Gonzalo, Adriana

A reconstruction of the "classical" linguistic transformational theory CLT

A Reconstruction of the "Classical" Linguistic Transformational Theory CLT*

Adriana Gonzalo†
Wolfgang Balzer‡

Abstract

We reconstruct "the classical transformational theory" of Chomsky, and fit it into the structuralist theory of science. We describe both the formal and the empirical features of this classical account, so that one basic hypothesis of this theory – where central notions are used – can be formulated, and in which Chomsky's "classical" distinction between surface structure and deep structure is clarified. In the empirical claim of this theory are words, sentences and high-structured entities in an inseparable way intertwined. We claim that the formal structure of a natural language is not approximately the same as that of an empirical theory in general. We clarify two special points which affect the structure of the notion of an empirical theory, namely: the delineation of intended applications and the fit between data and models. We hold that the concept of the empirical claim for a linguistic theory should be generalized in comparison with the "standard" structuralist approach.

Keywords: formal reconstruction - scientific theory - classical transformational linguistics - structuralist view

Resumen

Se reconstruye la "teoría clásica transformacional" de Chomsky en el marco de la teoría estructuralista de la ciencia. Se describen tanto los rasgos formales como los empíricos de la versión clásica, de modo de que puedan ser formuladas las hipótesis de la teoría, en las cuales se expresan las nociones centrales, y se clarifica la "clásica" distinción entre estructura superficial y estructura profunda. En la afirmación empírica de esta teoría están interconectadas inseparablemente palabras, oraciones y entidades de alta estructuración. Se sostiene que la estructura formal del lenguaje natural no es aproximadamente la misma que como se da en la teoría empírica en general. Se clarifican dos puntos que afectan la estructura de la noción de teoría empírica, esto es, la delimitación de las aplicaciones intencionales y el modo en que se ajustan los datos y los modelos. Se sostiene que el concepto de afirmación empírica para una teoría lingüística podría ser generalizada en comparación con la visión estructuralista "estándar".

Palabras clave: reconstrucción formal - teoría científica - lingüística transformacional clásica - concepción estructuralista

* Received: 3 January 2012. Accepted in revised version: 20 February 2012.

† Universidad Nacional del Litoral-CONICET. To contact the author, please write to: adriana.n.gonzalo@gmail.com.

‡ Ludwig-Maximilians-Universität, München. To contact the author, please write to: balmar@t-online.de.

Metatheoria 2(2)(2012): 25-49. ISSN 1853-2322.

1. Introduction

From a standard or generalized view, a natural language[1] is primarily both a tool for communication in a group of people, as well as an image which "the" person in a group internalizes, and which depicts her inner person and her surroundings. The individuals in a language group produce and hear sounds that store and transmit information and content. The sounds form the physical medium for storage and transport, including transport "from the outside in" and *vice versa*. In larger groups, the content of some sounds are turned into temporally stable signals, such as pictures, symbols, and written words and sentences. Linguistics is divided into main areas. Phonology is concerned primarily with sounds, while syntax is concerned with words (lexical items), categories, and sentences. Semantics investigates the meaning of words and sentences.

The sentences and words are held together by a system of rules, and are examined and depicted in detail. In this way, it is possible to differentiate expressions which have the form of sentences from other expressions purely on the basis of syntax. Special attention is paid to sentence generation.

On the one hand, in syntax expressions are broken down into primary components so that a sentence can be constructed out of words. On the other hand, this is complemented by an examination of the processes by which expressions are generated.[2]

A system of rules that further structures a language, dividing it into admissible and inadmissible expressions, is referred to as a grammar. In formal description, a grammar is also seen as a tool that uses an established number of rules to generate the set of all sentences in a language. Thus, the term "grammar" consists of at least three components: the set of sentences (and thereby the sets of words as well), the set of rules by which this set of sentences is generated, and a "causal" starting point, without which no generation can begin.[3]

From this linguistic environment, we want – for several reasons – to define and reconstruct a particular linguistic approach using principles from theory of science. In our paper, we are concerned with defining the identity of an empirical theory, delineating the actual intended applications, and defining the relationship between actual systems, data, and linguistic models. Since we have a meta-theoretic tool, the structuralistic theory of science,[4] at our disposal, it is also a goal to in-

[1] In the following, we will leave out the additional descriptor *natural*, as we will not discuss formal, artificial, or hybrid languages here.

[2] In all languages, words are also broken down into sequences of symbols (e.g. morphemes, phonemes, "letters").

[3] Our formulations in the last paragraphs are in several aspects, as one referee pointed out correctly, orthogonal to the spirit of **CLT**.

[4] Sneed (1971), Balzer, Moulines & Sneed (1987), Diederich, Ibarra & Mormann (1989, 1994), Balzer, Moulines & Sneed (2000). It is, of course, necessary that we keep the linguistic structuralism discussed here separate from that based on empirical theories.

tegrate a central approach in linguistics that has had great impact on structural aspects and aspects of history of science.

Our reconstruction primarily uses Chomsky's *Aspects of the Theory of Syntax*,[5] which is described as "classic" in many subsequent works. We therefore call the reconstructed theory the (Chomskian) classical linguistic transformational theory **CLT**. Of course we also studied the forerunners, especially Chomsky's works in the years 1953-1965 (Chomsky 1953, 1955, 1957, 1965), and the *American Structuralism* (Bloomfield 1933, Harris 1951, 1954, Hockett 1954, Wells 1947).

Our work in this paper contains only the central, structural part, which could be further enriched by logical, phonological, semantical, dynamic, historical, and sociological aspects. A more embracing perspective was recently published by Peris-Viñé (2011).[6] We profited also from some earlier articles (Gonzalo 2001, Quesada 1993, Peris-Viñé 1990, 1996, 2010).

Two questions in the philosophy of science, which were clarified satisfactorily for other scientific disciplines – such as physics, psychology, biology, and economics – were, in our opinion, left relatively open in the field of linguistics. These questions were, namely, how one can delineate an actual intended application (a language), and how exactly a language (i.e. an intended application) fits the linguistic models of a theory. Our goal was to elucidate both questions by way of structural means; in order to do so, we had to use the structuralistic instrument in great detail.

2. Some Structuralist Notions

We have modified the structuralistic "standard definition" of an empirical theory (Sneed 1971) in two points. First, we left out the auxiliary base sets, because in our example the hypotheses do not contain auxiliary elements (small numbers are integrated here into the set-theoretical apparatus). Secondly, we have generalized the definition of the specialization of theory-elements[7] in such a way so as to make possible a more realistic demarcation of actual intended applications in linguistics.

An empirical theory *T* consists of the *core K*, the *domain I of intended applications*, and the *approximation apparatus A*: $T = \langle K, A, I \rangle$. The core *K* contains the classes $\mathbf{M_p}$ (of *potential models*), $\mathbf{M}$ (of *actual models*), $\mathbf{C}$ (of *constraints*), and $\mathbf{M_{pp}}$ (of *partial potential*[8] *models*): $K = \langle \mathbf{M}_\text{p}, \mathbf{M}, \mathbf{C}, \mathbf{M}_\text{pp} \rangle$. The potential and actual models are set-theoretical structures of the form $\langle D_1,..., D_k, R_1,..., R_n \rangle$ where $D_1,\ldots, D_k$ are the *base sets* and $R_1,\ldots, R_n$ the *relations* of a (potential) model. We call these sets $D_1,\ldots, R_n$ *the components of* a (potential) model, and we write a (potential) model *x* as follows: $x = \langle v_1,..., v_s \rangle$ where $s = k + n$. A theory has a particular *type* which

[5] Without a reconstruction, we would be unable to speak of an empirical *theory* here.

[6] We cannot compare this approach to our paper here. This would afford a second article.

[7] See in general Balzer, Moulines & Sneed (1993).

[8] The addition of "potential" will be left out for this term in the following.

determines the form of the potential model. The class $\mathbf{M}$ of models is a subclass of $\mathbf{M_p}$ ($\mathbf{M} \subseteq \mathbf{M_p}$) representing the empirical hypotheses that are characteristic and valid for the models. The partial models are generated by the potential models by leaving out certain components $R_{nt+1}, \ldots, R_n$, namely the *theoretical terms*, i.e. $\mathbf{M}_{\text{pp}} = \{\langle D_1, \ldots, D_k, R_1, \ldots, R_{nt}\rangle / \exists R_{nt+1} \ldots \exists R_n (\langle D_1, \ldots, D_k, R_1, \ldots, R_{nt}, R_{nt+1}, \ldots, R_n\rangle \in \mathbf{M_p})\}$.

The *restriction function* $\mathbf{r}, \mathbf{r} : \mathbf{M_p} \rightarrow \mathbf{M_{pp}}$, removes the theoretical terms from the (potential) models. This function can be raised to the power set $\mathbf{r} : Pot(\mathbf{M_p}) \rightarrow Pot(\mathbf{M_{pp}})$, $\mathbf{r}(X) = \{\mathbf{r}(x) / x \in X\}$. We can then define the *ideal content*, CONT(*K*), of the core *K* of *T* for an empirical theory *T* as follows:

$$(1)\ Y \in \text{CONT}(K) \leftrightarrow Y \in Pot(\mathbf{M_{pp}}) \wedge \exists X (X \subseteq \mathbf{M} \wedge X \in \mathbf{C} \wedge Y = \mathbf{r}(X))$$

A theory *T* has no ideal content[9] iff the content of *T*, CONT(*K*), comprises the entire power set *Pot*($\mathbf{M}_{\text{pp}}$). In other words, every set of potential models can be embedded into a set of models X "with" constraints.

We make the idealized assumption that every intended application of *I*, from the perspective of theory *T*, consists of the *T*-non-theoretical components, and therefore "is" a partial model of T: $I \subseteq \mathbf{M_{pp}}$. Although an intended application of *T* is *a priori* an actual system, it is perceived by a group of researchers "through the lens" of their theory and its terms.

The approximation apparatus *A* is used to "approximately embed" the set *I* of intended applications into the ideal content of the core, in which the identity $Y = \mathbf{r}(X)$ formulated in (1) applies only approximately: $Y \approx \mathbf{r}(X)$. This approximate relation $\approx$ can be defined on three levels. First, a similarity relation $\sim$ (or more precisely "similarity of degree ε', $\sim_\varepsilon$) of components v, v' is defined, where the components v, v' must have the same form. Such a relation $v \sim v'$ is observed or measured in special points. A *n*-ary relation v can often be further treated approximately. If, for example, v and v' have two arguments, $v = R(-,-)$, $v' = R'(-,-)$, atomic formulas $R(x, y)$ and $R'(x', y')$ can often be reduced to the arguments x, x', y, y' and the relations of similarity $\sim_1, \sim_2 : x \sim_1 x', y \sim_2 y'$. Secondly, different types of similarity relations for partial models can be defined by joining individual components. Two partial models $y, y' \in \mathbf{M_{pp}}$ are similar $\simeq$ (or "similar to degree ε", $\simeq_\varepsilon$) iff all components y_i of y are similar to the corresponding components y'_i of y', that is $y_i \sim y'_i$ (or $y_i \sim_\varepsilon y'_i$). Thirdly, it is possible to define that two sets Y, Z of partial models are similar (of degree ε) iff $z \in Z, \exists y \in Y (z \simeq y)$, in other words: $Z \approx Y$.

If for Z we take specifically the set *I* of intended applications, there is for *I* a set X of "selected" models that can be approximately embedded. Thus, X becomes similar to a set $\mathbf{r}(X)$ of restricted partial models: $I \subseteq \mathbf{r}(X)$ We then arrive at the approximate content $\text{CON}^{\approx}_T(K)$ of the theory *T* and the corresponding approximate empirical claim:

---

[9] See for example Balzer, Moulines & Sneed (1987, p. 82).

(2) $I \in \mathrm{CON}^{\approx}_{T}(K)$, more precisely $\exists X(X \subseteq \mathbf{M} \wedge X \in \mathbf{C} \wedge I \approx \mathbf{r}(X))$.

In other words, all intended applications are incorporated into actual models so that all the models selected in this manner fulfill the constraints.

In linguistics, similarity relations are generally approached using statistical methods. In this period many statistical methods were used: elementary methods, like goodness of fit or χ-distributions, or advanced methods.[10] Using approximation, several statistical exceptions are admissible in the empirical claim; some intended applications lie merely in the vicinity of restricted models.

In Balzer, Moulines & Sneed (1987), a differentiation is made between two types of theories: *theory-elements* and *theory-nets*. A theory-net consists of a *basic-element* T_0 and a net of *specializations*. T_σ is a *specialization* of T_0 iff 1) T_σ is an empirical theory, 2) the class of models $\mathbf{M}_\sigma$ of T_σ is a subclass of $\mathbf{M}_0$, and 3) the set I_σ of intended applications for T_σ is a subset of I_0. In other words, the models of T_σ fulfill the hypotheses of T_0 as well as further additional hypotheses valid only for the particular intended applications of T_σ.

The subset relations "$\mathbf{M}_\sigma \subseteq \mathbf{M}_0$" and "$I_\sigma \subseteq I_0$" are generalized in Balzer, Moulines & Sneed (1993). To this end, potential models are restricted in a more general way. From a potential model one can replace a relation R_j by a sub-relation $R'_j (R'_j \subseteq R_j)$ and a base set D_i by a subset $D'_i (D'_i \subseteq D_i)$ of the base set. Therefore we can restrict a potential model in more ways, we can restrict it in other "dimensions of freedom". For a potential model x of the form $x = \langle v_1, \ldots, v_s \rangle$ of a given type τ, x' is a *partial structure* of x $(x' \sqsubseteq x)$ iff 1) $x' = \langle v'_1, \ldots, v'_s \rangle$; 2) for every $i \leq s$, $v'_i \subseteq v_i$; and 3) x' and x have the same type τ.

We define a generalized specialization T' of T_0 over a basic element $T_0 = \langle\langle \mathbf{M_p}, \mathbf{M}, \mathbf{C}, \mathbf{M_{pp}} \rangle, I \rangle$ as follows: T' has the form $\langle\langle \mathbf{M_p}', \mathbf{M}', \mathbf{C}', \mathbf{M_{pp}}' \rangle, I' \rangle$, and it holds that:

1) $\mathbf{M_p}' = \mathbf{M_p}$
2) $\forall x' \exists x (x' \in \mathbf{M}' \wedge x \in \mathbf{M} \wedge x' \sqsubseteq x)$
3) $\mathbf{C} = \mathbf{C}'$
4) $\forall y' \exists y (y' \in I' \wedge y \in I \wedge y' \sqsubseteq^* y)$

where $\sqsubseteq^*$ is the transitive closure of $\sqsubseteq$. As with theory-nets, we can also implement generalized theory-nets.

3. Trees, Rules, and Markers

Three essential terms of the theory **CLT** were used merely informally in the period discussed here.[11] Though these terms were continually applied, the formal defi-

[10] See, for instance, in this period, Bar-Hillel, Kasher & Shamir (1963), today e.g. Ho (2006). One of the linguistic methods is described in Clark (1992).

[11] Several formal definitions can be found in Chomsky's dissertation which was, however, first published in Chomsky (1975).

nitions were not important since the theoretical "superstructure" was constantly changing.

Informally, the notion of a tree plays a central role here. In a set-theoretic, simplified way, it consists a tree of a set of "nodes", a set of "start elements" and a set of "edges". Depicted in a normal way begins a tree from a start element and branches off "downward". Two types of trees are used in **CLT**. In the first type, the "lowest" nodes of a tree contain mostly words which, when in the right order, yield a sentence. Application of trees thus described presents two problems that are not easily comprehended "at a glance". First, with a sentence and its corresponding tree, it is often not possible to tell how the order of the words in the sentence is generated. Secondly, it is difficult to illustrate complexly structured sentences with a single tree. For this reason, "second level" trees are used in **CLT**, which Chomsky called transformation markers. In the follow we will use the term "frame" instead of "tree" to avoid misunderstandings and unsatisfiable expectations. We use the term "tree" only in a very special notion of "ordered tree" which is formally not a tree in the normal sense.

To keep this section short, we are using the concept of the *n*-tuple from set theory. An *n*-tuple $\langle x_1,\ldots, x_n\rangle$ is a sequence of symbols or components $x_i (i=1,\ldots, n)$ so that each x_i represents a set or a variable for sets. If all these components x_i are elements of a set X, one says that $\langle x_1,\ldots, x_n\rangle$ is an *n*-tuple for the set X. The set of all *n*-tuples for X is then defined by $X^n = \{\langle x_1,\ldots, x_n\rangle / \forall i \leq n\ (x_i \in X)\}$, and the set of all tuples for X by $X* = \cup_n X^n$.
By way of preparation, we will establish a general frame:

D1 $\langle N, \Psi, D\rangle$ is a frame iff the following is true:

1) N is a finite, non-empty set (of "frame-elements")
2) $\Psi \subseteq N$ (a set of "start elements")
3) $D \subseteq N \times N*$ (a set of "derivation rules")
4) for all rules $r \in D$ and all $x, y_1,\ldots, y_n \in N$, if $r = \langle x, \langle y_1,\ldots, y_n\rangle\rangle$,[12] then there exists $y_i \in \{y_1,\ldots, y_n\}$ such that: $y_i \neq x$.

A frame is usually drawn from the top downward, so that we find a start element $\sigma \in \Psi$ at the top. A rule r is a pair $\langle t_f, t_r\rangle$ of terms (Bourbaki 2004, Chap. IV) t_f, t_r where t_r is an *n*-tuple: $t_r = \langle y_1,..., y_n\rangle$. We call t_f the *fire-part* of the rule and t_r the *result-part* of the rule. The rule r finds the term t_f and fires, producing the resultant term t_r (the result-part). Put another way, a rule r generates the result part t_r using the fire-part t_f. By D1-3, a rule r always takes the form $\langle x, \langle y_1,\ldots, y_n\rangle\rangle$: $r = \langle x, \langle y_1,\ldots, y_n\rangle\rangle$. In the simplest case of n=1 a rule r has the form $\langle e, \langle e'\rangle\rangle$ or abbreviated: $\langle e, e'\rangle$, $r = \langle e, e'\rangle$. Condition D1-4 limits the rules to "generative" rules, i.e. something new is generated from the fire-part. These rules are used to enlarge a frame by – in general – "attaching" an additional frame-element below one of the "lower" frame-elements.

[12] In these rules, we often leave out the brackets around $\langle y_1,..., y_n\rangle$.

We define the concept of an ordered tree inductively. In this way the "lower" terminal frame-elements appear in the "correct" order. For this purpose, we use an order relation which remains implicit in D2, and which was likewise informally discussed even at that time.

D2 a) Inductive definition of ordered trees in the frame $\langle N, \Psi, D\rangle$.

i). for all $\sigma \in \Psi$, $\langle\{\sigma\}, \sigma, \langle\sigma\rangle, \varnothing\rangle$ is an ordered tree in the frame $\langle N, \Psi, D\rangle$.

ii). if $\langle K, \sigma, \langle b_1,\ldots, b_n\rangle, R\rangle$ is an ordered tree in the frame $\langle N, \Psi, D\rangle$ and if there exists $i \leq n$ and $r \in D$, such that there exists $m \geq 1$ and $e,\ldots, e_m \in N$, such that $r = \langle b_i, e_1,\ldots, e_m\rangle$, then $\langle K', \sigma', \langle b'_1,\ldots, b'_{n+m-1}\rangle, R\rangle$ is an ordered tree in the frame $\langle N, \Psi, D\rangle$ where the following is true: i) $K' = K \cup \{e_1,\ldots, e_m\}$, ii) $\sigma' = \sigma$, iii) $R' = R \cup \{r\}$, iv) $\langle b_1,\ldots, b_{i-1}, e_1,\ldots, e_m, b_{i+1},\ldots, b_n\rangle = \langle b'_1,\ldots, b'_{n+m-1}\rangle$.

b) The *set of ordered trees* is denoted by $OT(N, \Psi, D)$.

In an ordered tree $\langle K, \sigma, \langle b_1,\ldots, b_n\rangle, R\rangle, \langle b_1,\ldots, b_n\rangle$ is called *the basis* of the ordered tree, *K* is the set of frame-elements, σ is the start element, and *R* is a set of rules. By a rule $r = \langle b_i, e_1,\ldots, e_m\rangle$ the component b_i of a basis $\langle b_1,\ldots, b_n\rangle$ is replaced by $\langle e_1,\ldots, e_m\rangle$ so that the new basis has the form $\langle b_1,\ldots, b_{i-1}, e_1,\ldots, e_m, b_{i+1},\ldots, b_n\rangle$.

Lemma 1: If $\langle K, \sigma, \langle b_1,\ldots, b_n\rangle, R\rangle$ is an ordered tree in the frame (N, Ψ, D), with $\sigma \in \Psi$, then $K \in N$.

Proof: Following D2-i and D1-2), all lie within *N*. If in D2-ii) all $e_1,\ldots, e_m$ are elements of *N*, then, by D2-ii-i), it follows that $K' \subseteq N$.

An ordered tree is generated inductively step by step. In every step of induction the ordered tree is enlarged. It always begins with a start element, from which the "next line" of the order tree is created using a rule. Contained in the "currently bottom-most" line is the basis of the ordered tree, which was generated through prior application of the rule. This ordered tree can be used to construct a larger ordered tree as follows: one takes a rule $r \in D$ whose fire-part of *r* is identical to one of the frame-elements *e* of the basis, and whose result part $\langle e_1,\ldots, e_n\rangle$ is written below *e*. An edge is then drawn between *e* and every frame-element e_i. All these edges are systematically arranged by D2.

An induction process can be concluded in any step. Visually, the basis of the ordered tree may appear rather scattered, often making the "respective bottom-most" line of the ordered tree difficult to see. Both a frame-element from the basis of an ordered tree, as well as the rule by which one arrives at this frame element, can also appear "further up" in this ordered tree. For example, we start from $K = \langle\{a\}, a, \langle a\rangle, \varnothing\rangle$ and use the rule $r = \langle a, \langle a, b\rangle\rangle$ twice. From *K* and *r* we get $K_1 = \langle\{a, b\}, a, \langle a, b\rangle, \{\langle a, \langle a, b\rangle\rangle\}\rangle$ (D2-ii), and from K_1 and *r* we get $K_2 = \langle\{a, b\}, a, \langle a, b, b\rangle, \{\langle a, \langle a, b\rangle\rangle\}\rangle$. In K_1 the first component a of the basis $B_1 = \langle a, b\rangle$ is again replaced by $\langle a, b\rangle$, and B_1 changes to $\langle a, b, b\rangle$. In (D-ii-iv) we already used the abbreviation by which in a rule $\langle t_r, t_f\rangle$, with $t_f = \langle y_1,\ldots, y_w\rangle$,

the last pair of brackets is omitted. So $\langle t_r, \langle y_1, \ldots, y_w \rangle \rangle$ becomes $\langle t_r, y_1, \ldots, y_w \rangle$.

We distinguish frames of the first and the second type, and we denote theses frames by $\langle E, \Sigma, R^{pm} \rangle$ and by $\langle \Theta, \Xi, R^{tm} \rangle$. In a frame of the first kind we call the frame-elements *Chomsky-elements*, and we denote the *sets* of Chomsky-elements by E or E^{chy} (see below). In a frame of the second type are the frame-elements themselves ordered trees. These frames we write in the form $\langle \Theta, \Xi, R^{tm} \rangle$ where Θ is a set of ordered trees, as defined in D2, i.e. $\Theta \subseteq OT(N, \Psi, D)$. By this notation we can begin with the set E of Chomsky-elements, and an appertaining frame of the first kind $\langle E, \Sigma, R^{pm} \rangle$, and define ordered trees in the frame by $\langle E, \Sigma, R^{pm} \rangle$. In a second step we can form in D2 a restricted set Θ of ordered trees in the frame $\langle E, \Sigma, R^{pm} \rangle$: $OT\langle E, \Sigma, R^{pm} \rangle$. This set Θ of ordered trees is used now as a set of "complex, second-order" frame-elements in ordered trees of second-order.

In this way we can construct three kinds of markers which Chomsky used, namely phrase-markers, derived phrase-markers and transformation-markers.

D3 Let a frame of the form $\langle E, \Sigma, R^{pm} \rangle$ be given.

a) *pm* is a *phrase-marker* in the frame $\langle E, \Sigma, R^{pm} \rangle$ (abbreviated by: $pm \subseteq PM(E, \Sigma, R^{pm})$) iff *pm* takes the form $\langle K, \sigma, \langle b_1, \ldots, b_m \rangle, R \rangle$ and the following requirements are true:
   1) *pm* is an ordered tree in the frame $\langle E, \Sigma, R^{pm} \rangle$,
   2) $R \subseteq R^{pm}$, and
   3) $\{b_1, \ldots, b_m\} \subseteq E$.

b) The set $DPM(E, \Sigma, R^{pm})$ of *derived phrase-markers* is defined inductively.
   i) if x is a phrase-marker, then x is a derived phrase-marker.
   ii) if $\langle K, \sigma, \langle b_1, \ldots, b_m \rangle, R \rangle$ is a derived phrase-marker and $\langle K', \sigma', \langle b'_1, \ldots, b'_n \rangle, R' \rangle$ is a phrase-marker, and if there exists $j \leq m$ such that $b_j = \sigma'$, then $\langle K \cup K', \sigma, \langle b_1, \ldots, b_{j-1}, b'_1, \ldots, b'_n, b_{j+1}, \ldots, b_m \rangle, R \cup R' \rangle$ is a derived phrase-marker.

A derived phrase-marker is created when one phrase-marker is embedded into another. That is, a frame-element from the basis of the first marker is replaced with the entire second marker, and the basis of the first marker is extended by incorporating the basis of the second marker in the "correct" location.

Lemma 2: $PM(E, \Sigma, R^{pm}) \subseteq DPM(E, \Sigma, R^{pm})$.
Proof: D3-b-i.

We define transformation-markers as ordered trees of second level whereby a special frame of the form $\langle \Theta, \Xi, R^{tm} \rangle$ is given. An element of Θ (a frame-element) is, as said above, a derived phrase-marker, and a start element from Ξ is a derived phrase-marker. For simplicity, we identify the set of frame-elements of Θ with the full set $DPM(E, \Sigma, R^{pm})$ of all derived phrase-markers.

(3) $\Theta = DPM(E, \Sigma, R^{pm})$ and $\Xi \subseteq \Theta$.

a)The set $TM(\Theta, \Sigma, R^{tm})$ of *transformation-markers* is defined inductively.

i) if $\boldsymbol{k}$ is a derived phrase-marker and $\boldsymbol{k}$ is an element of Ξ, then $\langle\{\boldsymbol{k}\}, \boldsymbol{k}, \langle\boldsymbol{k}\rangle, \varnothing\rangle$ is a transformation-marker.

ii) if $\langle\{\boldsymbol{k}_1,\ldots,\boldsymbol{k}_n\}, \sigma, \langle\boldsymbol{b}_1,\ldots,\boldsymbol{b}_m\rangle, \boldsymbol{R}\rangle$ is a transformation-marker, and if there exist $j \leq m$ and $\boldsymbol{y} \in \Theta$ such that $r = \langle\boldsymbol{b}_j, \boldsymbol{y}\rangle \in R^{tm}$ then $\langle\{\boldsymbol{k}_1,\ldots,\boldsymbol{k}_n,\boldsymbol{k}\}, \sigma, \langle\boldsymbol{b}_1,\ldots,\boldsymbol{b}_{j-1},\boldsymbol{k},\boldsymbol{b}_{j+1},\ldots,\boldsymbol{b}_m\rangle, \boldsymbol{R} \cup \{\langle\boldsymbol{b}_j,\boldsymbol{k}\rangle\}\rangle$ is a transformation-marker.

ii) if $\langle\{\boldsymbol{k}_1,\ldots,\boldsymbol{k}_n\}, \sigma, \langle\boldsymbol{b}_1,\ldots,\boldsymbol{b}_m\rangle, \boldsymbol{R}\rangle$ is a transformation-marker, and if there exist $j \leq m$ and $\boldsymbol{k}, \boldsymbol{k'} \in \Theta$ such that $\langle\boldsymbol{b}_j, \boldsymbol{k}, \boldsymbol{k'}\rangle \in R^{tm}$, then $\langle\{\boldsymbol{k}_1,\ldots,\boldsymbol{k}_n,\boldsymbol{k},\boldsymbol{k'}\}, \sigma, \langle\boldsymbol{b}_1,\ldots,\boldsymbol{b}_{j-1},\boldsymbol{k},\boldsymbol{k'},\boldsymbol{b}_{j+1},\ldots,\boldsymbol{b}_m\rangle, \boldsymbol{R} \cup \{\langle\boldsymbol{b}_j,\boldsymbol{k},\boldsymbol{k'}\rangle\}\rangle$ is a transformation-marker.

Finally, we establish the sequences of Chomsky-elements, which we, like Chomsky, call *strings*.

D4 x is a *string structure* $(x \in SS)$ iff there exist $Str, \circ, E, \Lambda$, such that the following is true:

1) $x = \langle Str, \circ, E, \Lambda\rangle$
2) Str is a set (of "strings")
3) $\circ : Str \times Str \rightarrow Str$ is an associative[13] function ("concatenation")
4) E is a non-empty, finite set, $\varnothing \neq E \subset Str$ and $\Lambda \in E$
5) $\forall s \in Str\ \exists\, e_1,\ldots,e_n \in E\,(s = \circ(e_1, \circ(e_2,\ldots,\circ(e_{n-1}, e_n)\ldots)))$
6) $\forall s \in Str(\circ(s, \Lambda) = s = \circ(\Lambda, s))$
7) $\forall s, s' \in Str(s \neq \Lambda \neq s' \rightarrow s \neq \circ(s, s') \neq s')$

Each pair of strings s_1, s_2 are combined into one new string s by way of $\circ : \circ(s_1, s_2) = s$. Since the function $\circ$ can be applied iteratively, 5) requires that it be possible to combine all strings of *Str* of Chomsky-elements. In the following, we will write the strings like this: $s_1 \circ \ldots \circ s_n$. The Chomsky-elements are handled as special cases of strings by 4); that is, every Chomsky-element of *E* is also a string. The Chomsky-element Λ, the *blank*, is purely an aid to be able to distinguish two strings s_1, s_2 from the concatenated string $(s_1 \circ s_2)$. It is often difficult to differentiate between strings and *n*-tuples in a practice based on structure.

D5 If $pm = \langle K, \sigma, \langle b_1,\ldots,b_m\rangle, R\rangle$ is a phrase-marker or a derived phrase marker in the frame $\langle E, \Sigma, R^{pm}\rangle$, $z = \langle Str, \circ, E, \Lambda\rangle$ a string structure, and $tm = \langle \boldsymbol{K}, \sigma^*, \langle\boldsymbol{b}_1,\ldots,\boldsymbol{b}_m\rangle, \boldsymbol{R}\rangle$ a transformation-marker in the frame $\langle\Theta, \Xi, R^{tm}\rangle$, we define:

a) end(*pm*) is the *end string* of *pm* iff $\boldsymbol{end}\,(pm) = b_1 \circ \ldots \circ b_m$.

b) $\boldsymbol{end}$(*tm*) is the *second-level end string* of *tm* iff there exists $K^m, \sigma^m, b^m{}_1,\ldots, b^m{}_{u_m}, R^m$ such that $\boldsymbol{b}^m = \langle K^m, \sigma^m, \langle b^m{}_1,\ldots,b^m{}_{u_m}\rangle, R^m\rangle$ is a derived phrase marker, and $\boldsymbol{end}\,(tm) = b^m{}_1 \circ \ldots \circ b^m{}_{u_m}$.

---

[13] I.e. $\forall s_1 \forall s_2 \forall s_3 \in Str\,(s_1 \circ s_2) \circ s_3 = s_1 \circ (s_2 \circ s_3)$.

4. Chomsky's Bases and Basis

In Chomsky (1965), a distinction is made between six types of Chomsky-elements (*W*, *LC*, *PC*, *F*, *CS*, *H*) of which strings are made up.[14] *W* is the set of *words*[15] in a language, and *H* a set of *auxiliary symbols* used in the theory **CLT**. *LC* is a set of *lexical categories*, and *PC* a set of *phrase categories*. In the English language – and in many other languages – we find, for example, lexical categories such as *noun*, *verb*, *adverb* etc., and phrase categories such as *nominal phrase*, *verbal phrase*, *sentence*, etc. For example, in English *The girl*, *Tom*, and *Many dead trees* are nominal phrases, while *play* and *reads a book* are verbal phrases. *F* is a set of *features*[16] which express syntactic roles of lexical items (see below). Here we use only three auxiliary symbols +, -, ◊ from the set *H*, which are needed for the construction of complex symbols. Other auxiliary symbols – e.g. special brackets # ... # which delimit partial derivations in Chomsky's rules – are not used here. *CS* is a set of complex symbols that are defined by *F* and *H*.

From the words, phrases and sentences can be created. A sentence is a string of words where the string fulfills other characteristics as well, as described in the following models. In a first approximation, a phrase is a "part" of a sentence, a sequence of concatenated words that, taken together, express a meaning.

Sub-categories can be created from the features with the help of rules. For the sake of simplicity, we subsume here the lexical categories under the features as well. In this way (see (D7-6) and Lemma 4, Sec. 5) the lexical categories are also subsumed under the sub-categories. In English there is, for example, the lexical category *noun*, within which a differentiation is made between subcategories such as *animate* or *common*. For *common*, there are sub-categories *countable/uncountable* or *abstract/concrete*. These divisions sometimes lead to a classification. There are, however, other sub-categories that can only be accommodated in a multi-dimensional lattice. The problem of sub-categories was discussed at length in Chomsky (1965). A set of the sub-categories contains elements of very different kinds. This set could not be precisely delineated. Chomsky had therefore chosen another way, harking back to Halle (1962) who used matrices. The term *sub-category* is replaced with a technical term *complex symbol*, which is used to generate ordered trees.

In a generation process a sub-category is chosen from a lexical category (or from a sub-category) by a rule which uses a feature. A feature is formally a Chomsky-element. Informally, one can view it as a prescription to come from one lexical item to a more special item. For example, in the English language the lexical category *noun* has features like *animate*, *countable*, *human*. A special word like *child* is a *noun*,

[14] In Chomsky (1965), a seventh type, the *grammatical formatives*, is also used, but it has no formal consequences there.s_1.

[15] Words from *W* are treated as such givens in Chomsky (1965) that they are not considered especially worth mentioning.

[16] See Chomsky (1965, Sec. 2.3).

and semantically speaking child has some properties, which can also expressed by the terms *animate*, *human*, and in a certain sense also by *countable*. In other words, we can come from a *noun* to a more special class of expressions. In the same way we can proceed from *verb* to a more special kind of verb, like for example *read*, which has features such as *predicative*, or *transitive*.

Formally, we define the set CS(*F*) of complex symbols of *F* somewhat more generally and inductively, whereby we add the prefixes +, - , or ◊ to some of the complex symbols.[17] A complex symbol of the form +*cs* means that in the formation of a ordered tree, the symbol *cs* must be used at this stage; -*cs* means that the symbol *cs* cannot be used at this stage; and ◊ *cs* means that *cs* can be used as one alternative. D6 The set CS(*F*) of *complex symbols* of *F* is defined inductively.

1) every $f \in F$ is a complex symbol.
2) if $cs_1, \ldots, cs_n \in CS(F)$ and $\xi_1, \ldots, \xi_n \in \{+, -, \Diamond\}$, then $[\xi_1 cs_1, \ldots, \xi_n cs_n] \in CS(F)$

A sub-category is determined as follows: in an ordered tree that is still in the process of being formed, a Chomsky-element *e* from the basis of the ordered tree is replaced with a complex symbol *cs*. Specifically, if *cs* is a feature *f*, then following a lexical insertion rule (see below), *e* is replaced with a word that has the feature *f*. For example, if *e* is a *noun* one can replace *noun* by *man*, but also by *water*. To restrict several alternatives, we can first replace *noun* by a special feature, like *human*, and use the prefix +, to replace *noun* by *+human*. In a next step we find a rule in which *man* can be used, but there is no rule to replace *human* by *water*.

We summarize all these elements in the set E^{chy} of Chomsky-elements:

(4) $E^{chy} = W \cup LC \cup PC \cup F \cup CS \cup H$.

When generating ordered trees and markers, a distinction is made in **CLT** between two main types of rules, R^{pm} and R^{tm}, as previously introduced: rules used to create phrase-markers, and rules used to create transformation-markers.

Further distinctions can be made within the set of rules for phrase-markers. In Chomsky (1965) and in additional works, three[18] sub-types are used: the *normal phrase rules* (R^{pn}), the *sub-category rules* (R^{sc}), and the *lexical insertion rules* (R^{lx}), which are normally worked through in this order when generating a sentence. Furthermore, the set R^{pm} can be defined by these three sub-types, i.e. $R^{pm} = R^{pn} \cup R^{sc} \cup R^{lx}$. We call all these rules Chomsky-rules.

(5) $R^{chy} = R^{pm} \cup R^{tm} = R^{pn} \cup R^{sc} \cup R^{lx} \cup R^{tm}$.

[17] A standard form, which is used today in most computer language manuals, was developed from the original, specialized matrix approach of Halle (1962).

[18] An additional type of rule that leads to the phonetic level, and that is essential in Chomsky (1957) and other texts, is mentioned merely as an aside in Chomsky (1965). In principle, these rules could be embedded without difficulty into the models formulated here. This is presumably also the reason that the phonetic level is not further discussed in Chomsky (1965).

Chomsky calls the set of all rules that create phrase-markers the *base* for the phrase-markers.[19]

For phrase rules, there is always the phrase category S (the sentence category) that appears as the start element when generating sentences. Chomsky mentions the other modes, such as interrogative or imperative, only briefly.
The rules r^{pm} for phrase-markers have the general form:

(6) $r^{pm} = \langle e, e_1, ..., e_n \rangle$ where $e, e_1, ..., e_n \in E$ and $r^{pm} \in R^{pm} \subseteq E \times E*$.

By a normal phrase rule $r^{pn} = \langle e, e_1, ..., e_n \rangle \in R^{pn}$, a phrase category $e \in PC$ is replaced with a tuple $\langle e_1, \ldots, e_n \rangle$ whose components are phrase categories or lexical categories, i.e. $R^{pn} \subseteq PC \times (PC \cup LC)*$.

As previously discussed, the sub-category rules replace the lexical categories with complex symbols. The reason for this replacement is to specialize the lexical categories in a natural way. In general, the sub-category rules take the form shown in (6). More specifically, a sub-category rule can have, first of all, the form $\langle lc, [lc, \xi_1 f_1, ..., \xi_n f_n] \rangle$ where $lc \in LC$, $f_1, ..., f_n F$ and $\xi_1, \ldots, \xi_n \in \{+, -, \Diamond\}$.

This means that the fire-part consists of a lexical category, and the result-part consists of a complex symbol $[lc, \xi_1 f_1, \ldots, \xi_n f_n]$ whose first component is in fact the lexical category *lc*. Using steps of induction a rule can have the general form: $\langle cs, [\xi_1 cs_1, \ldots, \xi_n cs_n] \rangle$, where *cs*, cs_1,..., cs_n are complex symbols. These rules can be nested when applied. For example, a *noun* can be specialized to become [*noun*, *+common*], then further to become [*noun*, *+common*, *+count*,...]. In general, *cs*, cs_1,..., cs_n are features which occur in rules $r^{sc} \in R^{sc}$. We typify the set R^{sc} as follows: $R^{sc} \subseteq LC \times CS(F)$.

The further type of rules, the lexical insertion rules, sets itself apart from the first three types in two points. The rules of the three types previously discussed are normally read in the direction "from top to bottom", that is, from the phrase categories to the words. These rules generally extend the basis of a marker. By contrast, a lexical insertion rule is read in a way of a normal lexicon: *from* a word *to* various (sub-) categories or other linguistic elements. With regard to content, *noun*, for example, can be instantiated through many different words such as *man*, *tree*, *thing*, *category*, etc. Since we have arranged the phrase rules – and all other rules – formally from left to right, a lexical insertion rule is read from right to left, but is still "processed" from left to right.

The second point of interest is that in Chomsky (1965, p. 84), one and only one schematic lexical rule is formulated by which complex symbols can be replaced by feature matrices. This rule can be said to be universal, a point constantly emphasized by Chomsky. We use here a simpler formulation which is not universal. We express straight away the different feature matrices by different "local" insertion

[19] On p. 17 of Chomsky (1965), these essential terms: *base* and *basis* are introduced and then immediately qualified, e.g. on p. 18.

rules. The price for this is that we must give up universality at this point. The first three types of rules are applied in all – or at least in many – languages, but whether all insertion rules are universal is still a matter of discussion.

A lexical insertion rule r begins with a complex symbol, which is replaced during the generation of an ordered tree with a word or several words. Formally, a lexical insertion rule takes the form $\langle x, w_1, \ldots, w_n \rangle$ where $w_1, \ldots, w_n \in W$ and $x \in CS(F)$ The set R^{lx} of lexical insertion rules is typified as follows: $R^{lx} \subseteq CS(F) \times W^*$. 0

The components W, $CS(F)$ and R^{lx} form an initial basis for a lexicon[20] for a particular language: $L = \langle W, CS(F), R^{lx} \rangle$.

The rules for transformation-markers are more complex. The entities that are transformed in these rules are phrase-markers. Structurally, these rules r^{tm} for transformation-markers take the form:

(7) $r^{tm} = \langle x, y_1, \ldots, y_n \rangle$ where $x, y_1, \ldots, y_n$ are derived phrase-markers.

A complex transformation in a frame[21] is made up of simple "elementary" transformations.

The way in which Chomsky illustrated transformation rules in the period discussed here did not hold for long. We describe these rules merely informally.[22] In a transformation-marker, partial frames can be defined whose frame-elements once again have the structure of ordered trees. In this case, the start element of the partial frame is a frame-element that lies "further down" in the complete frame, and the basis of the partial frame is a sequence of frame-elements that lie "further up" on the complete frame. Since the frame-elements of the partial frame are derived phrase-markers, the transformation is divided into several elementary transformations, such that every elementary transformation is performed using one of the transformation rules of the form (7). The question of whether or not the order of the elementary transformations influenced the result, and whether or not the corresponding "elementary" transformations are independent of one another, was discussed during this period, though without lasting results.[23] We typify the transformation rules in the following way: $R^{tm} \subseteq DPM \times DPM^*$.

During the period discussed here, there was an array of other approaches for how transformations could or should be described. Empirical and formal perspectives played a role here. For example, in Chomsky (1957), such rules were illustrated using merely a few examples in a fragment of the English language, or in Chomsky (1965) informally using trees. Chomsky (1961) contains a half formal approach; Chomsky (1953) further formally develops the transformation approach of Harris (1954). Mathematical models came later, such as Ginsburg and Partee

[20] In English, for example, Hornby (1974), *Oxford Dictionary*.

[21] See e.g. Chomsky (1961, p. 131).

[22] Chomsky also proceeded in this manner at this time, see for example Chomsky (1961).

[23] See Chomsky (1965, p. 98)

(1969). In the mathematical models, the transformations are mostly represented using bijective functions from the Chomsky-elements or from parts of strings. In short, a string of the form $e_1 \circ \ldots \circ e_n$ is transformed into a string $e_{\varphi(1)} \circ \ldots \circ e_{\varphi(n)}$, where φ is a bijection between the order indexes.

The transformation rules do not necessarily have to appear in every derivation. For example, in the sentence *I go*, no transformation rule is used. The same holds true for the sub-category rules.

Using these distinctions, we can define the *closed* derived phrase-markers and the *preterminal* strings from a derived phrase-marker. We say that a derived phrase-marker is closed iff all Chomsky-elements from the basis of the marker are words. In this case, these words from the basis can potentially be concatenated into a sentence. We denote the set of closed derived phrase-markers with $CDPM(E, \Sigma, R^{pm}, W)$. From a graphic standpoint, a closed (derived) phrase-marker can illustrate the structure of a complete sentence. In such cases, the marker describes the way in which the sentence was constructed. In the general case, the basis of a phrase-marker also contains "variables", namely other Chomsky-elements inscribed in the frame-elements of the basis. We call all other derived phrase-markers *open* phrase-markers.

Lemma 3: $CDPM(E, \Sigma, R^{pm}, W) \subseteq DPM(E, \Sigma, R^{pm})$
Proof: D3-a-3), (5), D3-b-i, ii).

If the basis of a derived phrase-marker *pm* consists only of complex symbols, we call this marker's end string the *preterminal string* (of the phrase-marker). Finally, we say that *s* is a *preterminal string in the derived phrase-marker pm* iff there exists a substructure *pm** of *pm*, such that 1) *pm** and *pm* have the same start element, and that 2) the end string of *pm** is the preterminal string of *pm*.

5. The Formal Part of **CLT**

With these initial preparations, we can formulate an empirical theory in the way of structuralism: "the" classical linguistic transformational theory (**CLT**).

In addition to the components discussed, there are three others that are central to the model of **CLT**: the set *Sent* of *sentences* (in a language), the *deep structure ds*, and the *surface structure ss* (of sentences). The set of sentences is independent of the set of end strings generated from the markers. We emphasize this independence since this is hardly mentioned in the field of linguistics. The two other terms *ds* and *ss* were introduced by Chomsky.

D7 x is a potential model of the classical linguistic transformation theory ($x \in \mathbf{M_p}(\mathbf{CLT})$) iff there exists W, $Sent$, PC, LC, F, CS, H, Str, $\circ$, Λ, Σ, Ξ, S, R^{pn}, R^{sc}, R^{lx}, R^{tm}, R^{pm}, ds, ss, E^{chy}, R^{chy}, DPM, TM, such that the following is true:

1) $x = \langle Sent, E^{chy}, Str, \circ, \Lambda, \Sigma, \Xi, S, R^{chy}, DPM, TM, ds, ss \rangle$

2) $E^{chy} = W \cup PC \cup LC \cup F \cup CS \cup H$, $R^{chy} = R^{pn} \cup R^{sc} \cup R^{lx} \cup R^{tm}$ and $R^{pm} = R^{pn} \cup R^{sc} \cup R^{lx}$
3) $\langle E^{chy}, \Sigma, R^{pm} \rangle$ is a frame and $DPM = DPM(E^{chy}, \Sigma, R^{pm})$ (see D1, the set of derived phrase markers, D3-b).
4) $\langle DPM(E^{chy}, \Sigma, R^{pm}), \Xi, R^{tm} \rangle$ is a frame (see D1).
5) *Sent* is a non-empty set ("sentences").
6) *W*, *PC*, *F*, *H* are pairwise disjunct sets $LC \subseteq F$ and *W*, *PC*, and *LC* are not empty.
7) $S \in PC \cap S$
8) $\{+, -, \Diamond\} \subseteq H$
9) $CS = CS(F)$ (see D6)
10) $R^{pn} \subseteq PC \times (PC \cup LC)^*$, $R^{sc} \subseteq LC \times CS(F)$, $R^{lx} \subseteq CS(F) \times W^*$, and $R^{tm} \subseteq DPM \times DPM^*$
11) $TM = TM(DPM(E^{chy}, \Sigma, R^{pm}), \Xi, R^{tm})$ is the set of transformation markers and $\Xi \subseteq DPM(E^{chy}, \Sigma, R^{pm})$ (See D3-c).
12) $\langle Str, \circ, E^{chy}, \Lambda \rangle \in SS$ (a string structure, see D4).
13) $ds: Sent \rightarrow DPM$
14) $ss: Sent \rightarrow TM$

In D7-7 the special symbol S is on the one hand a phrase category and on the other hand a special start element (see D1).

Lemma 4: $LC \subseteq CS(F)$.
Proof: D7-5), D6-1).

D7-13 only states that the deep structure of a sentence takes the form of a (derived) phrase-marker, and D7-14 that the surface structure of a sentence takes the form of a transformation-marker. Using hypotheses in D8, these components will now be supplied with content.

D8 x is a *model of the classical linguistic transformational theory* ($x \in \mathbf{M}(\mathbf{CLT})$) iff there exist *W*, *Sent*,*PC*, *LC*, *F*, *CS*, *H*, *Str*, $\circ$, Λ, Σ, Ξ, *S*, R^{pn}, R^{sc}, R^{lx}, R^{tm}, R^{pm}, *ds*, *ss*, E^{chy}, R^{chy}, *DPM*, *TM* such that: $x = \langle Sent, E^{chy}, Str, \circ, \Lambda, \Sigma, \Xi, S, R^{chy}, DPM, TM, ds, ss \rangle$ and the following is true:

1) $x \in \mathbf{M}_{\mathbf{p}}(\mathbf{CLT})$
2) for every $sent \in Sent$ there exist $w_1, ..., w_n \in W$, such that $sent = w_1 \circ \ldots \circ w_n$
3) for every $w \in W$ there exists $s \in E^{chy}$, such that $\langle s, w \rangle \in R^{lx}$
4) there exist $w_1, \ldots, w_n \in W$ such that $w_1 \circ \ldots \circ w_n \notin Sent$
5) there exists $e \in PC^*$ such that $\langle S, e \rangle \in R^{pn}$
6) for every $r \in R^{pn}$ there exist $pc \in PC$ and $e \in (E^{chy})^*$, such that $r = \langle pc, e \rangle$
7) for every $pc \in PC$ there exists $e \in (E^{chy})^*$ such that $\langle pc, e \rangle \in R^{pn}$
8) for every $r \in R^{tm}$, $pm, pm' \in DPM$ and $b_1, \ldots, b_n, b'_1, \ldots, b'_m \in E^{chy}$, the following is true: if $r = \langle pm, pm' \rangle$, $\boldsymbol{end}(pm) = b_1 \circ \ldots \circ b_n$,

$\mathbf{end}(pm') = b'_1 \circ \ldots \circ b'_m$ then $m, n > 1$, and m and n, and the sets $\{b_1, \ldots, b_n\}$, $\{b'_1, \ldots, b'_m\}$ are approximatively equal[24]

9) every feature $f \in F$ is used in a rule from R^{sc}

10) for every $sent \in Sent$ there exists a preterminal string s, $s = e_1 \circ \ldots \circ e_m$ in $ds(sent)$, such that there is no e_i for which a rule $r \in R^{pn} \cup R^{sc} \cup R^{tm}$ exists with $r = \langle e_i, y \rangle$.

11) for every $sent \in Sent$ there exist

11-1) a derived phrase-marker $pm \in DPM(E^{chy}, \{S\}, R^{pm})$ such that *sent* is the end string of *pm* and *end*(*ds*(*sent*))=*sent*, or

11-2) a transformation-marker $tm \in TM$, such that *sent* is the second-level end string of *tm* and **end** (*ss*(*sent*))=*sent*.

Lemma 5: In D8-11-1), *pm* and in D8-11-2) the "last" markers of *ss*(*sent*) are closed derived phrase-markers.
Proof: D5), D8-2).

In other words, the hypotheses state the following: in accordance with 3), every word in a lexical insertion rule is used. 4) says that there are series of words that are not sentences. Without 5), there would be models whose set of sentences Sent is empty. Every phrase rule in 6) begins with a phrase category, and 7) states that all phrase categories in these rules are used as well. 8) expresses a kind of conservation law not found explicitly in Chomsky's texts, but which we consider important in distinguishing transformations from the other rules and other methods of generation. Ideally, the sets of Chomsky-elements and the end strings of two transformation-markers involved in a transformation rule are identical. For example, in English, one method of transforming a *verb* from the *active* to the *passive* form is to insert the additional word *be* (or other variants thereof). The relations of similarity $\sim_1, \sim_2$ in footnote (25) and in D8-8) are therefore essential.

In Chomsky (1965, p. 84), the term preterminal string is defined for a sentence in which the creation of the sentence is divided into two segments. On the one hand, the preterminal string completes the generation of the sentence up to the point that, in all additional sentences, only lexical insertion rules are applied. On the other hand, the preterminal string must be generated with the help of sub-category rules,[25] such that the Chomsky-elements of the preterminal strings are complex symbols. As long as the system of grammatical rules in a particular language is not made more explicit than this, it is hardly to be expected that a preterminal string of a sentence can be uniquely determined. We have therefore grasped the content in 10) in such general terms that, viewed from a sentence's preterminal string, only lexical insertion rules can be used.

---

[24] The similarity relations $\sim_1$, $\sim_2$ can be easily defined using the approximation apparatus *A* of **CLT**, see e.g. Balzer & Zoubek (1994): $m \sim_1 n$ and $\{b_1, \ldots, b_n\} \sim_2 \{b'_1, \ldots, b'_m\}$.

[25] The differentiation between context-free and context-related sub-categories was greatly discussed in the 1960's.

Hypothesis 11) represents the central statement of Chomsky's "classical" approach. Informally, every sentence *sent* from the set *Sent* (of a given language) can be generated as follows: a phrase-marker or a derived phrase-marker pm_n is generated such that pm_n begins with the start element *S*. If *n* equals 1, the creation of the sentence ends immediately. The deep structure *ds*(*sent*) is identical to the phrase-marker pm_n, and the end string of pm_n is identical to the sentence *sent*. If *n* is greater than 1, several open (derived) phrase-markers pm_1,..., pm_{n-1} and a closed phrase-marker pm_n are generated. The open phrase-markers thereby form the basis of a transformation-marker *tm*. The transformation-marker's second-level end string is then identical to the sentence *sent*. In both cases, a "circle" is closed at the end of the generation process. The end string of the surface structure *ss*(*sent*) is identical to the sentence with which the process started. In other words, the surface structure's end string can be obtained using a process in which the end string "somehow originates" from "the given" sentence *sent*. It is essential that this process has a deep structure for the given sentence.

A few other simple formal statements follow directly from the hypotheses. If no sub-categories are used, then the fire-part is a lexical category for every lexical insertion rule. If no other features are used besides the lexical categories, the set of complex symbols contains only the features, and no sub-categories are employed. The set *CS* is uniquely determined by *F* and $\{+, -, \Diamond\}$. Finally, we will formulate a (trivial) theorem from the theory of science:

Theorem: There exists a model for **CLT**.

From the standpoint of the theory of science, a second component is important. The term *constraints* for a theory[26] was not yet used at the time when **CLT** was developed, but it was discussed using a different vocabulary. A constraint for **CLT** is a class of combinations (sets) of potential models described by a "second-level" hypothesis. We would like to introduce briefly four constraints that cannot really be counted as parts of the reconstruction, but which are nevertheless relevant for the theory treated here.

The first constraint **C**1 pertains, however, to the phonetic level not discussed here. This constraint states that in all languages the set of utterances can be represented with the help of the same set of phonologic elements. A second constraint **C**2 combines groups of languages whose words and sentences are written using the same alphabet, e.g. written languages using Latin, Cyrillic, Hebrew, or other alphabets, or those which use Chinese characters. A third constraint **C**3 defines groups of languages that can be analyzed using the same set of "syntactic-semantic features". In Chomsky (1965), several sub-category rules are established which are used in the same way for all lexicons of languages in this group. For example, a differentiation is made between the *nouns* in *human* vs. *non-human* or *male* vs. *female*, and between the *verbal phrases* in *transitive* vs. *intransitive*.

[26] See Balzer, Moulines & Sneed (1987, Chap. II).

A final constraint **C4** contains sets ("groups") of languages that use the same categories and rules. Chomsky argued at length that this constraint **C4** is universal, i.e. that there are some categories and rules that apply to all languages. For the theory of science, this would mean that this "universal constraint" **C4** is not a well demarcated part of **CLT**, for in the following formulation, this constraint **C4** would be identical to the power set of models: **C4** = *Pot*(**M**). According to our current state of knowledge, this question remains open, e.g. in Han-Chinese or in Japanese. One example of this constraint is the subject-verb connection in the family of Indo-European languages, as in: *Peter goes*. For a component k from $x \in \mathbf{M_p}(\mathbf{CLT})$, we also write k_x. **C4** is the *Indo-European constraint* of **CLT** iff there exist $PN, VP, V \in PC$ such that: 1) $\mathbf{C4} \subseteq Pot(\mathbf{M_p}(\mathbf{CLT}))$ and 2) $\forall X(X \in \mathbf{C4} \leftrightarrow \forall x, y(x, y \in X \rightarrow \{NP, VP\} \subseteq PC^x \rightarrow \{NP, VP\} \subseteq PC^x \cap PC^y \wedge V \in LC^x \cap LC^y \wedge \{\langle S, \langle NP, VP\rangle\rangle, \langle VP, V\rangle\} \subseteq (R^{pn})^x \cap (R^{pn})^y))$.

6. The Empirical Part of **CLT**

As described in Sec. 2, the approximative empirical claim for **CLT** has the form of an existential quantification $\exists X(X \subseteq \mathbf{M} \wedge X \in \mathbf{C} \wedge I \approx r(X))$ whose content can be summarized as follows: 1) all natural languages examined in **CLT** are analyzed using the same categories and rules, 2) the analysis process and the creation of sentences follows the four types of rules in the "correct" order, 3) the sentences analyzed and identified, as well as their corresponding words, are matched with sentences from a model.

This approximative empirical claim of **CLT** uses the set of intended applications described with terms that were already used before **CLT** was created. In addition to these terms, "new" **CLT**-theoretical terms are used which were coined especially for this theory.[27] The **CLT**-non-theoretical terms for **CLT** are identified using methods which already existed before **CLT**, and we do not need to discuss them here. The sentences, words, the concatenation relation $\circ$ and the blank Λ can be determined independently of **CLT**. This likewise applies for the terms *PC*, *LC*, S, R^{pn}, and R^{lx}, which existed prior to **CLT**.

Even without a more exact description of the criteria which a theory must fulfill (Sneed 1971, Moulines 1985), one can see that the functions ds and ss are **CLT**-theoretical. Both the content and the form of these functions can only be analyzed with the help of the theory **CLT**. This also applies, in principle, to the set of transformation rules and to the term R^{sc}, which was used for the first time in **CLT**.[28] By contrast, it is difficult to say whether or not the term *Str* is theoretical relative to **CLT**. Expressed informally, the term *Str* and the corresponding set

[27] See e.g. Balzer, Moulines & Sneed (1987, p. 47).

[28] We have not examined the relationship between the transformation terms of Harris (1954) and Chomsky.

of strings would have to be divided into two "halves". This also holds true for the cognates of *Str*, the sets *F*, *H*, *CS*. Those strings created using only elements from *W*, *PC*, *LC*, *S* can be determined without **CLT**. The "mixed" strings also containing elements from the new sets *F*, *H*, *CS* or the partially new set Σ can be delineated more precisely only with the help of the theory **CLT**. This short illustration remains somewhat unsatisfying, since we have not described any methods of determination more precisely.

D9 y is a partial model of $\mathbf{CLT}\,(y \in \mathbf{M_{pp}}\,(\mathbf{CLT}))$ iff there exists *W*, *Sent*, *PC*, *LC*, *F*, *CS*, *H*, *Str*, $\circ$, Λ, Σ, Ξ, *S*, R^{pn}, R^{sc}, R^{lx}, R^{tm}, *ds*, *ss*, E^{chy}, R^{chy}, *DPM*, *TM* such that

1) $\langle Sent, E^{chy}, Str, \circ, \Lambda, \Sigma, S, R^{chy}, DPM, TM, ds, ss\rangle \in \mathbf{M_p}\,(\mathbf{CLT})$
2) $y = \langle W, Sent, PC, LC, Str, \circ, \Lambda, \Sigma, S, R^{pn}, R^{lx}\rangle$

F, *CS*, *H*, R^{tm}, R^{sc}, *ds* and *ss* are the **CLT**-theoretical terms.

The sets of markers are explicitly defined by the potential models' other components. We have listed them explicitly as components merely for ease of reading and understanding.

The intended applications of **CLT** anchor the empirical claim within reality. Viewed very idealistically, an intended application "is" a natural language, and from a structural point of view, a partial model. A partial model consists of a tuple (a "list") of sets. The interesting elements of these sets are data. Within the empirical theory, a datum that can be formulated with the help of an atomic sentence, terms for the base sets of the theory **CLT** and with the help of "names" (i.e. expressions for elementary entities, "nominal phrases"), is observed, defined, measured, or developed. Metaphorically speaking, the actual facts are pressed through a filter so that within the theory **CLT**, they can be described as data using very few sentence forms. These "filtered" data for **CLT** are comprised of many atomic sentences consisting of the base terms of $\mathbf{CLT}, \circ, \Lambda, \Sigma$, *S*, *ds*, *ss*; of elements from *Sent*,...*E*..., and of elements from R^{chy} and from the defined terms *DPM* and *TM*. In other words, the many forms of data that exist and are produced in linguistics are reduced to a few basic forms.

The methods for defining and measuring that were used at the time of **CLT** are manifold; they are not discussed in Chomsky (1965) in depth.[29] They span from physical, technical methods by which utterances are recorded and documented, to different types of recognition, notation, and systemization of expressions, to participatory methods in which the experimenters are actively integrated into the sessions. A very large group of methods applied at the sound level –and deeper– could not even be described in our reconstruction. We believe, however, that in principle, the results of such methods can easily be transferred to the purely syntactic level *without* changing the structural composition of our model.

[29] In our article we focus on the structure of the hypotheses and the models. We hope we can supplement this lacuna by a future paper.

In the field of linguistics, a data set arising from a particular application is called a *corpus of data*. If, contrary to fact, we assume that the linguistic analyses in the area of **CLT** generate results in the standardized form described, we can say that a corpus of data forms a nucleus of a substructure of a partial model of **CLT**. Realistically, the actual data from a corpus of data must be "translated" into a standardized set of atomic formulas of **CLT**, often requiring that additional elements be added.

Using the structuralist theory of science, we thus idealize an area which predates the models and theoretical hypotheses. Still other idealizations are used here and/or in the linguistic community.

A second idealization is abstracted from the component of time that is essential to every natural language. The words, the pronunciation, and the linguistic rules of a language change to a certain extent over time. The linguistic works written in Chomsky's time and earlier employed a strategy, which Saussure described as "synchronic", of creating an initial static image. We have likewise employed this strategy in our reconstruction.

A third idealization makes a rough delineation between similar natural languages. Are, for example, the English and the American languages the same, or the German and the Swiss? This leads directly to the sets of words and rules that differ in two models; this then leads, among other things, to Chomsky's dissertation (Chomsky 1975), which discusses different linguistic levels left implicit in Chomsky (1957, 1965).

A fourth idealization levels the manifold aspects of natural languages. Different groups of individuals, such as children, adolescents, and adults, or educated and uneducated people, speak the same language differently. There are also many dialects, such as the Texan dialect, for example, or dialects from New York, which can easily be differentiated from one another. This means from our structuralistic point of view that the sets of words, pronunciations and rules can change a bit. From Chomsky's point of view, this normal idealization is even more enhanced, since he views these sets as highly stable and generalized.

The global constraint $\mathbf{C}1 \cap \mathbf{C}4$ appears to be only partially corroborated. If one leaves this constraint out, the claim in (1) deteriorates into "local" claims that each apply to a partial model (a "language"). $\exists X(X \subseteq \mathbf{M} \wedge I = r(X)) \leftrightarrow \forall y(y \in I \leftrightarrow \exists x(x \in \mathbf{M} \wedge \mathbf{r}(x) = y))$. That $i_0 \in I \wedge \exists x(x \in \mathbf{M} \wedge \mathbf{r}(x) = i_0)$ is true for a particular natural language i_0 seems unlikely to us, even without special knowledge of linguistics.

7. The Theory-Net of **CLT**

We can identify two "extreme points" in an ideal empirical claim in (1). In the first, a theory has *no* content; in the second, it is falsified by the theory's intended

systems, and thereby by its data. For theories that develop into theory-nets, there is a third alternative. In a theory-net, the base element can be without content, and yet be very successful as the framework for a net of easily falsifiable specializations. For several theory-nets, it was possible to confirm that the basic element has no content,[30] and that over time, specializations are falsified and thereby removed from the net.[31]

We could not precisely define the formal content of **CLT**. We can only conjecture that the basic element **CLT** itself has little content. By contrast, the specialization for a particular natural language is so rich in content that it could be swiftly falsified using data.

The concept of specialization was not known in the period discussed here. We cannot actually say that we are reconstructing specializations "of" **CLT**, since the "latent sub-theories" of **CLT** are, formally, simply specializations of **CLT**. In the following, we will simply continue to speak of specializations *of* **CLT**. The numerous specializations that develop effortlessly from **CLT** cannot be discussed here. In an *addendum* to Chomsky (1957), a clearly formulated forerunner system for **CLT**, a system of rules (without sub-categories) for the English language, is described. The rules of this one specialization alone – without annotations – fill several pages. We can presume that specializations of **CLT** exist for every natural language. The theory-net for **CLT** is thus fairly large, and we can therefore mention it only very briefly. A specialization of **CLT** contains additional transformation rules and subcategory rules that apply only to particular languages. In the simplest case, it has only one single intended application. Other specializations apply to groups of languages, such as the German, Latin, or Slavonic language groups.

In the original concept of specialization in Balzer, Moulines & Sneed (1987), an intended application consisted of all non-theoretical components of a model. With **CLT**, it was necessary that the set of sentences in the partial models, and thereby in the intended applications, appear in full. This leads to a twofold problem in **CLT**. On one hand, all the strings ("hypothetical sentences") generated in a model had to be present in a given intended application. On the other hand, it was necessary to be able to generate every expressed sentence that was part of the application using the rules. This identity of hypothetical and expressed sentences, established per definition as such, is hardly realistic. Using the hypotheses of **CLT**, it is possible to generate very long strings that no person could utter, and there are printed sentences in a particular language that, in all likelihood, could not be generated from the area of intended sentences using the rules of a specialization of **CLT**.[32]

The concept of *generalized specializations* introduced at the end of Sec. 2 solves both problems. A partial model is "reduced" by removing parts in such a way that

[30] For example in classical particle mechanics and in thermodynamics, Balzer, Moulines & Sneed (1987).

[31] For more precise information, see Bartelborth (1996, Chap. VII).

[32] There are, of course, variations in which a constant limits the elements in a string.

the "rest" remains of the original type. In the generalized empirical claim of **CLT**, there can now be two types of problems of fit. How can we fit sentences which were actually expressed to hypothetical sentences which we find in a model of **CLT**. In the first type of problem, there is a sentence expressed that does not lie within the hypothetical set of sentences. This is the standard case of falsification. An actual expressed sentence cannot be generated in a model of the theory. In the second type of problem, there is a generated string that was not expressed.

The first problem we solve as follows: we introduce a new specialization in which there are models that – relative to the basic-element **CLT** – can generate "new" sentences. We then choose a model of the new specialization, such that the set of sentences describing the intended application is a *subset* of the set of sentences belonging to the new model. However, we cannot guarantee that this method always works. In principle, it could happen that the "stubborn" sentence from the intended application cannot be generated using any set of rules possible within the framework of the basic-element of **CLT**. This leads to the formal question whether the theory **CLT** has empirical content (see Balzer, Moulines & Sneed 1987, p. 92). We were unable to answer this formal question. If this theory has empirical content, we would presume that the empirical case of a "found" sentence which cannot be generated by the rules of **CLT** would occur only with very low probability. We wish to point out that within a theory-net, two theories – specializations – can be formally inconsistent. Within a net, many specializations (or in the case of **CLT**, many different systems of rules) can be introduced and verified. The second type of problem is solved by creating subsets and reducing the problems to problems of the first kind.

In **CLT**, a connection is made between a corpus of data and a partial model of a specialization of **CLT** using the approximation relation of **CLT**. On one hand, the data are prepared and standardized in such a way, that a set of "empirically determined" sentences ("ep-sentences") and their elements, the words, is given. This treatment, which usually occurs at the phonetic level, can be fairly involved. On the other hand, a model and the set of sentences generated theoretically from this model ("th-sentences") are given. Depending on the number of ep-sentences from the corpus that do not fit, a decision is made as to whether or not the corpus fits to a partial model at hand. This decision normally depends upon a constant applied –often without further explanation especially in **CLT**.

The generalized theory-nets have other positive aspects. First, it is possible that several intended applications be generated from a single actual system. For example, one can formulate a "global" specialization of **CLT** that refers to an actual system in which the English language is spoken. It is possible, however, to actually delineate several real partial systems. Regions such as Scotland, India, and the Bronx, USA, have particular units that can be described through special rules. In this way, further specializations of **CLT** are generated, as well as intended applications limited to partial systems. Interestingly, using this generalization it is also

possible that a linguistic corpus is identical to an intended application. Secondly, one can also ontologically delineate different intended applications originating from the same system. A boundary – regardless of how it originally came to be – between two or more subsets is supported empirically with respect to different objects. Third, the rules in Chomsky (1957) form a good example for these points. We guess that one cannot generate all English sentences using these rules. If this is true the rules in Chomsky (1957) generate only a fragment of the English language. However, additional rules could be applied or used to replace others. Extensions of the set of rules inconsistent with the rules used before can be tested as well. All these possibilities can be illustrated using specializations of **CLT**. Fourthly, probably all empirically analyzed specializations of **CLT** are falsifiable.

8. Conclusion

The examination of a no longer current ("dead") theory, newly formulated and in a clearer representation, leads, from the perspective of the theory of science, to two new insights.

First, 60 years ago, the representation of the concept of an empirical theory was fairly similar to the linguistic formulation of the concept of grammar. An empirical theory was viewed as a deductive, closed set of sentences derived from hypotheses and data ("observation sentences"). This essential concept was used directly in linguistic grammars, such that questions from the theory of science could be discussed "directly". There, these questions were often reduced to their syntactic aspects.[33] The question of delineating "one" language was hardly discussed. In our formulation using the theory of science, the theory **CLT** cannot be seen as – or even reduced to – a pure grammar. Rather, it becomes clear that an intended application ("a language") cannot be adequately delineated by means of a deductively closed system.

Secondly, we were able to fit the empirical portions of **CLT** into our structuralist framework in such a way that a language can be made to fit a model using actual data. Contact between the data and the data sets and the linguistic models, which appears to be a given for many linguists, could not, from the standpoint of the theory of science, be made without difficulty. It was necessary to use specifically the concept of the intended application (Sneed 1971) to bring order to the many systems of data, and to assign them to the actual parts of reality. This problem is intensified in the field of linguistics, since there the linguistic elements in the linguistic models are, on the one hand, essential elements from which most other model components are composed, but on the other hand are often filled with

[33] Chomsky made, for example, a classification in which a theory is divided into three success levels: the observational, descriptive, and explanatory levels of a theory. Chomsky (1962, p. 63). However, this discussion did not yield any lasting results for the theory of science.

theoretical content of the highest level. In these cases, theory laden data cannot be fit to a model using the usual statistical methods at the observational level, such as goodness of fit. We found this result noteworthy for the philosophy of science. For **CLT**, we were unable to make contact between a corpus of data and a model using the structuralist "standard" approximation relation, but were rather required to represent the fit between data and models – conveyed by the intended applications – using generalized theory-nets.

We were able to clearly distinguish between the statistical method (Clark 1992) usual in linguistics, with which data and corpora of data are collected and produced, and the fit between the intended applications and models of the theory, which are given using empirical claims, approximation, and the net concept.

References

Balzer, W., Moulines, C. U. and J. D. Sneed (1987), *An Architectonic for Science*, Dordrecht: Reidel.

Balzer, W., Lauth, B. and G. Zoubek (1993), "A Model for Science Kinematics", *Studia Logica* 52: 519-548.

Balzer, W., Sneed J. D. and C. U. Moulines (eds.) (2000), *Structuralist Knowledge Representation. Paradigmatic Examples*, Amsterdam-Atlanta: Rodopi.

Balzer, W. and G. Zoubek (1994), "Structuralist Aspects of Idealization", in Kuokkanen, M. (ed.), *Idealization VII: Structuralism, Idealization and Approximation*, Poznan Studies in the Philosophy of the Sciences and the Humanities 42, Amsterdam: Rodopi, pp. 57-79.

Bar-Hillel, Y., Kasher, A. and E. Shamir (1963), *Measures of Syntactic Complexity, Report for U.S. Office of Naval Research*, Jerusalem: Information Systems Branch.

Bartelborth, T. (1996), *Begründungsstrategien*, Berlin: Akademie Verlag.

Bloomfield, L. (1933), *Language*, New York: Holt, Rinehart and Winston.

Bourbaki, N. (1968), *Theory of Sets*, Berlin - Heidelberg: Springer.

Chomsky, N. (1953), "Systems of Syntactic Analysis", *The Journal of Symbol Logic* 18: 242-256.

Chomsky, N. (1957), *Syntactic Structures*, Berlin-New York: Mouton de Gruyter, 2nd ed. 2002.

Chomsky, N. (1961), "On the Notion of Grammar", *Proceedings of the Twelfth Symposium in Applied Mathematics* XII: 6-24; reprinted in Fodor & Katz (1965), pp. 119-136.

Chomsky, N. (1962), "Current Issues in Linguistic Theory", *Ninth International Congress of Linguists*; reprinted in Fodor & Katz (1965), pp. 50-118.

Chomsky, N. (1965), *Aspects of the Theory of Syntax*, Cambridge, Mass.: MIT.

Chomsky, N. (1975), *The Logical Structure of Linguistic Theory*, New York and London: Plenum Press.

Clark, R. (1992), "The Selection of Syntactic Knowledge", *Language Acquisition* 2: 83-149.

Diederich, W., Ibarra, A. and T. Mormann (1989), "Bibliography of Structuralism 1971-1988", *Erkenntnis* 30: 387-407.

Diederich, W., Ibarra, A. and T. Mormann (1994), "Bibliography of Structuralism II 1989-1994 and Additions", *Erkenntnis* 41: 403-418.

Fodor, J. A. and J. J. Katz (eds.) (1965), *The Structure of Language*, Englewood Cliffs NJ: Prentice-Hall.

Ginsburg, S. and B. Partee (1969), "A Mathematical Model of Transformational Grammars", *Information and Control* 15: 297-34.

Gonzalo, A. (2001), "Cambios modeloteóricos en la lingüística chomskiana. Una reconstrucción desde la concepción estructural de la ciencia", Dissertation, Universidad de Buenos Aires.

Halle, M. (1962), "Phonology in Generative Grammar", *Word* 18: 54-72.

Harris, Z. (1951), *Methods in Structural Linguistics*, Chicago: University of Chicago Press.

Harris, Z. (1954), "Distributional Structure", *Word* 10: 146-162; reprinted in Fodor & Katz (1965), pp. 34-49.

Ho, R. (2006), *Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS*, London-New York: Chapman and Hall/CRC.

Hockett, C. F. (1954), "Two Models of Grammatical Description", *Word* 10: 210- 231.

Hornby A. S. *et al.* (eds.) (1974), *Oxford Dictionary. Oxford Advanced Learner's Dictionary of Current English*, Oxford: Oxford University Press.

Moulines, C. U. (1985), "Theoretical Terms and Bridge Principles: A Critique of Hempel's (Self-)Criticisms", *Erkenntnis* 22: 97-117.

Peris-Viñé, L. M. (1990), "First Steps on the Reconstruction of Chomskyan Grammar", in Díez, A., Etcheverrìa, J. & A. Ibarra (eds.), *Structures in Mathematical Theories*, San Sebastián: Servicio Editorial Universidad del País Vasco, pp. 83-87.

Peris-Viñé, L. M. (1996), "Caracterización de las nociones básicas de la Gramática de Chomsky", *Ágora* 15(2): 105-124.

Peris-Viñé, L. M. (2010), "Estructura parcial de la gramática estándar del castellano", in Peris-Viñé, L. M. (ed.), *Filosofía de la Ciencia en Iberoamérica: Metateoría Estructural*, Madrid: Tecnos, pp. 223-256.

Peris-Viñé, L. M. (2011), "Actual Models of the Chomsky Grammar", *Metatheoria* 1: 195-225.

Quesada, D. (1993), "Grammar as a Theory. An Analysis of the Standard Model of Syntax within the Structural Program", in Díez, A., Etcheverrìa, J. & A. Ibarra (eds.), *Structures in Mathematical Theories*, San Sebastián: Servicio Editorial Universidad del País Vasco, pp. 175-182.

Sneed, J. D. (1971), *The Logical Structure of Mathematical Physics*, Dordrecht: Reidel.

Stanley Peters Jr., P. and R. W. Ritchie (1973), "On the Generative Power of Transformational Grammars", *Information Sciences* 6: 49-83.

Wells, R. S. (1947), "Immediate Constituents", *Language* 23: 81-117 and 169-180.