ROUTLEDGE REVIVALS

# Morphophonemics of Modern Hebrew

Noam Chomsky

ROUTLEDGE REVIVALS

# Morphophonemics of Modern Hebrew

Noam Chomsky



Routledge
Taylor & Francis Group
Routledge Revivals

First published in 1979
by Garland Publishing, Inc

This edition first published in 2011 by Routledge
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
711 Third Avenue, New York, NY 10017

*Routledge is an imprint of the Taylor & Francis Group, an informa business*



**Publisher's Note**
The publisher has gone to great lengths to ensure the quality of this reprint but points out that some imperfections in the original copies may be apparent.

**Disclaimer**
The publisher has made every effort to trace copyright holders and welcomes correspondence from those they have been unable to contact.

A Library of Congress record exists under ISBN: 0824096886

ISBN 13: 978-0-415-68792-8 (hbk)
ISBN 13: 978-0-203-35745-3 (ebk)

# Morphophonemics of Modern Hebrew

Noam Chomsky

Garland Publishing, Inc. ▪ New York & London
1979

**Library of Congress Cataloging in Publication Data**

Chomsky, Noam.
Morphophonemics of modern Hebrew.

(Outstanding dissertations in linguistics ; 12)
A revision of the author's thesis (M.A.),
University of Pennsylvania, 1951.
Bibliography: p.
1. Hebrew language—Morphophonemics.
I. Title. II. Series.
PJ4600.C58 1979 492.4'5 78-66579
ISBN 0-8240-9688-6



All volumes in this series are printed on acid-free,
250-year-life paper.
Printed in the United States of America

December, 1951 Revision of

Morphophonemics of Modern Hebrew

Avram Noam Chomsky

"A Thesis

In Linguistics

Presented to the Faculty of the

Graduate School of the University of

Pennsylvania in partial fulfillment of

the requirements for the degree of

Master of Arts"

June, 1951

TABLE OF CONTENTS

## 0. Introduction

A grammar of a language must meet two distinct kinds of criteria of adequacy. On the one hand it must correctly describe the 'structure' of the language (i.e., it must isolate the linguistic units, and, in particular, must distinguish and characterize just those utterances which are considered 'grammatical' or 'possible' by the informant, including as a special subclass those of the analyzed corpus. On the other hand it must meet requirements of adequacy imposed by its special purposes (e.g., pedagogical, as a basis for comparative study, etc.), or, in the case of a linguistic grammar having no such special purposes, requirements of simplicity, economy, compactness, etc.[1] Thus the linguistic analysis of a language *L* can be described as the process of determining the set of 'grammatical' or 'significant' sentences of *L* (i.e., of determining the extension of the predicate 'grammatical in *L*'), or, in other words, it is the process of converting an open set of sentences--the linguist's incomplete and in general expandable corpus--into a closed[2] set--the set of grammatical sentences--and of characterizing this latter set in some interesting way. Accordingly we might distinguish and consider separately two aspects of the linguistic analysis of a language, a process of 'discovery' consisting of the application of the mixture of formal and experimental procedures constituting linguistic method, and a process of 'description' consisting of the construction of a grammar describing the sentences which we know from step one to be grammatical, and framed in accordance with the criteria related to its special purposes.

Although the distinction between the processes of discovery and description is clear enough in the case of grammars with special purposes, it is perhaps less clear in the case of a linguistic grammar constructed solely in accordance with considerations of elegance, since the process of discovery itself can perhaps best be understood as the process of constructing a tentative grammar specifying the grammatical sentences by listing the linguistic elements and their permitted arrangements on various levels.[3] Furthermore it is clear that considerations of elegance are operative in the original process of discovery, i.e., that they have a distinct place in the framing of the procedures of linguistics themselves. Thus in setting up such linguistic elements as morphemes (a process of discovery) we must consider properties of the linguistic elements themselves (e.g., perhaps minimization of their number) and properties of the statements describing these elements and their relationships (e.g., perhaps minimization of their number),[4] and the same is true on other levels of linguistic analysis. This consideration amounts to the requirement that the predicate 'grammatical in *L*' (and in general, the procedures of linguistic analysis) be defined and analyzed in a metalanguage to the language in which grammars are written, and consequently in a meta-metalanguage to the language L under analysis. Thus one of the considerations involved in setting up linguistic elements in a particular way, and consequently, in determining what are in fact the grammatical sentences, will be the total simplicity of the grammar in which these elements appear.

However it will still be useful to consider the processes of discovery and description separately. For the most reasonable way to approach the investigation and analysis of the notions of simplicity in terms of which 'grammatical in L' is defined (i.e., those notions of elegance that are relevant to the very formulation of the procedures of linguistics) seems to be to assume, for some language, that the grammatical sentences are fixed (i.e., that the process of discovery has been completed) and to determine the effect on grammar-formulation of explicit considerations of simplicity imposed on the grammatical statement.[5]

The outline of Modern Hebrew grammar given below is an example of the second step in linguistic analysis, artificially isolated. It is assumed that the sole purpose of the grammar is to generate a closed body of sentences, these having already been determined. Hence the grammar must be designed in such a way as to be the most efficient, economical, and elegant device generating just these sentences.

The grammar consists of the following parts:

1. A syntactic statement giving permitted arrangements of morphemes in sentences.

2. A morphemic constituency statement giving permitted arrangements of morphophonemes in morphemes.
3. A series of morphological and morphophonemic statements transforming any grammatical sequence of morphemes into a sequence of phonemes.
4. A phonemic statement (transforming phoneme sequences into phone sequences).[6]

The effect of the first two parts is to give the permitted sequences of morphemes by presenting sequences of 'morpheme names', some of them in morphophonemic spelling. The first part will only be sketched here, and the second[7] and the fourth will be entirely omitted. The third part will be given in detail. Beginning with a sequence of morphemes from parts one and two, each statement of the third part of the grammar specifies certain changes which must be undergone by any sequence of a certain shape. It will appear that an order is imposed on the statements, relative to certain criteria of simplicity. Thus the statements are ordered so as to present a maximally simple grammar. The actual demonstration of adequacy given below must be taken in a limited sense only. What is shown is that any single interchange of consecutive statements will necessitate changes which increase the complexity of the grammar.[8] Thus the simplicity of the system is at what might be called a 'relative maximum' with this ordering of statements. It is not excluded that some complicated set of interchanges of the statements might give a simpler grammar, or in fact, that a total recasting in different terms might be more elegant. Thus this investigation is limited in that only one 'dimension' of simplicity is considered, viz., ordering. Actually a complete demonstration would have to show that the total simplicty is greatest with just the given ordering, segmentation, classification, etc.

For the formulation of any relatively precise notion of simplicity, it is necessary that the general structure of the grammar be more or less fixed, as well as the notations by means of which is constructed. We want the notion of simplicity to be broad enough to comprehend all those aspects of simplicity of grammar which enter into consideration when linguistic elements are set up. Thus we want the reduction of the number of elements and statements, any generalization, and, to generalize the notion of generalization itself, any similarity in the form of non-identical statements, to increase the total simplicity of the grammar. As a first approximation to the notion of simplicity, we will here consider shortness of grammar as a measure of simplicity, and will use such notations as will permit similar statements to be coalesced. To keep this notion of simplicity from reducing to an absurdity, the notations must be fixed in advance, and must be chosen to be neutral to any particular grammar, except with respect to the considerations they are chosen to reflect.[9]

Given the fixed notation, the criteria of simplicity governing the ordering of statements are as follows: that the shorter grammar is the simpler, and that among equally short grammars, the simplest is that in which the average length of derivation of sentences is least.

## 1. Notation

The grammar, then, will be a set of transformation statements each of which transforms a given representation of a sentence into a more specific one.[10] If $\alpha$, $\beta$, $\gamma$, with or without subscripts and primes, stand for any sequences (or zero, henceforth Ø) of the elements appearing in statements (e.g., sequences of phonemes, morphemes, phrases, etc., including brackets, dots, etc.), then the basic transformation statements of the grammar will be of the form:

(1) $\alpha \longrightarrow \beta$, where ...

where $\alpha$ and $\beta$ contain no notational elements but are simply sequences of the elements set up to represent parts of sentences (phonemes, morphemes, etc.). This means that $\alpha$ is transformed by this statement into $\beta$, when conditions ... obtain.

If $\alpha=\alpha_1\beta_1\gamma$ and $\beta=\alpha_1\beta_1'\gamma$, we rewrite (1) as:

(2) $\beta_1 \longrightarrow \beta_1'$, in environment $\alpha_1$___$\gamma$, where ...

The notational devices which will actually be used should be introduced definitionally (by so-called 'contextual definitions') be describing a procedure to convert each expression using these notations into a sequence of simple expressions of the form (1) or (2) (which is reducible to (1)) where no notational elements appear. Two kinds of brackets--{},[]--and two kinds of parentheses--(), <>--will be employed as follows:

N1. A statement '... $\left\{\begin{matrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{\underline{n}} \end{matrix}\right\}$...' is an abbreviation for

(i) '... $\alpha_1$ ...', (ii) '... $\alpha_2$ ...', ..., ($\underline{n}$) '... $\alpha_{\underline{n}}$...',

*in that order*. If two sets of brackets with a different number of rows appear, either can be expanded first. If two or more sets of brackets of this form with the same number of rows appear, then they are expanded simultaneously, the $\underline{k}^{th}$ row of the first concurring with the $\underline{k}^{th}$ row of the second. For example

$$'\alpha_1 \left\{\begin{matrix} \alpha_2 \\ \alpha_3 \end{matrix}\right\} \alpha_4 \longrightarrow \beta_1 \left\{\begin{matrix} \beta_2 \\ \beta_3 \end{matrix}\right\}, \text{ where } \ldots \left\{\begin{matrix} \gamma_1 \\ \gamma_2 \end{matrix}\right\} \ldots'$$

stands for

(i) $'\alpha_1\alpha_2\alpha_4 \longrightarrow \beta_1\beta_2$, where ... $\gamma_1$ ...'

(ii) $'\alpha_1\alpha_3\alpha_4 \longrightarrow \beta_1\beta_3$, where ... $\gamma_2$ ...'

*in that order*. To indicate how many rows a given set of brackets have, '---' is written where no element $\alpha$ occurs.

Thus '$\left\{\begin{matrix}\alpha_1\\ ---\\ \alpha_2\end{matrix}\right\}$' has three rows.

N2. N1 holds in exactly the same form for [].

N3. A statement containing one or more elements in main parentheses () is an abbreviation for two statements, one in which *all* of the parenthesized elements appear, and one in which *none* of the parenthesized elements appear, *in that order*. For example

'$\alpha_1(\alpha_2)\alpha_3 \longrightarrow \beta_1\beta_2(\gamma_1(\gamma_2))$, where ...(---)...'

stands for

(i) '$\alpha_1\alpha_2\alpha_3 \longrightarrow \beta_1\beta_2\gamma_1(\gamma_2)$, where ...---...'

(ii) '$\alpha_1\alpha_3 \longrightarrow \beta_1\beta_2$, where ......'

with (i) preceding (ii), and (i) in turn standing for two statements by the same process of development.

N4. A statement containing one or more elements in parentheses <> is an abbreviation for the conjunction, in any order, of all statements with zero or more of the parenthesized elements omitted. For example

'$\alpha\langle\beta\rangle\gamma \longrightarrow \alpha_1\langle\beta_1\rangle$'

stands for

(i) '$\alpha\beta\gamma \longrightarrow \alpha_1\beta_1$'

(ii) '$\alpha\beta\gamma \longrightarrow \alpha_1$'

(iii) '$\alpha\gamma \longrightarrow \alpha_1\beta_1$'

(iv) '$\alpha\gamma \longrightarrow \alpha_1$'

taken in any order. Order does not happen to be important in the statements of the grammar in which <> is used. But it could be, and an order could be imposed. Note that the appearance of a single <> is just like that of a single () (except that order is imposed in the second case).

It remains to give the interpretation for the cases where several of these four notations co-occur in one statement. To do this we have to give an order of priority, stating which of co-occurring notations is to be expanded first, so as to have a unique interpretation for each statement. The order of development follows these two principles:

N5. No brackets or parentheses are expanded if enclosed within brackets or parentheses. I.e., at each step in the development of a sentence only main brackets or parentheses may be developed.

N6. If there is more than one set of main brackets or parentheses, they are developed in the order (i) {}, (ii) [], (iii) (), (iv) <>; i.e., in exactly the order in which they were introduced by N1-4.

This now gives us an explicit step by step procedure for converting each statement of the grammar into an ordered sequence of statements of form (1) or (2). Notice that the case of co-occurring brackets of the same kind with the same number of rows is analogous in interpretation to matrix multiplication, while co-occurring brackets of different kinds give essentially the Cartesian product.

One other point concerning co-occurrence of various notations needs clarification, namely, occurrence of brackets and parentheses within other brackets or parentheses.

N7. Each set of brackets or parentheses is treated as a single element when inside of a containing set. E.g., the main bracket in '$\left\{\begin{matrix}\left\{\begin{matrix}\alpha_1\\ \alpha_2\end{matrix}\right\}\beta \\ \left[\begin{matrix}\gamma_1\\ \gamma_2\\ \gamma_3\end{matrix}\right]\end{matrix}\right\}$' has two rows.

N8. In accordance with customary practice, a set of brackets with a single row is used to give the membership of a class. Thus $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is the class containing as members $\alpha_1, \alpha_2, \ldots, \alpha_n$. A statement of the form '$\alpha=\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$' is interpretable in terms of (1) and (2). It can be taken as an abbreviation for '$\alpha\longrightarrow\alpha_1$ or $\alpha\longrightarrow\alpha_2$ or ... or $\alpha\longrightarrow\alpha_n$'. We write '$\alpha_1\varepsilon\alpha$', '$\alpha_2\varepsilon\alpha$', etc., to indicate that $\alpha_1$ is a member of $\alpha$, $\alpha_2$ is a member of $\alpha$, etc. '$\alpha_i$' will be taken to designate the $i^{th}$ member of $\alpha$ considered to be ordered from left to right as given, i.e., $\alpha_i$, and '$\alpha$' will be used as a variable ranging over members of the class $\alpha$. If $\alpha_i$ itself contains variables or brackets, then any explicit expanded expression produced by developing $\alpha_i$ is taken to be a member of $\alpha$. E.g., if

$$\alpha=\{\alpha_1, \alpha_2\left\{\begin{matrix}\alpha_3\\ \alpha_4(\alpha_5)\end{matrix}\right\}\}$$

then $\alpha_1\varepsilon\alpha$, $\alpha_2\alpha_3\varepsilon\alpha$, $\alpha_2\alpha_4\alpha_5\varepsilon\alpha$, $\alpha_2\alpha_4\varepsilon\alpha$.