one example for each of the normal or typical, 'low entropy' sequences is added for illustration):

TABLE 3

1. A 1 pr	∅	s	<u>.1207</u>	5. B 1 pr	∅	s	.0000
_____	—	i	<u>.5893</u>	_____	—	i	.2000
_____	<i>d</i>	s	.0290	_____	<i>d</i>	s	.0000
_____	—	i	.2610	_____	—	i	<u>.8000</u>
<i>Ich sage, ich bin krank</i>				<i>Ich weiss, dass ich krank bin</i>			
2. A 1 pt	∅	s	<u>.7084</u>	6. B 3 pr	∅	s	.0000
_____	—	i	.0616	_____	—	i	.0500
_____	<i>d</i>	s	.1748	_____	<i>o</i>	s	.0190
_____	—	i	.0552	_____	—	i	<u>.9310</u>
<i>Ich sagte, ich sei krank</i>				<i>Er weiss, dass er krank ist</i>			
3. A 3 pr	∅	s	<u>.6468</u>	7. B 1 pt	∅	s	.0270
_____	—	i	.1232	_____	—	i	.0730
_____	<i>d</i>	s	.0851	_____	<i>d</i>	s	.0630
_____	—	i	.1449	_____	—	i	<u>.8370</u>
<i>Er sagt, er sei krank</i>				<i>Ich wusste, dass ich krank war</i>			
4. A 3 pt	∅	s	<u>.7821</u>	8. B 3 pt	∅	s	.0105
_____	—	i	.0079	_____	—	i	.0595
_____	<i>d</i>	s	.1827	_____	<i>d</i>	s	.1395
_____	—	i	.0273	_____	—	i	<u>.7905</u>
<i>Er sagte, er sei krank</i>				<i>Er wusste, dass er krank war</i>			

Note that if one should arbitrarily decide to disregard the choice of *dass* vs. zero conjunction and choose only one conjunction for the output code irrespective of governing verb type, the figures for the subjunctive/indicative relationship would differ much more: The probabilities for subjunctive dependent clauses would then run as follows: 1: 0.1497, 2: 0.9832, 3: 0.7325, 4: 0.9648, 5: 0.0000, 6: 0.0190, 7: 0.900, and 8: 0.1495. At least the modal distinction relative to the various GVs would thus have to be incorporated into the output code, and if generation of maximally 'natural' translation texts were to be regarded as an important desideratum, as I think it should, the conjunctive distinction would also need to be taken care of. The solutions of the particular problems which were stated above (how to translate into German the English sentences (1) ... (8) can now be read directly out of Table 3, keeping in mind that *glauben* is an A-verb, and that *wissen* is a B-verb:

- (1) *er glaubte, er habe ihn gesehen*
- (2) *er glaubt, er habe ihn gesehen*
- (3) *ich glaubte, ich hätte ihn gesehen*
- (4) *ich glaube, ich habe ihn gesehen*
- (5) *er wusste, dass er ihn gesehen hatte*
- (6) *er weiss, dass er ihn gesehen hat*
- (7) *ich wusste, dass ich ihn gesehen hatte*
- (8) *ich weiss, dass ich ihn gesehen habe*

This study was based on more than four thousand examples of German indirect discourse sentences of the type under consideration. The differences between the two parts of the corpus (and also between each of these and Läftmann's corpus) were so small that we feel justified in assuming that accumulation of new data will not change the results obtained here in any important respects. Our findings clearly suggest that adequate translation of English indirect discourse sentences is indeed very much at variance with the various mentalistic statements about 'correct' choices of mode as found in most traditional grammars, for instance in Jude: 'Der Gebrauch der beiden Modi hängt nicht von der äusseren Satzform oder einzelnen Konjunktionen, auch nicht von dem Begriff des Verbs im Hauptsatz, sondern *allein von den Absichten des*

Sprechenden ab' (p. 148 italics mine). If this statement reflected actual usage, it is obvious that no 'correct' mechanical transfer into German of e. g. the English sentence 'he said he was sick' would be possible, at least not on a sentence-by-sentence basis, since English does not formally specify the reaction of the speaker to the proposition, thought, wish etc. that he is reporting. Neither does German, *as a rule*, which raises the problem as to the amount of distortion that would be present in our mechanically obtained translations from English indirect discourse sentences.

From a careful examination of the texts from which the second part of the corpus was gathered, it was found that there is probably about as high an incidence of nonadherence to the mentalistic grammarians' laws in modern German novels as there would be in texts obtained through mechanical translation. Thus one finds dozens of instances where the indicative is employed although it is quite clear from the context that the subjunctive would be called for according to the meaning-related 'rules'. Occasionally the 'revealing' parts of the context are even very close to the indirect discourse sentence, in which case the discrepancy between rule and usage is particularly apparent, cf. the following examples chosen at random from a large number of such instances of *Sprachsünden*:

Hatte er nicht gedacht, dass es ein Wink war, ein Zeichen,
das das Schicksal an ihn gegeben hatte? Ach, und in
Wahrheit war es nichts als ein Irrtum (Fussenegger, p. 195).

Vielleicht haben euch welche gesagt, dass die Menschen
gleich sind . . . Wer nur ein Lot Verstand in seinem Hirn-
kasten hat, wird sagen: nein! (Habeck, p. 82).

Wie könnt Ihr behaupten, Herr Kommandant, dass meine
Aussage falsch ist. Frau Gräfin war bei dem Herrn . . .
Das hat seine lautere Richtigkeit (Niebelschütz, pp. 171—72).

Similarly, it is not all difficult to find indirect discourse sentences in which the message is explicitly represented as being true in the im-

mediate context, and yet the subjunctive mode occurs in the finite verb of the dependent clause:

Die Leute sagen, er habe sein Leben vergeudet. Das ist wahr
(Fussenegger, p. 145)

Tonio spricht die lautere Wahrheit, wenn er hinüberuft, wir
seien glücklich, sie zu sehen (Holthusen, p. 112).

Ja, ich war da, log er, aber so, dass man sah, er löge
(Helwig, p. 27).

"Wird im Stall sein," sagte Joan. "Ich werde ihn rufen."
"Ich tue das," sagte Leslie. "Ich gehe einstweilen in das
Haus." Joan sah ihm eine Weile nach. Dann ging sie in den
Stall, um dem Alten zu verständigen, dass Leslie . . . da sei
(Müller, p. 170).

It is obvious from the context that Leslie was there and that Joan was telling the old man the truth.

Examples such as these indicate that the modal distinction in indirect discourse sentences is basically one that can be adequately stated in terms of the dominant selective habits of the speakers or writers, and that it is on the whole an essentially formal, dependent distinction rather than one which reflects conscious semantic differentiation on the part of the reporter (writer). It is in other words a distinction that carries little information content, one which has little 'surprise value', one which has gone a long way toward full grammaticalization, i. e. toward zero entropy within specified frames of grammatical context.

REFERENCES

- O. Erdmann, *Grundzüge der deutschen Syntax*, 1 (Stuttgart, 1886).
H. W. Fowler, *A. Dictionary of Modern English Usage*, 15th printing
(London, 1954).
G. Fussenegger, *Die Brüder von Lasawa* (Salzburg, 1948).
G. Gaiser, *Zwischenland* (München, 1949).

- F. Habeck, *Der Tanz der sieben Teufel* (Wien, 1950).
 W. Helwig, *Das Wagnis* (Hamburg, 1947).
 H. E. Holthusen, *Das Schiff* (München, 1956).
 A. Hoppe, *Inhalte und Ausdrucksformen der deutschen Sprache* (Frankfurt, 1955).
 W. K. Jude, *Deutsche Grammatik*, 3rd ed. (Braunschweig, 1954).
 E. Låftmann, *Verbets modus i indirekt anföring i modern tyska* (Stockholm, 1919).
 A. Müller, *Die verlorenen Paradiese* (Mannheim, 1950).
 W. von Niebelschütz, *Verschneite Tiefen* (Berlin, 1940).
 C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (Urbana, 1949).

ÜBER DIE RELATIVE HÄUFIGKEIT DER PHONEME DES SCHWEDISCHEN

Logopäd Mary Weiss
Phonetisches Institut, Uppsala, Schweden

Im Zusammenhang mit einer Untersuchung über die Wirkung des Hörtrainings auf die Sprachperzeption Gehörsgeschädigter ist der Versuch unternommen worden, die relative Häufigkeit der Phoneme des Schwedischen zu berechnen.¹⁾

Das Material bestand aus 22.000 Phonemen. Es bestand die Absicht, den Text nach Möglichkeit gewöhnlicher Konversation entsprechen zu lassen. Er setzte sich daher zum grössten Teil aus Gesprächen und Dialogen zusammen, nämlich: 11 Seiten aus Bertil Malmberg "Åke och hans värld" (S. 49—54 und 87—92), 9 Seiten aus Hazze Z. "Anna-Clara och hennes bröder" (S. 72—78 und 118—119) und 7 Seiten aus Olle Hedberg "Häxan i Pepparkakshuset" (S. 205—212). Es ist anzunehmen, dass die hieraus gewonnenen Prozentzahlen für Texte dieses Stiles repräsentativ sind. (Siehe Tab. I, II, III, IV, Mikrofilm.)

Der Text wurde langsam und deutlich ausgesprochen, jedes Wort für sich, also nicht fliessend. Dadurch wurde eine willkürliche und wechselnde Satzbetonung vermieden und es konnte ein Versuch, die Vokale in druckstarke und druckschwache aufzuteilen, gemacht werden. (Tab. VI, S. 44). Für jedes Wort wurde nur eine druckstarke Silbe angenommen. Nebenakzente wurden nicht berücksichtigt, sondern zu den druckschwachen Silben gerechnet. Hiedurch fiel jedoch die Prozentzahl für druckschwache Vokale niedriger aus als in fliessender Rede. Die Aussprache war die in Uppsala gewöhnliche, man kann sie als Umgangssprache bezeichnen oder, mit A. Noreen²⁾ "familiären Mittelstil" (familiär mellanstil). Es wurde mit 17 Vokalphonemen plus 4 speziellen Varianten plus unbetontem ə und mit 18 Konsonantphonemen plus den 5 Supradentalen, von denen jeder einzelne auch als zwei Phoneme gewertet werden kann, gerechnet.

¹⁾ Herrn Dozent G. Hammarström spreche ich an dieser Stelle meinen aufrichtigen Dank für wertvolle Anregungen bei der Ausführung dieser Arbeit aus.

²⁾ A. Noreen, *Vårt Språk*, I., S 29.