

*The Acquisition of German Syntax by Foreign Migrant Workers*¹

Heidelberger Forschungsprojekt "Pidgin-Deutsch"

INTRODUCTION

This chapter deals with the main aims, methods, and some results of a project undertaken at the University of Heidelberg in which we try to analyze the undirected **natural** acquisition of German by Spanish and Italian migrant workers. Their language may be considered as a set of pidginized varieties of German.² Because it shows some structural and functional similarities to colonial pidgins, we call it **Pidgin-Deutsch** ('Pidgin-German'), using a term coined by Clyne (1968), the first paper on this topic. The use of this term should not be misinterpreted, however; we

¹This paper is an extended discussion of the work reported in Heidelberger Forschungsprojekt "Pidgin-Deutsch" (HPD, 1977). See also Becker, Dittmar, and Klein (1977), Dittmar and Rieck (1976, 1977), and HPD (1975a, b, 1976). For the general sociolinguistic background, see Dittmar 1976, Klein 1974. This chapter is based—as are all our papers—on the work of the whole research group: Angelika Becker, Norbert Dittmar, Margit Gutmann, Wolfgang Klein, Bert-Olaf Rieck, Gunter and Ingeborg Senft, Wolfram Steckner, and Elisabeth Thielicke. The present formulation is by Klein and Dittmar. We are very grateful to David Sankoff who carefully corrected the manuscript and made a number of valuable suggestions.

²For a discussion of the "pidgin problem" see HPD (1975a, Chapter 2) and the papers in Klein (1975), particularly Meisel (1975).

employ it as a more or less convenient label for a rather complex and unexplored phenomenon. Whether or not the German spoken by foreign migrant workers is a true pidgin is a difficult question, requiring a great deal more knowledge both about pidgins in general and about foreign migrant workers' German in particular before it can be answered. One point should be made clear right from the beginning: Pidgin-German is by no means a stable language but is a rather heterogeneous system of varieties. In this respect it does not differ from any other language—if the languages themselves are taken into consideration instead of just regularized descriptions of them.

In the following section something will be said about the project itself, about its aims, and particularly about the process of language acquisition (or, strictly speaking, second language acquisition) and how to model it. In the third section we will explain our descriptive framework as it concerns syntax. We have developed a particular way of describing variation, the central concept of this procedure being that of variety grammar. Transitional grammars are particular cases of variety grammars. In the fourth section, a short outline of some empirical aspects of our work will be given, and in the final section, we will informally present some of our major findings.

MODELING THE ACQUISITION PROCESS

In January 1976, the foreign migrant worker population in West Germany amounted to 4.1 million people (including family members), approximately 850,000 of whom come from Italy or Spain. Most of them do not know a word of German when they arrive, but in their daily living, they learn what is most urgently needed, some of them eventually achieving a certain fluency. This is clearly an important problem from both linguistic and social points of view. The miserable social situation of foreign workers is due not only to economic factors, such as insecurity of employment, low-prestige work, and so on, but also in large measure to a rather thorough exclusion from the local social and political life. With some exceptions, they form a class of their own or, strictly speaking, classes of their own, there being often great social distances between nationalities, for example, between Italian and Turkish workers.

The social isolation of foreign migrant workers is closely connected with their linguistic isolation. It would be ridiculous, of course, to posit this as the sole explanation, and it would be even more ridiculous to imagine that their social isolation could be corrected by improving only their language skills. Nevertheless, it seems evident that a reasonable

solution to the social problem is impossible without a solution to the language problem. Hence, helping these workers to improve their communicative competence in the widest sense of the term is a necessary though not sufficient condition, and it is to just this task that linguists can contribute—perhaps. It is our opinion that this cannot be done without a careful analysis of the current processes of acquisition of German by foreign workers and of the various social and individual factors governing this natural, undirected kind of language learning. Our research encompasses the phonological, morphological, syntactic, and pragmatic levels, but the following considerations are confined to the role of syntax only within the process of language acquisition.³

In the present context, the term **language acquisition** refers only to second language learning in a social environment where the language to be learned is spoken—second language learning without explicit teaching. Language acquisition in this sense is a rather slow process with many intermediate stages, each stage being characterized by a set of grammatical rules the speaker or the group of speakers masters at a given time. These sets may be considered as particular varieties of the second language, varieties that may be correct or ridiculous in the opinion of an average speaker of that language. A highly simplified description of the whole process of language acquisition would then be in terms of a step-wise approximation, passing through a series of intermediate varieties in the direction of a **target variety** (or target varieties, if there is internal variation in the language depending on social environment). In most cases, the target variety is never reached, though the whole process moves in its direction. The specific nature of the different speech varieties, their similarities and differences, and the trajectory of the process are governed by a set of extralinguistic factors such as:

1. Time (i.e., duration of stay)
2. Kind of job
3. Location
4. Origin (i.e., mother tongue or dialect)
5. Degree of social relationship (intensity of contact)
6. Family status
7. Mobility
8. Sex
9. Age (at time of immigration)
10. Education
11. Individual attitudes (e.g., motivation).

³Some general aspects of communicative behavior are discussed in HPD (1975a, Chapter 4) and in HPD (1975b).

There are additional factors that might be important, but those listed here will suffice for the present discussion. Taken together, they constitute both a learning context and an individual disposition; correspondingly, they can be subdivided into **environment factors** and **bias factors**. Each learner is characterized by a set of specifications of these factors which determine a complex system of diverse acquisition conditions. Let us neglect for the moment all possible factors but one, namely, duration of stay, and turn to the general problem of how to describe the process of language acquisition along the dimension of time. It should be emphasized, however, that the choice of this factor is for illustrative purposes only. Indeed, we have found that duration of stay is completely overshadowed in explanatory value by other factors after about 2 years.

Let us imagine that each speech variety can be described by a grammar, say, a transformational grammar or a simple context-free grammar. Each of these grammars, which we call **transitional grammars**, characterizes the variety situated at a certain point along the process of language acquisition. This yields a series of grammars along the dimension of time, for example:



The subdivision of time is set by the linguist, who may refine it to be relevant and interesting for current purposes. A key problem becomes how to describe the transition from one grammar to the next. The simplest way to depict the relationships between the grammars is as follows: Form the (set-theoretical) union of all rule sets, i.e., all rules occurring in at least one grammar, and then after each interval of time indicate whether or not the rule in question occurs.

This leads to a presentation such as:

Rule	G_1	G_2	G_3	G_4	G_5
r_1	+	+	+	+	+
r_2	+	+	+	+	+
r_3	-	-	+	+	+
r_4	-	+	+	+	+
r_5	-	-	-	-	+
r_6	+	+	-	+	+
r_7	-	-	-	+	+
r_8	-	-	+	-	+
r_9	-	-	+	-	+
r_{10}	+	+	-	-	-

In this fragmentary (and fictitious) example the transitional grammar G_1 contains the rules r_1 , r_2 , r_6 , and r_{10} ; 6 months later, the speaker (or the group of speakers) has learned another rule, r_4 ; 1 year later, we note that in G_3 four additional rules, r_3 , r_7 , r_8 , and r_9 , have been added and two previous rules, r_6 and r_{10} , have been dropped, and so on. This kind of description in terms of rule adding and rule dropping is well known; it suffers from at least two crucial inadequacies:

1. The description of transition in terms of sudden qualitative changes from minus to plus or plus to minus is rather inaccurate. In fact, there is often a long period of time in which rules co-occur and it is only a gradual shift in usage frequency that leads to the replacement of one rule by another. Hence, the description should make use of the whole continuum of real numbers between 0 and 1 instead of just + and -.

2. In the simple model presented above, only one dimension of variation is taken into account, namely, duration of stay. This is clearly inadequate. There is variation, too, according to factors like origin, kind of job, age at the time of immigration, and so on. If we accept that the 11 factors just cited can influence the process of language acquisition, we cannot assume a one-dimensional space of variation. An adequate representation of this variation may well require a multidimensional space.

In developing a model having both of these capacities, we have tried to fulfill one more condition: The model must be simple enough to be operationally applied to a large amount of data.

A DESCRIPTIVE FRAMEWORK FOR VARIATION⁴

Two basic concepts are **space of varieties** and **probabilistic grammars**. Probabilistic grammars, developed mainly by Grenander (1967), Suppes (1972), and Salomaa (1969) are simply formal grammars with an index associated with each rule giving the probability of rule application. The details depend on the type of grammar and whether conditioned probabilities are taken into account; we shall not discuss this here. A space of varieties is an analytic grid sufficient to distinguish all possible speech varieties thought to exist. Suppose in a given domain of investigation there are three relevant factors of variation:

1. Sex, with two possible values s_1 (male) and s_2 (female)
2. Age (at the time of immigration) with, say, four possible values a_1 , . . . , a_4 , where $a_1 = 20$ to 30 years, $a_2 = 31$ to 40 years, etc.
3. Duration of stay, with five possible values d_2 . . . , d_5 , where $d_1 =$ from 6 months to 1 year; $d_2 = 1$ to 2 years, etc.

⁴The concept of variety grammar was introduced and defined in Klein (1974).

This leads to $2 \times 4 \times 5 = 40$ possible varieties, each one defined by a triple of factors; e.g., (s_2, a_2, d_2) is the speech variety of a woman who immigrated a year or two ago between the ages of 30 and 40. This variety may well be identical from a linguistic point of view to some other variety, perhaps to (s_1, a_3, d_2) , but whether or not this is the case is an empirical question. The entire set of triples constitutes the space of varieties on which the investigation is based. Of course, one cannot be sure that the space of varieties constructed in this way contains only and all the relevant factors and distinctions. It is simply a hypothesis about the relevant determinants of variation in the domain under study.

Next, representative data must be obtained for each variety within the space of varieties, and a grammar must be written, or part of a grammar, if the interest is only in a particular linguistic problem. This may be a context-free grammar, a context-sensitive grammar, a dependency grammar, a transformational grammar, or whatever, but it must be clearly defined. The result is a set of n grammars, if there are n varieties. The next task is to interrelate these grammars by establishing a **reference grammar** consisting of the union of all rule sets of the particular grammars. This reference grammar describes nothing; it is merely a useful analytical construct. It can generate each variety by the association to each rule of an appropriate number between 0 and 1. This number indicates the probability of application of this rule in the variety in question. A given rule may be applied in a certain variety with probability .9, which means—informally speaking—that it is an important rule in that variety. In another variety, its probability may be .2, that is, it is less probable that it occurs in the derivation of a sentence. In a third variety, its probability may even be 0, that is, it does not occur at all in that variety. The same grammar then is used to describe all varieties in a given space of varieties, and the differences among them are expressed by the differences of rule values.

DATA COLLECTION AND ANALYSIS

Our grammatical description is essentially based on interview material, although in our project we have also made use of participant observation data to study phenomena such as code switching, linguistic expression of social relationship, and so on.

We decided to interview 48 persons and to stratify the sample as follows:

1. 32 men and 16 women (this reflects approximately the actual distribution)
2. 24 Italian and 24 Spanish workers

3. 12 workers from each of the following duration-of-stay categories
 - a. Up to 2 years
 - b. From 2 to 4 years
 - c. From 4 to 6 years
 - d. More than 6 years.

(All the other factors mentioned earlier in the second section are also registered.) An informant then may be considered to represent a collection of specified factors, to be representative of a certain variety.

The interview took the form of casual but gently directed conversations which were recorded on a two-track recorder (Uher 210, Lavalrière microphones). Special techniques, most of them from Labov, Cohen, Robins and Lewis (1968), were used to avoid communicative disturbances and deviations from usual communicative behavior. From each interview, 15 minutes were transcribed in a simplified phonetic notation. (Excerpts from our transcriptions are published as an appendix to HPD [1975a].)

The next step consisted in developing a reference grammar including all rules applying at least once among our 48 texts. The rules are context-free. A complete syntax would require the addition of word-order transformations, but the treatment of these rules has been postponed to a later phase of our study. The grammar we worked with, after six or seven revisions follows (Figure 1).

In order to facilitate reference and discussion, a list of categories was added, but it should be kept in mind that this is a formal grammar, and symbols like AC or PROP have meaning by virtue only of the rules in which they occur. The grammar contains 101 context-free rules in all, grouped into 15 rule clusters. A rule cluster consists of all rules with the same left-hand symbol, the rules within a cluster being alternative ways of rewriting this symbol. For instance, in rule cluster 3, VC may be rewritten as VG or as PVL, where VG leads to sentences with (finite) verb or copula and PVL to sentence without (finite) verb or copula.

This reference grammar was used to parse 100 sentences from each informant's interview, and hence to compile relative frequencies of rule applications within each cluster. The parsing was recorded in the form of labeled bracketing, facilitating the counting of rule applications. We interpret the relative frequencies of rule applications as estimates of the probabilities within the variety the 100 sentences represent. These estimates may not be too accurate in some cases, because the number of occurrences is sometimes too small to ensure that the figures would not significantly change were a larger number of sentences to be considered. Nevertheless, the relative frequencies after 50 sentences and then after 100 sentences per informant were essentially the same.

The result of this procedure may be imagined to be a matrix consisting

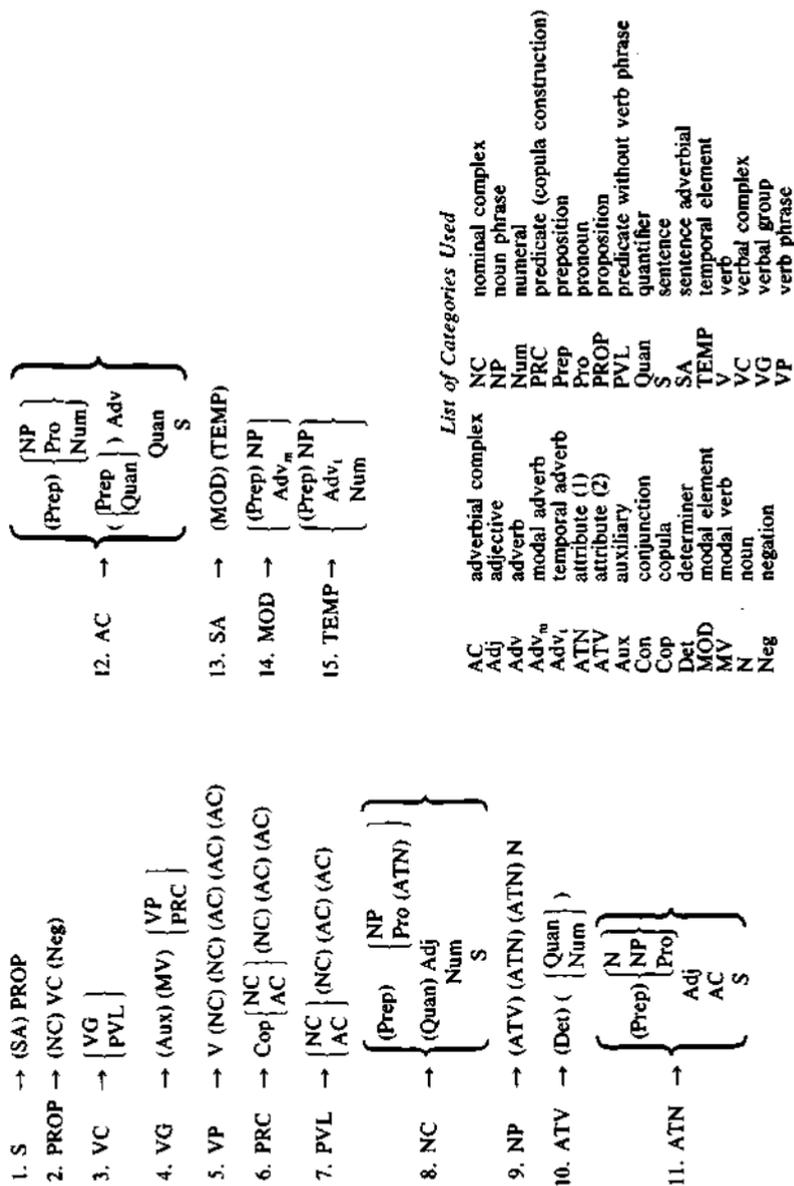


Figure 1. Reference grammar.

of 48 rows and 101 columns, where each row corresponds to an informant, or, strictly speaking, to a variety represented by that informant and each column to a rule of the reference grammar. Each cell of the matrix contains a number between 0 and 1, representing the application probability of the rule in the variety in question.

This basic matrix is a complete description of language variation during the process of language acquisition, insofar as this could be extracted from our data on 48 informants and as it is restricted to syntax and then only a part of syntax. It is not reprinted here because most of the rules show little or no variability, having similar or even identical values in all varieties. In some cases, this lack of variability may be caused by the small numbers of occurrences. Hence, for the next stage of the analysis we excluded (*a*) all rules with an average of less than 50 occurrences per informant and (*b*) all rules whose variability ranges over less than 40% of the possible range. (A rule's probability may range from 0 to 1. If it ranges from .3 to .5, it is excluded; if it ranges from 0 to .5, it is taken into account.) This somewhat arbitrary restriction is made to conserve time and effort, and should not materially affect the results. The remaining rules follow (Figure 2). Their probabilities for the 48 informants comprise the reduced matrix in Table 1.

Fine distinctions between different rules of the same cluster are neglected. For instance, the rules from 12.02 to 12.11 are collapsed together (Table 1). This matrix represents the syntax of our informants insofar as it is variable. It allows for a direct and precise study of the variability of particular rules and rule sets. Consider for example the values of rules 2.03 + 2.04 which represent the use of nominal complexes (NC) in subject position, or simply speaking the frequency with which sentence subjects are present. Informant SP-35 uses no subject noun (or pronoun) in more than 70% of all sentences, while Informant IT-01, the speaker closest to the standard dialect, always uses one; he applies the rule with a probability of 1. Or consider rule 3.01 which states one of the two possible ways to construct a simple sentence—with or without a (finite) verb or a copula. Whether a learner uses finite elements is surely an important fact of language acquisition. Informant SP-35 uses no verb (or copula) in most cases; he uses mainly predicative nouns, adverbs, or adjectives (in predicative function) without a copula. IT-01, on the other hand, uses verbs or copulas whenever possible. Between these extreme cases, there is a continuum of intermediate stages as evidenced by the spectrum of values in Table 1.

The values in a given row of the matrix indicate the informants' syntactic performance with respect to a certain set of rules. It seems interesting and useful to compute an overall syntactic index summarizing the infor-

2.01	PROP	→	VC	8.04	NC	→	Prep NP
2.02	PROP	→	VC Neg	8.05	NC	→	Prep Pro
2.03	PROP	→	NC VC	8.06	NC	→	Prep Pro ATN
2.04	PROP	→	NC VC Neg	8.07	NC	→	Adj
3.01	VC	→	VG	8.08	NC		Quan Adj
3.02	VC		PVL	8.09	NC	→	Num
4.01	VG		VP	8.10	NC		S
4.02	VG	→	Aux VP	9.01	NP	→	N
4.03	VG	→	MV VP	9.02	NP	→	ATV N
4.04	VG	→	Aux MV VP	9.03	NP	→	ATN N
4.05	VG	→	PRC	9.04	NP	→	ATV ATN N
4.06	VG	→	Aux PRC	9.05	NP	→	ATN ATN N
4.07	VG	→	MV PRC	9.06	NP	→	ATV ATN ATN N
4.08	VG	→	Aux MV PRC	10.01	ATV	→	Det
5.01	VP	→	V	10.02	ATV	→	Quan
5.02	VP	→	V NC	10.03	ATV	→	Num
5.03	VP	→	V NC NC	10.04	ATV	→	Det Quan
5.04	VP	→	V AC	10.05	ATV	→	Det Num
5.05	VP		V AC AC	12.01	AC		NP
5.06	VP	→	V AC AC AC	12.02	AC		Pro
5.07	VP	→	V NC AC	12.03	AC		Num
5.08	VP	→	V NC AC AC	12.04	AC		Prep NP
5.09	VP	→	V NC AC AC AC	12.05	AC		Prep Pro
5.10	VP	→	V NC NC AC	12.06	AC	→	Prep Num
5.11	VP	→	V NC NC AC AC	12.07	AC		Adv
5.12	VP	→	V NC NC AC AC AC	12.08	AC	→	Prep Adv
8.01	NC	→	NP	12.09	AC	→	Quan Adv
8.02	NC	→	Pro	12.10	AC	→	Quan
8.03	NC	→	Pro ATN	12.11	AC	→	S

Figure 2. Phrase structure rules.

mation given by all the individual rule values. There are several possible ways of doing this. For example, we might simply take an informant's average value for all rules, but for several reasons this is not too meaningful a procedure. The way we computed our syntactic index is more complicated and is explained in HPD (1976, Chapters 4 and 6). This index accords well with our intuitions about the syntactic elaborateness of our informants and the highest syntactic index values among them approach those of a group of native speakers we analyzed for the sake of comparison. Nevertheless, for the reasons developed earlier in the second section, we refrain from any theoretical interpretation of such a cumulative index. Here we use it merely to determine the order in which to present the informants.

Table 2 presents the correlation coefficients between some of the subrules used for the construction of the syntactic index (cf. Figures 1 and 2 and Table 1). The pattern which emerges from these values justifies to a

TABLE 1

Different Stages in 1 Acquisition of German by Migrant Workers: Prob-
 abilistic Values for ected Phrase Structure Rules and Syntactic Index

Informant	Rules								Syntactic index
	2.03 and 2.04	3.01	4.02 to 4.08	5.02 to 5.12	8.02 and 8.03	9.02 to 9.06	10.01	12.02 to 12.11	
IT-01	1.00	1.00	.83	.84	.61	.79	.88	.99	1.602
sp-11	.95	1.00	.89	.91	.60	.68	.82	.97	1.448
IT-31	.95	1.00	.60	.93	.54	.67	.80	.94	1.243
SP-29	.85	.99	.61	.89	.60	.72	.82	.91	1.234
SP-19	.94	1.00	.57	.83	.55	.82	.75	.92	1.233
IT-22	.84	.95	.42	.90	.64	.71	.69	.80	.961
SP-31	.89	.97	.47	.75	.47	.84	.79	.85	.956
IT-02	.74	.88	.47	.85	.55	.73	.79	.96	.934
SP-24	.85	.98	.47	.85	.70	.62	.46	.90	.831
SP-17	.94	.85	.41	.80	.47	.64	.80	.88	.765
IT-10	.73	.92	.45	.89	.48	.62	.72	.91	.709
IT-06	.82	.96	.51	.77	.53	.57	.55	.83	.527
IT-33	.77	.90	.60	.86	.43	.61	.50	.66	.375
IT-20	.70	.91	.24	.76	.52	.60	.58	.96	.374
SP-06	.61	.87	.38	.82	.51	.57	.58	.84	.291
SP-13	.72	.79	.15	.85	.42	.57	.61	.89	.208
IT-15	.63	.82	.27	.87	.35	.55	.67	.88	.200
SP-01	.62	.67	.40	.70	.51	.60	.68	.79	.135
SP-36	.65	.72	.21	.92	.33	.50	.59	.98	.116
IT-26	.57	.78	.12	.81	.49	.51	.69	.92	.112
IT-05	.73	.71	.17	.87	.40	.52	.69	.71	.059
IT-25	.64	.69	.42	.76	.49	.57	.58	.68	.031
IT-32	.64	.73	.07	.82	.40	.63	.57	.77	-.030
IT-18	.63	.65	.20	.81	.31	.76	.64	.62	-.033
SP-30	.72	.70	.07	.74	.35	.59	.62	.81	-.092
IT-16	.50	.75	.13	.71	.32	.67	.63	.92	-.092
SP-18	.61	.72	.04	.90	.43	.44	.67	.72	-.124
SP-26	.67	.66	.18	.75	.30	.46	.81	.81	-.128
IT-28	.55	.66	.11	.73	.41	.60	.67	.72	-.220
IT-07	.60	.68	.15	.80	.46	.42	.69	.66	-.225
SP-15	.66	.86	.16	.86	.41	.34	.35	.87	-.236
IT-29	.55	.69	.06	.57	.49	.52	.83	.74	-.265
SP-14	.62	.74	.05	.85	.34	.52	.59	.63	-.285
SP-12	.65	.65	.02	.67	.25	.73	.64	.73	-.290
IT-12	.56	.51	.18	.76	.45	.53	.50	.80	-.303
SP-09	.58	.76	.00	.76	.48	.58	.32	.68	-.384
IT-13	.65	.46	.02	.61	.35	.42	.71	.79	-.575
IT-09	.68	.43	.05	.82	.40	.41	.41	.66	-.610
SP-04	.55	.57	.00	.82	.26	.45	.57	.62	-.662
IT-23	.53	.41	.08	.69	.38	.60	.46	.56	-.727
IT-24	.52	.45	.13	.78	.24	.45	.59	.57	-.789
IT-08	.53	.21	.00	.95	.10	.52	.53	.64	-.851
SP-08	.45	.61	.15	.83	.33	.36	.26	.64	-.875
SP-02	.56	.42	.05	.70	.37	.48	.47	.33	-.997
SP-21	.42	.36	.00	.72	.23	.70	.52	.31	-1.035
SP-25	.50	.31	.00	.61	.31	.51	.24	.82	-1.064
SP-22	.72	.49	.00	.49	.30	.46	.42	.43	-1.126
SP-35	.29	.16	.00	.33	.15	.33	.05	.43	-2.396
Mean	.67	.71	.24	.78	.42	.57	.60	-.76	
Standard Deviation	.15	.22	.23	.12	.13	.12	.17	.17	

TABLE 2

Correlations (Pearson *r*) between (a) Some Important Subrules, (i) Subrules and Syntactic Index, and (a) Subrules and "Age at Time of Immigration" (A.T.I.) and "Duration of Stay" (D.S.)

Subrules	Subrules						Sociological variables	
	2.03 +2.04	3.01	4.02 +4.04	8.02	9.02, 9.04 +10.01	12.04, 12.05, 12.06 +12.08	A.T.I.	D.S.
2.03+2.04		-	-	-	-	-	-.40 ⁺	.07
3.01	.79 ⁺⁺						-.55 ⁺⁺	.16
4.02+4.04	.72 ⁺⁺	.62 ⁺⁺					-.38 ⁺	.23
8.02	.70 ⁺⁺	.81 ⁺⁺	.61 ⁺⁺				-.42 ⁺	.09
9.02+9.04 +10.01	.72 ⁺⁺	.65 ⁺⁺	.63 ⁺⁺	.50 ⁺⁺	-	-	-.48 ⁺⁺	.26
12.04+12.05 +12.06+12.08	.60 ⁺⁺	.66 ⁺⁺	.43 ⁺	.49 ⁺⁺	.48 ⁺⁺	-	-.42 ⁺	-.02
Syntactic index	.88 ⁺⁺	.92 ⁺⁺	.73 ⁺⁺	.80 ⁺⁺	.81 ⁺⁺	.65 ⁺⁺	-.56 ⁺⁺	.19

+ = significant at .01 level.

++ = significant at .001 level.

certain extent the construction of the syntactic index. The correlations of syntactic rules (and index) with two sociological variables (age at time of immigration [A.T.I.] and duration of stay [D.S.]) also shown in Table 2 will be discussed on pages 18-21.

Figure 3 illustrates the correlation of the rules NC->Pro (8.2) and VC->VG (3.01) in the form of a scattergram.

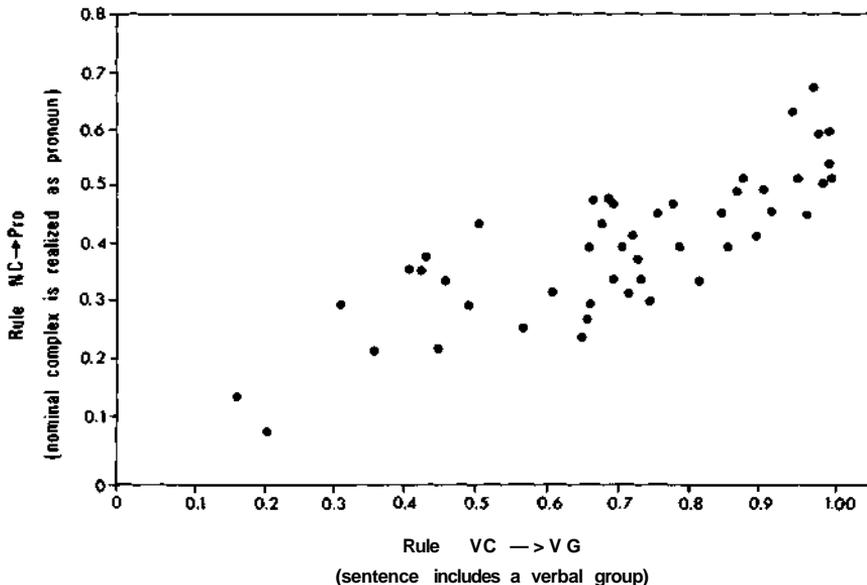


Figure 3. Correlation between rule NC -> Pro and rule VC -> VG (scattergram). Each point represents one informant, plotted according to frequency of the two rules.

OVERALL RESULTS AND CONCLUSIONS

Outline of Stages in the Acquisition of German Syntax

To obtain an accurate picture of syntactic development, the evolution of individual rules or rule clusters must be considered. We have made a detailed study of this kind, the results of which are summarized here in an informal and sketchy way. We concentrate here upon five principal areas of syntactic development: the structure of whole **propositions**, **verbal complexes (VC)**, **nominal complexes (NC)**, **adverbial complexes (AC)**, and **subordinate clauses**. Although the terminology we use differs somewhat from standard nomenclature, there should be no difficulty in understanding it.

1. The proposition
 - a. In the initial stage, propositions are formed without any finite element (verb or copula) and without a subject, e.g., /**kɪnda traɪ**/ '(I have) three children' or /**amə ta: amə ma futɪ**/, literally 'one day one mark fifty'.
 - b. The most advanced learner never uses propositions without a finite element or a subject; this corresponds to the usage of native speakers.
2. The verbal complex
 - a. The constituents of the verbal complex are learned in the following order: simple verb, copula, modal verb, auxiliaries. Combinations of modal verb, auxiliary + verb or copula are acquired very late.
 - b. In the early stages, verbs are complementized by only one nominal complex (direct or indirect object) or one adverbial complex. There is a very regular and steady increase in the number of complexes depending on the verb.
3. The nominal complex
 - a. Simple nouns (proper nouns, class nouns without an article or modifier, etc.) precede pronouns in order of acquisition.
 - b. In the beginning, noun phrases do not have any modifier or determiner. There is a steady process of elaborating complex noun phrases.
 - c. Within the class of determiners, there is a continuous shift from simple numbers (/svaɪ ma:/'two marks') and quantifiers (/ft: arbai/, 'much work') to articles, i.e., numbers and quantifiers predominate at first; articles occur mainly in later stages.
 - d. The first and most important adjectivals are adjectives. Prepositional phrases functioning adjectivally and relative clauses appear very late.
 - e. Nominal clauses ('that . . .', 'whether . . .') first appear in the middle stages.
4. The adverbial complex
 - a. The first adverbials are simple noun phrases without any preposition (/doɪsla/ 'in Germany, to Germany, for Germany'). This structure disappears rapidly. It is replaced by simple adverbs, prepositional phrases and adverbial clauses.
 - b. Prepositional phrases with nouns are learned before prepositional phrases with pronouns (/baɪ maɪn kole:ga/ 'with my colleague' before /baɪ ɪ:m/ 'with him').
5. Subordinate clauses

The acquisition of subordinate clauses shows a very clear and dis-

tinct order: Adverbial clauses are learned before nominal clauses, nominal clauses before relative clauses.

The whole process of the acquisition of syntax can then be described roughly as follows: The first utterances consist of simple or slightly expanded nominal complexes and/or adverbial complexes of a very simple kind. Then, the first finite verbs occur, sentences take on subjects, and the first pronouns are used. Verbal complexes and nominal complexes continuously increase in complexity during this process. Adverbial prepositional phrases and adverbs supplant simple noun phrases functioning as adverbials. Adverbial clauses, copulas, modal verbs, and adnominal prepositional phrases are learned. Only in the last stages is the expansion of verb or copula by auxiliaries and modal verbs learned. The same holds for the acquisition of nominal and relative clauses.

A Pattern of Overgeneralization: The Case of the Modal Verb *Müssen*

Overgeneralization of a specific form or rule of the target language during undirected second language learning process is a well-known fact. As an example, we will consider the acquisition of German modal verbs used as finite forms in connection with infinite verb forms. The co-occurrences of the finite modal and the infinite verb form are globally analyzed in our phrase structure grammar by rule 4.03 (see Figure 2, p. 10). For the application probability of this rule, we calculated the following values for the four groups of our sample⁵:

TABLE 3

Raw Scores and Application Probabilities for the Rule
VERBAL GROUP → MODAL VERB + VERB by Four Groups of Foreign Workers

Groups	Raw scores	Probabilities
I	4	.01
II	25	.03
III	94	.11
IV	72	.07

Table 3 shows a kind of crossover pattern for Group III. The rule is applied by this group with a rate substantially higher than not only Groups

⁵Each group includes 12 speakers. The sample was divided into groups on the base of the syntactic index. Consequently, Group IV covers IT-01 to IT-06, Group III IT-33 to IT-18, Group II SP-30 to SP-09, and Group I IT-13 to SP-35 (cf. Table 1). Group I shows the greatest and Group IV the smallest distance from the local Heidelberg vernacular.

I and II, but also Group IV, which corresponds most closely to the local vernacular and can be supposed to employ modal verbs in a wider and more differentiated range. In order to explain this crossover, it seems useful first of all to enumerate the lexical realizations of the modal verb rule. Altogether, we find the five modals *können*, *wollen*, *müssen*, *sollen*, and *mögen* appearing in a total of 195 sentences.

TABLE 4

Lexical Realization of Modal Verbs and Their Quantitative Proportions within the Rule MODAL VERB + VERB for Four Groups of Foreign Workers

Groups	Modal Verb					Total
	<i>mögen</i>	<i>sollen</i>	<i>müssen</i>	<i>wollen</i>	<i>können</i>	
I	–	–	–	3	1	4
II	–	–	13	9	3	25
III	–	1	75	12	6	94
IV	3	5	19	26	19	72
Total	3	6	107	50	29	195
Probabilities	.02	.03	.55	.26	.15	

Table 4 shows that (a) *müssen*, *wollen*, and *können* are the modals applied with greatest frequency; (b) *wollen* and *können* are applied at an earlier stage than *müssen*; (c) *mögen* und *sollen* are acquired very late; and (d) *müssen* contributes more than 50% to the applications of the rule. Most striking, however, is the fact that Group III applies *müssen* at a very high rate with 75 occurrences, 80% of their total modals. If we look at the values for Group IV, we note that *müssen*, *wollen*, and *können* are used in approximately the same proportion. The high value of *müssen* for Group III, which contrasts sharply with that of Group IV (cf. the crossover pattern of Figure 4), can then be supposed to be either a function of specific topics of discourse or of particular learner strategies. If the latter is the case, *müssen* acquires a broader meaning than standard German commonly allows.

A look at the interview passages of those learners in Group III who use the modal very often shows that particular verbal strategies rather than specific discourse topics seem to be responsible for the frequent application of *müssen*. We find that *muss* is apparently used as a substitute for morphological tense markers of the verb. It functions then as a feature of overgeneralization in the sense that it not only covers the meaning in standard German of *müssen* as an obligation to do something, but also the tense and aspect system of the verb.⁶

⁶The process of "overgeneralization" is discussed in Corder (1973, pp. 272-294).

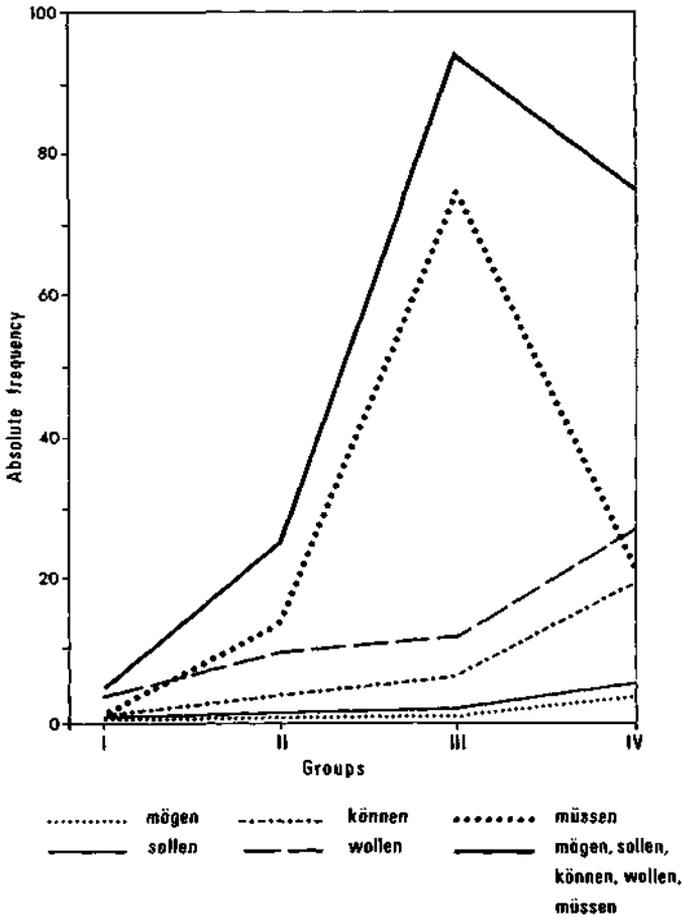


Figure 4. Crossover pattern.

The overgeneralized use of *müssen* can be explained in a preliminary way by **linguistic** and **sociological** arguments. Linguistically, *muss* functions as an **overgeneralized feature** and **covert error**. The application of *muss* here is overgeneralized because it covers a broader range of grammatical and semantic meaning than it does in German vernacular. The construction represents—in many cases—a covert error because the sentence is well formed on the surface, but erroneous in semantic deep structure (interpretation). The learner simply inserts *muss* in order to make the sentence acceptable, but the superficially correct formulation appears to be contextually and semantically inappropriate. The selection

of *muss* as a suitable formal expression of the learning strategy may be explained by the fact that *muss* has the comfortable property of having the same form in first and third person singular.

From a sociological point of view, the overgeneralized use of *müssen* has three interesting aspects. First of all, it seems to be the modal that foreign migrant workers hear most frequently in everyday communication, particularly at work. Though its use in lieu of the verbal tense and aspect system seems to be socially motivated, its significance is mainly linguistic—since the specific meaning of *müssen* as an expression of an obligation is unconsciously extended to other meanings. Thus, the concept of *müssen* has been only partially recognized. Second, the insertion of *muss* before the verb form serves to avoid the use of the unmarked verb forms which represent a socially stigmatized feature of the use of German by foreign migrant workers. The use of *muss* has then the function of increasing their social acceptability by "improving" their speaking behavior.

Finally, the excessive use of the *muss* + Verb construction reflects characteristics of a particular group of learners who have acquired a level of German which may not be very good from a normative point of view, but which is sufficient for the resolution of everyday communication problems. These learners have lived in Germany for 4 to 6 years and are between 20 and 30 years old. Their contact with Germans, their style of living, and their jobs seem to indicate that these learners are involved in a process of social adaptation. These results, however, must be considered preliminary, awaiting further data collection and analysis as must our observation that the overgeneralized use of *müssen* is more apparent among women than among men.

Social Correlates of the Syntactic Acquisition Process

Finally, we outline the connection between the syntactic performance in German of the 48 Italian and Spanish workers and the social environment. In order to isolate factors which favor or hinder the process of second language learning, we have correlated the syntactic indices of the 48 informants with extralinguistic variables (cf. points 1-11 on p. 3). Correlations of the syntactic data with extralinguistic parameters show that the acquisition process is governed by the following six variables listed here in the order of their decreasing influence:

1. Contact with Germans during leisure time ($\eta = .64$)⁷

" η " describes the correlation between an independent nominal variable and a dependent metrical variable: "In this case, ' η ' is the most sensitive correlation coefficient [Benninghaus, 1974, p. 230]." This correlation coefficient turned out to be the most appropriate for the kind of data we are dealing with.

2. Age at time of immigration ($\eta = .57$; $r = .56$)
3. Contact with Germans at place of work ($\eta = .53$)
4. Professional training in the country of origin ($\eta = .42$)
5. Education (years of attendance at school) ($\eta = .35$; $r = .33$)
6. Duration of stay ($\eta = .28$; $r = .20$).

"Contact with Germans" and "age at time of immigration" seem then to be the most important factors governing the level of second language

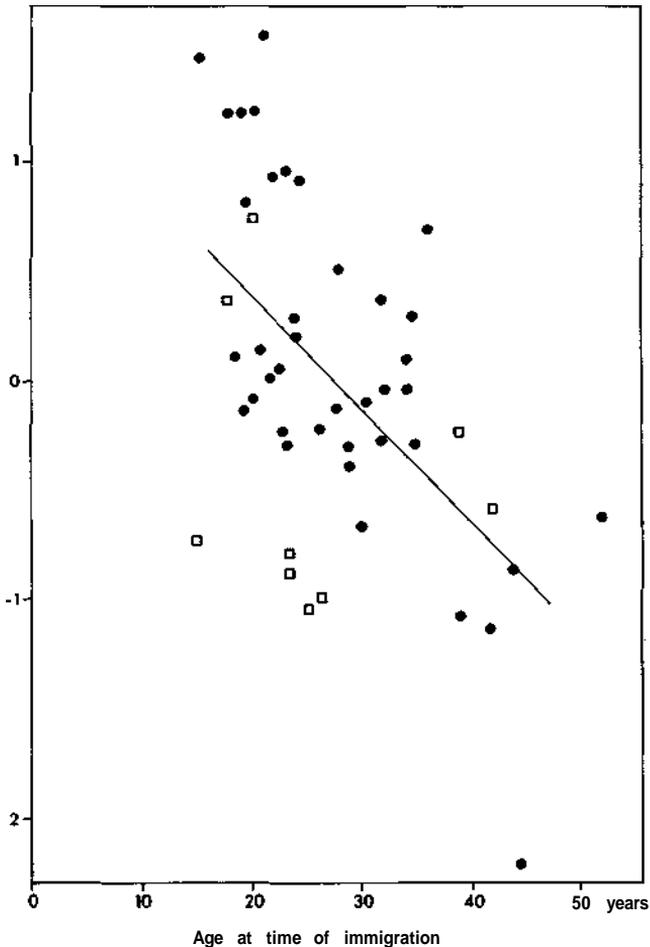


Figure 5. Scatter diagram: 48 foreign workers according to syntactic index and age at time of immigration. Informants with a duration of stay of less than 2.4 years are marked with open squares.

learning performance. At the beginning of our study, we hypothesized that "duration of stay" would play a major role in the process of acquisition of the second language. It turns out that this factor is only significant for the first 2 years of stay. After this, its effect is overridden by other social factors.

Figures 5 and 6 give some idea of these correlations and illustrate the differences between the correlation of the syntactic index with "age at the

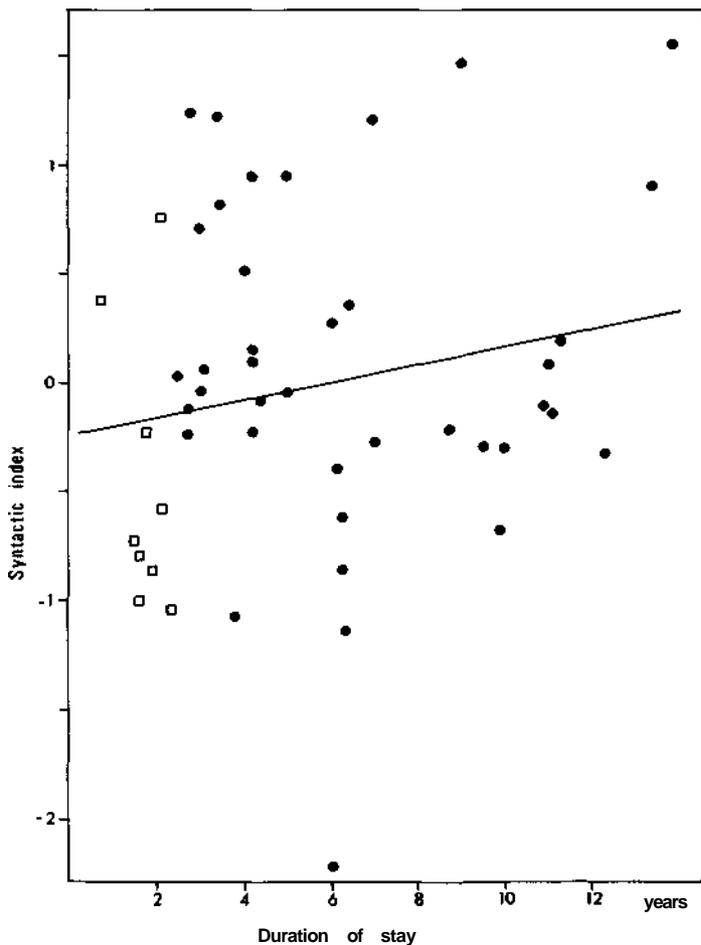


Figure 6. Scatter diagram: 48 foreign workers according to syntactic index and duration of stay. Informants with a duration of stay of less than 2.4 years are marked with open squares.

time of immigration" and "duration of stay." Correlation coefficients between the syntactic index and subrules, on one hand, and "age at time of immigration" and "duration of stay," on the other, are found in Table 2.

REFERENCES

- Becker, A., Dittmar, N., & Klein, W. Sprachliche und soziale Determinanten im kommunikativen Verhalten ausländischer Arbeiter. In U. Quasthoff (Ed.), *Sprachstruktur-Sozialstruktur. Zur linguistischen Theoriebildung*. Kronberg/Ts: Scriptor Verlag, 1977.
- Benninghaus, H. *Deskriptive Statistik*. Stuttgart: Teubner, 1974.
- Clyne, M. Zum Pidgin-Deutsch der Gastarbeiter. *Zeitschrift für Mundartforschung*, 1968, 35, 130-139.
- Corder, P. *Introducing applied linguistics*. Harmondsworth: Penguin, 1973.
- Dittmar, N. *Sociolinguistics. A critical survey of theory and application*. London: Arnold, 1976.
- Dittmar, N., & Rieck, B.-O. Reihenfolgen im ungesteuerten Erwerb des Deutschen. Zur Erlernung grammatischer Strukturen durch ausländische Arbeiter. In R. Dietrich (Ed.), *Aspekte des Fremdspracherwerbs. Beiträge zum 2. Fortbildungskurs 'Deutsch als Fremdsprache', Heidelberg 1975*. Kronberg/Ts: Scriptor Verlag, 1976, 119-145.
- Dittmar, N., & Rieck, B.-O. Datenerhebung und Datenauswertung im Heidelberger Forschungsprojekt 'Pidgin-Deutsch ausländischer Arbeiter'. In U. Bielefeld, E. W. B. Hess-Lüttich, & A. Lundt (Eds.), *Soziolinguistik und Empirie. Beiträge zum Berliner Symposium 'Corpusgewinnung und Corpusauswertung'*. Frankfurt: 1977, 59-96.
- Grenander, U. Syntax-controlled probabilities. Technical report, Brown University, Division of Applied Mathematics, 1967.
- Heidelberger Forschungsprojekt "Pidgin-Deutsch" (HPD). *Sprache und Kommunikation ausländischer Arbeiter. Analysen, Berichte, Materialien*. Kronberg/Ts: Scriptor Verlag, 1975. (a)
- Heidelberger Forschungsprojekt "Pidgin-Deutsch" (HPD). Zur Sprache ausländischer Arbeiter: Syntaktische Analysen und Aspekte des kommunikativen Verhaltens. *Zeitschrift für Literaturwissenschaft und Linguistik*, 1975, 5 (18), 78-121. (b)
- Heidelberger Forschungsprojekt "Pidgin-Deutsch" (HPD). *Untersuchungen zur Erlernung des Deutschen durch ausländische Arbeiter. Research report III*. Germanistisches Seminar, University of Heidelberg, 1976.
- Heidelberger Forschungsprojekt "Pidgin-Deutsch" (HPD). Transitional grammars in the acquisition of German by Spanish and Italian workers. In Jürgen M. Meisel (Ed.), *Langues en contact—Pidgins—Creoles—Languages in contact*. Tübingen, 1977.
- Klein, W. *Variation in der Sprache. Ein Verfahren zu ihrer Beschreibung*. Kronberg/Ts: Scriptor Verlag, 1974.
- Klein, W., Ed. *Sprache ausländischer Arbeiter. Zeitschrift für Literaturwissenschaft und Linguistik*, 1975, 5 (18).
- Labov, W., Cohen, P., Robins, C., & Lewis, J. *A study of the non-standard English of Negro and Puerto Rican speakers in New York City* (2 vols.). Final Report, Cooperative Research Project 3288. Washington, D. C.: U.S. Office of Health, Education, and Welfare, 1968.
- Meisel, J. M. Ausländerdeutsch und Deutsch ausländischer Arbeiter. Zur möglichen Ent-

stehung eines Pidgins in der BRD. *Zeitschrift für Literaturwissenschaft und Linguistik*, 1975, 5 (18), 9-53.

Salomaa, A. Probabilistic and weighted grammars. *Information and Control*, 1969, 15, 529-544.

Suppes, P. Probabilistic grammars for natural languages. In D. Davidson & G. Harman (Eds.), *Semantics of natural language*. Dordrecht: Reidel, 1972, 741-762.