CONSONANT COMBINATORICS IN GERMAN

MART RANNUT

Dept. of Computational Linguistics
Institute of Language and Literature
Tallinn, Estonia, USSR 200106

ABSTRACT

In the paper we present a list of possible (existing as well as typologically relevant non-existing) consonant sequences in German and their unified phonological distributional description. The works of L.Hirsch-Wierzbicka /7/ and V.Tarамets's group /26/ were used as source material for sketching out phonotactic rules, methodologically the work is based on Rannut /17/. The initial version was prepared as a computer program which was reprocessed and improved to obtain a better final result. The aim of the work was to find a set S' maximally similar to the set S of German consonant sequences, fulfilling the condition $|S-S'|<\epsilon$, where ϵ is minimal. The final version of phonotactic restrictions is presented in the form of a context-sensitive grammar.

1. INTRODUCTION

Consonant sequences in German and regularities of their formation have previously been studied and described by Twaddell /24/, Menzerath /11,12/, Moulton /13, 14/, Trubetzkoy /23/, Seiler /19,20/, Tanaka /22/, Hirsch-Wierzbicka /17/ and Taranets's group /26/. Bierwisch /1/, Wurzel /25/ and Copeland /13/ have described them by means of generative grammars. In addition, we have used the data provided by Kelz /9/, Stock /21/ and Kästner /10/.

The main research object of Twaddell was the structure of German words consisting of one syllable. He provided the rules of word-initial and word-final consonant clusters. When studying the structure of consonant sequences in the middle of the word (he found 394 of them) he came to the conclusion that their structure follows certain regularities. Menzerath followed the work of Twaddell in studying monosyllabic words, Moulton and Trubetzkoy studied certain aspects of them (Moulton studied the correlation between the structure of a consonant cluster and the length of the preceding vowel, Trubetzkoy formulated the distributional rules of the word-initial consonant cluster), Seiler constructed its structural formula and the work was completed by Hirsch-Wierzbicka, who applied a computer and added a functional load to every case. After Twaddell the work in studying word-internal consonant sequences was continued by Tanaka. He examined the structure of intermediate clusters, which he defined as clusters between two vowels, not taking into account word boundary signals. From this material he drew 5 statistically significant conclusions.

Taranets has studied the same problem with an aim of finding articulatory correlates to the syllable boundary. Bierwisch and Wurzel have studied certain morphonological aspects of the problem while Copeland has given separate distributional rules for coda and onset.

In our paper we deal with the word-internal consonant sequence as a unit with its own structure. Word-initial and word-final consonant clusters will not be thoroughly discussed as there are above-mentioned fundamental works available on this topic. Neither are those clusters difficult to generate when applying a few additional rules to the word-internal consonant sequences (see p.9).

2. PRINCIPLES OF DESCRIPTION

The principles of the description of consonant sequences are as follows:
1) The composition and occurrence of German consonant sequences are determined by regularities, part of which are general linguistic while the rest are characteristic of the phonological system of the German language, the latter being primary with regard to the general linguistic ones.
2) Syllable structure is regarded as primary with respect to phonemes and capable of dictating phoneme sequences in it (see /5/). The consonant sequences containing a syllable boundary are restricted by the syllable structure as well as by consonant sequence constraints. Such a model can be considered as a submodel of the model of the rhythmic pattern of speech (see /18/). As the structure of the syllable is influenced by the morphological component, the deviations resulting from this are regarded as a unnatural heterogenous set Q_n (see p.7).

3) In working out an integral model of phonotactic restrictions we have proceeded from word-internal consonant sequences as they are more regular than word-initial and word-final consonant clusters. By extending the restrictions of word-internal consonant sequences to word-initial and word-final consonant clusters and applying a few additional rules we have obtained a unified system of phonotactic restrictions of German consonant sequences (cf./15,16/).
4) Unlike the above studies which have dealt with the analysis of consonant sequences with a view to discovering the rules lying as the basis for the structure, this investigation aims at a synthesis of consonants by means of a generative model, the purpose of which is to obtain an adequate final result with the help of a minimal number of rules.
5) Considered in the study are only consonant sequences occurring in native German words while the difference between a German word and a foreign one has been made on the basis structure (see /6/).
6) The structure model of consonant sequences contains only predominating rules. As the phonological component is constantly affected by the environment through the adoption of foreign words and the phonetic peculiarities of other languages, added to its grammar are more and more rules which can be opposed to the existing ones. At the same time the language dislodges part of the rules to retain the level of homogeneity necessary for its existence. For that reason from the synchronic point of view a language is a complex of rules opposed to each other and some of them dominate over the others. In the case of opposing rules, considered in constructing the phonological generator are the rules which dominate over the others. The choice of rules in this case has been made with the help of functional load.
7) The present phonotactic description is based on Duden's /4/ transcription, containing the following consonant phonemes:

/p t k
b d g
pf ts tʃ
dʒ
f s ç x h
v z ʒ
m n ŋ
l
r
j/

As the consonants dʒ, tʃ and ʒ do not appear in native German consonant clusters, they are not taken into account in the inventory of our generative grammar.
8) The present work is not concerned with consonant sequences generated at compound component junctions, regarding only sequences in the words consisting of stem and affix(es), The latter are taken from Duden/4/

3. DEFINITIONS

A consonant sequence is an arbitrary sequence of different consonants within the boundaries of one simple word.

A consonant cluster is a syllable-internal consonant sequence. A syllable is a sequence of phonemes between two successive syllable boundaries within one word. Syllable boundary is fixed according to traditional German grammar.

A fortis structure is a consonant sequence containing stop consonants and/or affricates.

A lenis structure is a consonant sequence containing neither stop consonants nor affricates.

A base structure is a sequence of non-terminal symbols.

A surface structure is a sequence of German consonant phonemes.

A pronunciation strength structure is a sequence of non-terminal symbols marking pronunciation strength classes (see /8/).

4. GENERAL PRINCIPLE OF THE GENERATOR

A generator of consonant sequences represents a formal grammar which determines one phonological subsystem of a language - the phonotactics of consonant sequences - and gives its exact description. According to Chomsky's /2/ grammar the consonant sequences generator can be described by the formula D=(I,E,T,P), where D is the consonant sequences generator, E - the ultimate number of non-terminal symbols (the auxiliary symbols of the base structure), T - the number of terminal symbols (the consonant phonemes of the surface structure) different from E, P - the ultimate number of restriction (X) and derivation rules (Y), which describe the process of generating consonant sequences and restrict sequences not characteristic of the language, and I - the initial symbol.

E comprises the auxiliary symbols of the first- and second-level base structures (M1,K2,M3,K4,M5, 1,2,3,4,5, G,L,N,V,F, K,A,B,R,Q_1,Q_2). T comprises the consonant phonemes occurring in sequences in the surface structure(the 3rd level) of the German language (k,p,t,g,b,d,pf,ts,s,f,ʃ,ç,x,rm, h,v,z,m,n,ŋ,l,r,j,rn) and a blank or empty string (ϵ). Different variants are separated in the description by the symbol (,). The length of a consonant sequence, i.e. the number of phonemes in the sequence between two vertical strokes is marked by a number following the mark of equation. The symbol C stands for any consonant phoneme, * is a string of consonants comprising from 0 to 5 phonemes and $*_1$ is a string of consonants comprising from 1 to 5 consonant phonemes.

The generator has been compiled on the basis of the combinatorical regularities of German consonant sequences and it constructs the existing consonant sequences whose structure is acceptable to the phonological system of the German language as well as the non-existing consonant sequences which are typologically relevant to the existing ones. It also determines the position of syllable boundary. The work of the generator is based on the application of the hierarchical character of the phonological system of the language where restrictions are applied on all levels. This makes it possible with few restrictions to obtain from 5.5 million potential sequences ($\prod_{k=1}^{5} C_k = 23*23*23*22*21 = 5621154$) a result that in number approximately corresponds to the one in reality. The generator does not pretend to psychological reality or "natural" processes in the language, but represents a black-box-type model in which the application of phonological rules provides a result close to linguistic realities.

5. FORMAL DESCRIPTION OF THE GENERATOR

The formal description of the generator of German consonant sequences is given in the form of a grammar where Y marks the derivation rules and X the restriction rules, and the number following them denotes the hierarchical level. The number in brackets refers to the subsection where the respective operation is presented in more detail. The rules with the number 0 (e.g. X-1.0) point out those consonant strings which do not correspond to the definition of consonant sequences.

Y-1	$I \rightarrow 1\ 2\ 3\ 4\ 5$
(6)	$1 \rightarrow M_1, \epsilon$
	$2 \rightarrow K_2, \epsilon$
	$3 \rightarrow M_3, \epsilon$
	$4 \rightarrow K_4, \epsilon$
	$5 \rightarrow M_5, \epsilon$
X-1.0	$C \rightarrow \epsilon$
Y-2	$M_1 \rightarrow L,N,Q_1$
(7)	$K_2 \rightarrow K,A$
	$M_3 \rightarrow F,R$
	$K_4 \rightarrow K,A,B$
	$M_5 \rightarrow G,L,N,V,Q_2$
X-2.1.1	$Q_1{*}\ Q_2 \rightarrow \epsilon$
2	$*A*\ A* \rightarrow \epsilon$
3	$*BQ_2 \rightarrow \epsilon$
	$*BV \rightarrow \epsilon$
X-2.1.4	$*FB* \rightarrow \epsilon$
	$*AB* \rightarrow \epsilon$
	$*KB* \rightarrow \epsilon$
X-2.2.1	$Q_1*R* \rightarrow \epsilon$
	$N*R* \rightarrow \epsilon$
	$*KR* \rightarrow \epsilon$
	$*AR* \rightarrow \epsilon$
2	$*RK* \rightarrow \epsilon$
	$*RB* \rightarrow \epsilon$
	$*RA* \rightarrow \epsilon$
X-2.3.1	$G* \rightarrow \epsilon$
2	$*CCG \rightarrow \epsilon$
3	$BG \rightarrow \epsilon$
	$AG \rightarrow \epsilon$
	$RG \rightarrow \epsilon$
	$FG \rightarrow \epsilon$
X-2.4	$*FA* \rightarrow \epsilon$
X-2.5	$Q_1AK* \rightarrow \epsilon$
	$Q_1KA* \rightarrow \epsilon$
	$Q_1KK* \rightarrow \epsilon$
X-2.6.1	$C_1C_2C_3C_4* \rightarrow \epsilon$, if $C_n = B,A$
2	$C_1C_2C_3C_4C_5 \rightarrow \epsilon$ if $C_5 = N,Q_2,V$- if $C_1 = Q_1$
Y-3.1	$G \rightarrow j$
(8)	$L \rightarrow l,r$
	$N \rightarrow$ n,m,ŋ,rn,rm
	$V \rightarrow$ v,z,h
	$R \rightarrow$ l,m,n
	$F \rightarrow$ s,f,ʃ,x
	$B \rightarrow$ b,g,d
	$A \rightarrow$ pf,ts
	$K \rightarrow$ k,p,t,ç
X-3.0	$*C_iC_j* \rightarrow \epsilon$, if $i=j$
X-3.1	*Cŋ $\rightarrow \epsilon$
	$*Cx* \rightarrow \epsilon$
X-3.2.1	ŋC $\rightarrow \epsilon$, if $C \neq$ v,z,h,l,n,k
2	ŋ$C_1C_2* \rightarrow \epsilon$, if $C_1 \neq$ k,s
X-3.3	$xC* \rightarrow \epsilon$, if $C \neq$ t,ts
X-3.4	Cç* $\rightarrow \epsilon$, if $C \neq$ l,r,n
X-3.5	$*kf* \rightarrow \epsilon$
	$*pf* \rightarrow \epsilon$
	çf $\rightarrow \epsilon$
	$*tf* \rightarrow \epsilon$
	$*tsf* \rightarrow \epsilon$
	$*pff* \rightarrow \epsilon$
	$*sf* \rightarrow \epsilon$
	$*mf* \rightarrow \epsilon$
X-3.6	$Q_1fCC \rightarrow \epsilon$
X-3.7	$*gm \rightarrow \epsilon$
	$*bm \rightarrow \epsilon$
	$*dm \rightarrow \epsilon$
	$*km \rightarrow \epsilon$
	$*pm \rightarrow \epsilon$
	$*fm \rightarrow \epsilon$
	$*fm \rightarrow \epsilon$
	$*tsm \rightarrow \epsilon$
	$*pfm \rightarrow \epsilon$

Se 63.1.3

X-3.8	mʃC* $\rightarrow \epsilon$
	$*Csm \rightarrow \epsilon$
	$*sp* \rightarrow \epsilon$, if $* \neq \epsilon$
X-3.9	$*CCpn \rightarrow \epsilon$
	$*CCkn \rightarrow \epsilon$
X-3.10	$*tsp* \rightarrow \epsilon$
	$*pfp* \rightarrow \epsilon$
	$*kp* \rightarrow \epsilon$
	$*tp* \rightarrow \epsilon$
	çp $\rightarrow \epsilon$
	$*fp* \rightarrow \epsilon$
X-3.11	$mpC_1C_2 \rightarrow \epsilon$, if $C_2 \neq t$
	$lpC_1C_2 \rightarrow \epsilon$, if $C_2 \neq t$
	$rpC_1C_2 \rightarrow \epsilon$, if $C_2 \neq t$
X-3.12	$mtCC \rightarrow \epsilon$
	$Q_1tCC \rightarrow \epsilon$
X-3.13	$Cpt* \rightarrow \epsilon$
	Cçs* $\rightarrow \epsilon$
X-3.14	$*tk* \rightarrow \epsilon$
	$*pk* \rightarrow \epsilon$
	$*pfk* \rightarrow \epsilon$
	$*tsk* \rightarrow \epsilon$
	ʃk $\rightarrow \epsilon$
	$*fk* \rightarrow \epsilon$
X-3.15	$*sk* \rightarrow \epsilon$, if $* \neq e$
	çk $\rightarrow \epsilon$, if $* \neq e$
X-3.16	$Cpf* \rightarrow \epsilon$, if $C \neq m,Q_1,r,n$
X-3.17	$rpfC \rightarrow \epsilon$, if $C \neq l,r$
	$npfC \rightarrow \epsilon$, if $C \neq l,r$
	$Q_1pfC \rightarrow \epsilon$, if $C \neq l,r$
X-3.18	$CCs* \rightarrow \epsilon$
	$CCts* \rightarrow \epsilon$
X-3.19	$C_1tsCC \rightarrow \epsilon$, if $C_1 \neq n,r,l$
X-3.20	$*tts* \rightarrow \epsilon$
	$*kts* \rightarrow \epsilon$
	$mts* \rightarrow \epsilon$
X-3.21	*çr $\rightarrow \epsilon$
	$*sr \rightarrow \epsilon$
	$*tsr \rightarrow \epsilon$
X-3.22	$C_1C_2C_3r \rightarrow \epsilon$ if $C_2 \neq$ ʃ
X-3.23	*kʃ* $\rightarrow \epsilon$
	pʃ $\rightarrow \epsilon$
	tʃ $\rightarrow \epsilon$
	çʃ $\rightarrow \epsilon$
	tsʃ $\rightarrow \epsilon$
	pfʃ $\rightarrow \epsilon$
X-3.24	$Q_1s* \rightarrow \epsilon$
	$ms* \rightarrow \epsilon$
	$*tss* \rightarrow \epsilon$
X-3.25	$rsC* \rightarrow \epsilon$, if $C \neq t$
X-3.26	$mg* \rightarrow \epsilon$
X-3.27	$*pfz \rightarrow \epsilon$
	$*tsz \rightarrow \epsilon$
	*ʃz $\rightarrow \epsilon$
	$*fz \rightarrow \epsilon$
	$*sz \rightarrow \epsilon$
	$*pz \rightarrow \epsilon$
	$*tz \rightarrow \epsilon$
	*çz $\rightarrow \epsilon$
X-3.28	$*pfv \rightarrow \epsilon$
	$*fv \rightarrow \epsilon$
	$*pv \rightarrow \epsilon$
	$*tv \rightarrow \epsilon$
	*çv $\rightarrow \epsilon$
X-3.29	*ʃCl $\rightarrow \epsilon$
	*ʃCn $\rightarrow \epsilon$
	*ʃCQ $\rightarrow \epsilon$
X-3.30	$*pfsC \rightarrow \epsilon$, if $C \neq t$
	$*tsC \rightarrow \epsilon$, if $C \neq t$
	$rnC \rightarrow \epsilon$, if $C \neq t$
X-3.31	$ktC \rightarrow \epsilon$
	$ptsC \rightarrow \epsilon$
	$pftC \rightarrow \epsilon$
X-3.32	$npC* \rightarrow \epsilon$, if $C=s,n,Q_2$
X-3.33	$Q_1Cn \rightarrow \epsilon$, if $C=$ t,p,ts,f,b,d
X-3.34	$Q_1Cl \rightarrow \epsilon$, if $C=t,d$
X-3.35	$npn \rightarrow \epsilon$
X-3.36	$rmCCC \rightarrow \epsilon$
	$rmC \rightarrow \epsilon$
X-3.37	$rmC_1C_2 \rightarrow \epsilon$, if $C_1 \neq s$
	$rnC_1C_2 \rightarrow \epsilon$, if $C_1 \neq s$
X-3.38	$C_1C_2C_3C_4 \rightarrow \epsilon$, if $C_1 = k,t$
X-3.39	$C_1C_2C_3C_4C_5 \rightarrow \epsilon$,
	if 1. $C_n =$ n,x,ç,f,ŋ,m,v,Q_n,k
	v 2. $C_1 = s$
	v 3. $C_2 = t$
X-3.40	$C_1C_2C_3C_4C_5 \rightarrow \epsilon$, if $C_5 \neq l$
Y-3.2	$Q_1 \rightarrow$ s,f,nt etc.
	$Q_2 \rightarrow$ ç,b,h,t,f,.etc.

Se 63.1.4