

La gramática como ciencia

Carlos Muñoz Pérez
cmunozperez@filo.uba.ar
Universidad de Buenos Aires

1. Introducción

Empecemos con una definición operativa, casi de sentido común:

- (1) *Lingüística (definición informal)*
La ciencia que estudia el lenguaje.

¿Pero qué entendemos por “científico” cuando nos referimos al lenguaje?

Une étude est dite scientifique lorsqu'elle se fonde sur l'observation des faits et s'abstient de proposer un choix parmi ces faits au nom de certains principes esthétiques ou moraux. "Scientifique" s'oppose donc à "prescriptif". (Martinet 1960, apud Newmeyer 1985).

La distinción entre descripción y prescripción no fue el único parámetro para definir a la lingüística como ciencia.

Where we deal with man we have not yet progressed beyond the animistic view. In spite of the fact that the measured outcome of human activity (as, say, in lifting weights or in a metabolism test) never conflicts with the physical conditions, we insist that within each man and yet not as a part of his body there exists an invisible, supra-physical double, a "soul," "spirit," "mind," "will," "consciousness," or what not, which determines the actions of the man. Conveniently, of course, this invisible double is not subject to the regular sequences of events which science has found in the rest of the universe. Accounting with magic ease for anything and everything that may happen, animism obscures every problem and drugs us into accepting our ignorance and helplessness in the face of human conduct. (Bloomfield 1933: 555).

Un momento: estas nociones de “cientificidad” no se parecen en nada al criterio de demarcación introducido por Popper: el falsacionismo.

¹ El título original del libro fue *Methods in Structural Linguistics*. El hecho de que la presencia de la palabra “métodos” no haya implicado ninguna revisión en el contenido

A partir de este pequeño problema, intentaremos empezar a responder dos preguntas:

- ¿a partir de qué momento el estudio de la gramática se transforma en ciencia en el sentido popperiano?
- ¿cómo es la práctica de este tipo de gramática?

2. El estructuralismo descriptivista norteamericano

(Vamos a caracterizar un poco al llamado *estructuralismo norteamericano*. Hagamos esto teniendo en mente algunos términos: (i) inducción, (ii) explicación y (iii) contrastabilidad).

El mote de “descriptivista” ya nos dice bastante acerca de cómo desarrollaba sus investigaciones esta escuela lingüística.

Bloomfield was not himself particularly interested in methodology, but many of his students and followers were, and procedures were soon developed to find or "discover" the various elements which combine to make up the forms of a language. These procedures all had in common the division of the stream of speech into successively smaller and smaller units (segmentation) and the grouping together of the units that occur at a given level of analysis (classification). The attempts to codify these procedures and the disputes that raged over those attempts dominated the discussions and publications of the period from roughly 1935 to 1955, and gave rise to the perception that linguistics, as practiced by the Bloomfieldian structuralists is a 'taxonomic' science. (Langendoen 1983).

La obra más importante de Zellig Harris, *Structural Linguistics*¹, empieza con las siguientes palabras:

This volume presents methods of research used in descriptive, or, more exactly, structural linguistics. It is thus a discussion of the operations which the linguist may carry out in the course of his investigations, rather than a theory of structural analyses which result from these investigations.

Es, por supuesto, una total exageración decir que la lingüística estructural era completamente “metodológica” o puramente “taxonómica”. Había, por supuesto,

de la obra es un indicio del carácter metodológico (y no puramente teórico) del estructuralismo post bloomfieldiano.

predicciones teóricas. Consideren el carácter falsacionista de la siguiente observación de Hockett:

The simplest and earliest assumption about the relation between morphemes and phonemes was that a morpheme is **composed** of phonemes: the morpheme *cat* is composed of the phonemes /k/, /æ/, /t/ in that arrangement.... This conception is satisfactory as long as we confine our attention to certain parts of the material in any language, but it breaks down when we are confronted with such phenomena as English *knife*: *kniv(s)* or *(dog)s*; *(cat)s* and *ros(es)*. The very criteria that lead us to reject a morphemic identification of *cat* and *cat(alog)* ... also lead us to want to call *knife* and *knife-*, or /z/, /s/, and /iz/, the same morpheme. Such identifications are incompatible with an acceptance of the composed of relation between morphemes and phonemes. (Hockett 1961, apud Langendoen 1983).

En un artículo titulado *The Status of Linguistics as a Science*, Edward Sapir observa:

The more we devote ourselves to the comparative study of the languages of a primitive linguistic stock, the more clearly we realize that **phonetic law and analogical leveling** are the only satisfactory key to unravelling of the development of dialects and languages from a common base. Professor Leonard Bloomfield's experiences with Central Algonquian and my own with Athabaskan leave nothing to be desired in this respect and are a complete answer to those who find it difficult to accept **the large scale regularity of the operation of all those unconscious linguistic forces which in their totality give us regular phonetic change and morphological readjustment on the basis of such change** (Sapir 1929).

La defensa de las leyes de cambio de los sonidos (i.e., proposiciones de carácter teórico sobre el comportamiento de los fonos) contrasta fuertemente, sin embargo, con el resto del artículo. Para Sapir, la lingüística debe ser considerada una ciencia por el valor que el estudio del lenguaje tiene para otras disciplinas (e.g., antropología, ciencias sociales, psicología, etc.).

It is peculiarly important that linguists, who are often accused, and accused justly, of failure to look beyond the pretty patterns of their subject matter, should become aware of what their science may mean for the interpretation of human conduct in general. Whether they like it or not, they must become increasingly concerned with the many

² Para una discusión de este esquema, con énfasis en el papel de las teorías de la computabilidad, véase el artículo de Peregrín Otero (1970) disponible en el Campus.

anthropological, sociological, and psychological problems which invade the field of language.

El problema en esta manera de entender la lingüística fue señalado por Chomsky:

As I interpret most of the more careful proposals for the development of linguistic theory, [...] they **attempt to state methods of analysis that an investigator might actually use, if he had the time, to construct a grammar of a language directly from the raw data**. I think that it is very questionable that this goal is attainable in any interesting way, and I suspect that any attempt to meet it will lead into a maze of more and more elaborate and complex analytic procedures that **will fail to provide answers for many important questions about the nature of linguistic structure**.

Las partes subrayadas hacen referencia a dos problemas en el enfoque general del descriptivismo: (i) su uso generalizado de la lógica inductiva y (ii) su carácter taxonómico (y no explicativo).

3. El carácter falsacionista de Estructuras Sintácticas

Más allá del debate (totalmente estéril) de si la gramática generativa significó una revolución en el sentido de Kuhn, creo que la siguiente formula ayuda a entender el contexto de desarrollo de dicho marco teórico:

- (2) Estructuras Sintácticas = CCII + Popper + Teorías de la computabilidad²
- No voy a explicar a qué refiere el término *Constituyentes Inmediados* (CCII).
 - No voy a volver sobre Popper.
 - Las teorías de la computabilidad refieren al desarrollo de *funciones* o *algoritmos* para el procesamiento de información.

Lo que sigue es una breve exposición de los capítulos 3 a 5 de Chomsky (1957).

Tomemos una oración como (3):

- (3) El chico fue al supermercado.

Dado que la oración (3) puede ser concebida como una secuencia lineal de palabras (muy a lo Saussure³), podríamos pensar que el lenguaje consiste en un algoritmo que

³ De hecho, el material de prácticos de *Sintaxis* (Giammatteo) está orientado a mostrar las consecuencias de una sintaxis basada únicamente en relaciones *in praesentia*.

procesa relaciones de adyacencia lineal. A este tipo de algoritmo se lo denomina *gramática de estados finitos*.

- (4) El chico fue al supermercado.
1 2 3 4 5

La idea es que el algoritmo funciona de estrictamente lineal: se procesa la palabra 1 primero, luego la palabra (2), luego la palabra (3) y así. Cuando estamos en el paso 1, la única información a la que podemos acceder en toda la cadena de elementos es la que está en el paso 2. Cuando estamos en el paso 2, sólo podemos acceder a la información del paso 3.

Supongamos que postulamos la siguiente hipótesis:

- (H1) La sintaxis de las lenguas naturales se corresponde con un algoritmo de estados finitos.

¿Qué tipo de predicciones hace una hipótesis de este tipo?

La predicción inmediata es que no puede haber relaciones sintácticas “a distancia”. Es fácil mostrar que esta predicción es errónea.

- (5) El chico [que conoce a mis amigos] fue al supermercado.

La demostración de Chomsky es más complicada. Él ejemplifica un lenguaje con dos ítems léxicos a y b, que tiene como “corpus” expresiones como las siguientes:

- (6) a. ab, aabb, aaabbb (*n* veces a y *n* veces b).
b. aa, bb, abba, baab, aaaa, bbbb, aabbaa (la segunda parte de la expresión es el “reflejo” de la primera)
c. aa, bb, abab, baba, aaaa, bbbb, aabaab (toda secuencia de a y b es seguida por una secuencia idéntica).

Una gramática de estados finitos no puede capturar estos patrones (e.g., ¿cómo a partir de adyacencia se podría explicar que si de un lado tengo 3 letras a del otro lado tengo que tener la misma cantidad?). Según Chomsky, patrones similares a estos se observan en las lenguas naturales (ejemplos parecidos a (5)).

- (7) a. If S_1 , then S_2
b. Either S_3 , or S_4 .
c. The man who said that S_5 is arriving today.

Para colmo, señala Chomsky, podemos hacer cosas más complicadas todavía:

- (8) If, either (7c), or S_4 , then S_2

Bien, ¿qué hicimos hasta acá?

- (9) *Reconstruyendo nuestros pasos (o los de Chomsky 1957)*
a. Postulamos una hipótesis H1: que la sintaxis de las lenguas obedece relaciones de adyacencia (las relaciones sintagmáticas son *in praesentia*, no hay “sintaxis incorpórea”).
b. Derivamos una predicción empírica de dicha hipótesis: no puede haber relaciones sintácticas a distancia.
c. Contrastamos empíricamente la predicción, mostrando que la hipótesis H2 es falsa.

¿A qué noción de *ciencia* está suscribiendo Chomsky a través de los pasos de su razonamiento?

Si vamos a “jugar el juego de la ciencia”, no nos basta con decir que H1 está mal: hay que proponer una alternativa superadora. Aquí entran los CCIL. Para decirlo mal y pronto, las oraciones de las lenguas naturales obedecen una estructura jerárquica, en la que las palabras se agrupan en *sintagmas* de complejidad y extensión creciente.

(10)

El	rey	francés	es	un	monarca	ficticio

Otra manera de representar (10) es ir “poniéndoles nombre” a cada uno de los sintagmas y especificar cuáles son sus constituyentes inmediatos.

- (11) Oración → Sintagma Determinante + Sintagma Verbal
Sintagma Determinante → Determinante + Sintagma Nominal
Sintagma Nominal → Nombre + Adjetivo
Sintagma Verbal → Verbo + Sintagma Determinante

Llamemos al algoritmo que procesa estructuras de este tipo *gramáticas de estructura de frase*. Podemos, así, postular una nueva hipótesis:

- (H2) La sintaxis de las lenguas naturales se corresponde con un algoritmo de estructura de frase.

La H2 nos permite lidiar satisfactoriamente con los datos de (7) y (8) (no pienso hacer las cajitas ni escribir las reglas, queda en ustedes), por lo que es una hipótesis

superadora con respecto a H1. Ahora bien, ¿qué tipo de predicciones se siguen de esta teoría?

Los propios estructuralistas habían observado algún problema con la noción de *constituyente inmediato*. El par de (12), por ejemplo, es de Hockett (1971).

- (12) a. [SUJETO Juan] [PREDICADO viene con ustedes].
b. [PREDICADO Viene] [SUJETO Juan] [PREDICADO con ustedes].

- (13) a. [SUJETO Juan] [PREDICADO compró una bicicleta para María].
b. ¿[PREDICADO Qué compró] [SUJETO Juan] [PREDICADO para María]?

Hockett denomina estos casos *constituyentes inmediatos discontinuos* (un verdadero oximoron).

Los ejemplos de Chomsky son un más complicados y menos orientados a contrastar la H2. Supongamos que tenemos las siguientes oraciones:

- (14) a. La escena – de la película – fue en Chicago.
b. La escena – de la obra – fue en Chicago.

Chomsky observa que es un procedimiento muy creativo del inglés (y de español, como vemos acá) formar una única oración a partir de esas dos:

- (15) La escena – de la película y de la obra – fue en Chicago

Si quisiéramos caracterizar este proceso gramatical, la conjunción, dice Chomsky, la manera más simple de hacerlo sería a través de una regla como la siguiente:

- (16) Si S_1 y S_2 son oraciones gramaticales, y S_1 difiere de S_2 sólo en que X aparece en S_1 e Y aparece en S_2 , y X e Y son constituyentes del mismo tipo en S_1 y S_2 respectivamente, entonces S_3 es también una oración gramatical, donde S_3 es el resultado de reemplazar X por $X+Y+Y$ en S_1 .

El problema con (16), señala Chomsky, es que no hay manera de plasmar esta regla en una gramática de estructura de frase, por lo que una teoría basada en dicho tipo de algoritmo pierde una generalización importante.

(El segundo ejemplo que da Chomsky en contra de las gramáticas de estructura de frase es un clásico de la gramática moderna: el análisis de los auxiliares del inglés. No lo desarrollamos acá (i) por su extensión y (ii) porque implicaría introducir supuestos de morfofonémica. Su lectura, sin embargo, se recomienda para todo interesado en la teoría de la sintaxis).

El tercer ejemplo de Chomsky es el análisis de las pasivas.

- (17) a. Juan pateó la pelota.
b. La pelota fue pateada por Juan.

(Como estoy seguro que todos conocen el análisis transformacional de la pasiva, voy a ir directo al grano). Observen el siguiente patrón:

- (18) a. [SUJETO Juan] admira la sinceridad.
b. *[SUJETO La sinceridad] admira a Juan.

Es un hecho evidente que un nominal como *la sinceridad* no puede ser sujeto de un verbo como *admirar*, por lo que cabría postular ciertas restricciones sobre los posibles sujetos de *admirar*. Sean cuales sean estas restricciones, las formas pasivas deben incluir restricciones sobre sus sujetos que funcionen de modo opuesto:

- (19) a. [SUJETO La sinceridad] es admirada por Juan.
b. *[SUJETO Juan] es admirado por la sinceridad.

Una manera de explicar los patrones de (18) y (19) sería estableciendo dos conjuntos distintos de restricciones, uno para las formas activas u otro para las pasivas. Pero esto no haría más que duplicar de forma redundante cierto tipo de información semántica. Otra posibilidad sería derivar la forma pasiva de la activa a partir de una regla:

- (20) Si S_1 es una oración gramatical de la forma
 $SD_1 - FLEX - V - SD_2$
Entonces la secuencia correspondiente de la forma
 $SD_2 - ser - FLEX - PASIV - V - por - SD_1$
También es gramatical.

La regla de (20) resuelve en forma elegante y parsimoniosa el problema de tener que duplicar las restricciones sobre los sujetos del verbo *admirar*.

Como saben, Chomsky denominó *transformaciones* a las reglas del tipo de (16) y (20). Una transformación T opera sobre una determinada secuencia con una determinada estructura de constituyentes y la convierte en una nueva secuencia con una nueva y derivada estructura de constituyentes. Podemos, a partir de esto, formular una nueva hipótesis:

- (H3) La sintaxis de las lenguas naturales consiste en una gramática de reglas de estructura de frase enriquecida con reglas transformacionales.

El problema ahora es elegir entre H2 y H3. Noten que el problema no es que H2 no pueda de ningún modo generar los patrones ejemplificados a partir de (12), sino que

la forma en que puede hacerlo es recurriendo a reglas particulares y *ad hoc* (noten el parecido a los supuestos auxiliares que toman las teorías científicas cuando aparecen datos que van contra sus predicciones). La hipótesis H3, en cambio, captura en forma simple las generalizaciones que se desprenden de los datos y permite establecer nuevas predicciones a ser testeadas.

4. Sobre los objetivos de la teoría lingüística

Si bien el contenido de los capítulos II a V de Chomsky (1957) fue muy importante para el desarrollo de la lingüística, el carácter rupturista de la obra aparece recién en el capítulo VI.

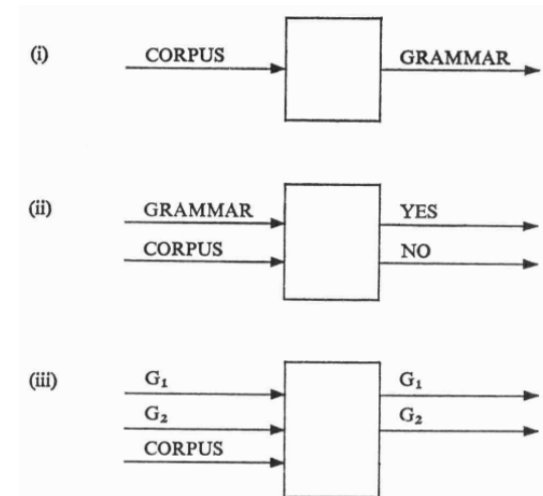
Our fundamental concern throughout this discussion of linguistic structure is the problem of justification of grammars. ***A grammar of the language L is essentially a theory of L.*** Any scientific theory is based on a finite number of observations, and it seeks to relate the observed phenomena and to predict new phenomena by constructing general laws in terms of hypothetical constructs such as (in physics, for example) "mass" and "electron." Similarly, a grammar of English is based on a finite corpus of utterances (observations), and it will contain certain grammatical rules (laws) stated in terms of the particular phonemes, phrases, etc., of English (hypothetical constructs). These rules express structural relations among the sentences of the corpus and the indefinite number of sentences generated by the grammar beyond the corpus (predictions). ***Our problem is to develop and clarify the criteria for selecting the correct grammar for each language, that is, the correct theory of this language.*** (Chomsky 1957).

Chomsky distingue dos criterios: (i) el criterio de *adecuación externa* (i.e., las oraciones generadas por la gramática de L deben ser aceptables para el hablante nativo de L) y (ii) la *condición de generalidad* (i.e., la gramática de L debe ser construida a partir de una teoría específica de la estructura lingüística en la cual términos como *fonema* y *frase* sean definidos independientemente de cualquier lengua)⁴.

Existe, por tanto, un componente general de la labor del gramático (i.e., la teoría) cuyas predicciones deberán ser válidas con respecto a múltiples dominios empíricos (i.e., las lenguas particulares).

Notice that neither the general theory nor the particular grammars are fixed for all time, in this view. Progress and revision may come from the discovery of new facts about particular languages, or from purely theoretical insights about organization of linguistic data — that is, new models for linguistic structure. (Chomsky 1957).

En otras palabras, nuevas hipótesis pueden provenir tanto de datos nuevos de las lenguas particulares como de la búsqueda de teorías más parsimoniosas. Ahora bien, hay varias maneras en que una teoría puede relacionarse con sus datos.



- (i) Refiere a la teoría de la estructura lingüística como un *procedimiento de descubrimiento de gramáticas*.
- (ii) Refiere a la teoría de la estructura lingüística como un *procedimiento de decisión de gramáticas*.
- (iii) Refiere a la teoría de la estructura lingüística como un *procedimiento de evaluación de gramáticas* (su output es la gramática más adecuada entre G₁ y G₂ para un determinado corpus).

Una teoría como (i) se basa principalmente en la *inducción* como mecanismo de postulación de gramáticas: se toman los datos y, a partir de un conjunto de procedimientos, se postula una gramática.

⁴ El proyecto que se retoma aquí es el de, por ejemplo, la *Gramática General y Razonada de Port Royal*: dicha gramática se proponía enunciar principios generales para todas las lenguas que ofrecieran explicaciones "profundas" de sus propiedades.

Una teoría como (iii), en cambio, utiliza el método *hipotético-deductivo*: la teoría ofrece un determinado conjunto de primitivos con el solo fin de comparar la adecuación de *n* gramáticas hipotéticas al corpus.

En el abandono de la inducción se encuentra el aspecto más revolucionario de *Estructuras Sintácticas*. Como señala Newmeyer:

Chomsky's "true argument" dealt with nothing less than the very nature of linguistic theory. No issue is as important as the relevant criteria for theory evaluation, since a radically revised evaluation procedure entails a theory with a radically revised ontological basis. To abandon a procedure based on *induction*, and to adopt one based on *generality*, is to break from past practice at its most fundamental point; it requires one to cease thinking about a grammar as an operationally-derived synthesis of a corpus, and to begin regarding it as a theory of a language. (Newmeyer 1985).

Referencias

Todas las obras citadas se encuentran en la carpeta de la Unidad 2 del Campus Virtual, excepto por Hockett (1971).

HOCKETT, C. (1971). *Curso de Lingüística Moderna*. Buenos Aires: Eudeba.