

# COMPLEJIDAD Y NIVELES DE GRAMATICALIDAD EN LA ENSEÑANZA/APRENDIZAJE DE LENGUAS: MODELOS FORMALES

M. Dolores Jiménez López<sup>1</sup>

Adrià Torrens Urrutia

*Universitat Rovira i Virgili, Tarragona*

## RESUMEN

*Presentamos tres modelos para calcular la complejidad lingüística y los niveles de gramaticalidad que pueden ser útiles en la enseñanza/aprendizaje de lenguas: una herramienta de inferencia gramatical para calcular la complejidad lingüística; un formalismo que genera la gramática de lenguas desconocidas a partir de la gramática de lenguas ya descritas; y una gramática que analiza cualquier estructura independientemente de su nivel de gramaticalidad. Estos modelos pueden simplificar diversas tareas en la enseñanza de idiomas como los procesos de evaluación, el análisis contrastivo, el estudio de la interlengua o el análisis de errores.*

Palabras clave: complejidad lingüística, gramaticalidad, enseñanza/aprendizaje de lenguas

## ABSTRACT

*We present three formal models to calculate linguistic complexity and levels of grammaticality that may be useful in language teaching/learning: a grammatical inference tool that calculates linguistic complexity; a formalism that generates the grammar of unknown languages from the grammar of languages already described; and a grammar that allows syntactic analysis of any structure irrespective of its level of grammaticality. These models can simplify different tasks in language teaching such as assessment processes, contrastive language analysis, interlanguage study or error analysis.*

Keywords: linguistic complexity, grammaticality, second language teaching/learning

---

<sup>1</sup> Este trabajo se ha realizado en el marco del proyecto FFI2015-69978-P (MINECO/FEDER, UE) financiado por el Ministerio de Economía y Competitividad y el Fondo Europeo de Desarrollo Regional.

## 1. INTRODUCCIÓN

En este trabajo, proponemos utilizar en la enseñanza/aprendizaje de segundas lenguas modelos formales definidos para calcular la complejidad lingüística y los niveles de gramaticalidad. Estos modelos pueden ayudar a automatizar distintas tareas en la enseñanza de idiomas como el análisis contrastivo de lenguas, el estudio de la interlengua, el análisis de errores o los procesos de evaluación.

El análisis contrastivo pone de relieve las diferencias y semejanzas entre dos lenguas, ayuda a identificar los problemas con los que se encontrará el aprendiz y puede mejorar los métodos en la enseñanza de lenguas (Fisiak, 1981). Disponer de herramientas que realicen automáticamente la comparación simplifica la tarea y posibilita la obtención de resultados más exactos.

El análisis de errores ofrece datos útiles al profesor de lenguas porque “orienta al docente sobre dónde debe hacer hincapié en su labor profesional” (Bustos, 1998, p. 11). Identificar los errores y sistematizarlos es una ardua tarea que puede simplificarse si utilizamos modelos formales que automaticen el proceso.

Interesante en la enseñanza/aprendizaje de lenguas es el concepto de interlengua, definido como “el sistema lingüístico cognitivo, específico, de una lengua no materna” (Baralo, 2004, p. 369). Cada aprendiz posee, en cada fase de aprendizaje, un sistema específico que no es ni el de su lengua materna ni el de la L2. Disponer de herramientas que caractericen de manera precisa y automática la interlengua puede proporcionar al docente información de gran utilidad.

La evaluación del aprendiz es una tarea muy delicada que persigue objetivos diferentes: determinar el nivel del estudiante para incluirlo en una clase; valorar la competencia adquirida; analizar las necesidades/dificultades para aplicar estrategias de corrección. Utilizar instrumentos que automáticamente realicen la evaluación garantiza la objetividad y agiliza la tarea.

## 2. NIVELES DE GRAMATICALIDAD

Aunque distintas corrientes lingüísticas han considerado la posibilidad de hablar de gradualidad en las cuestiones gramaticales, no es fácil encontrar teorías en las que realmente se considere la gramática como un objeto no-discreto. Los modelos sintácticos estándar presuponen la gramaticalidad como una noción

categoría. En contraste con esas ideas, encontramos formalismos como las gramáticas de propiedades que permiten hablar de gramaticalidad difusa.

### 2.1 Gramáticas de propiedades

Las gramáticas de propiedades se presentan como un modelo, basado en el uso de restricciones, para dar cuenta del conocimiento que un hablante tiene de una lengua con independencia del valor de gramaticalidad de los inputs generados (Blache y Balfourier, 2001). *Restricciones, descripciones y construcciones* son las nociones principales sobre las que se define este formalismo.

Las *propiedades* expresan relaciones entre las categorías de una lengua. En la información sintáctica existen regularidades que pueden describirse de forma independiente. Se trata de especificar esas regularidades usando restricciones o propiedades.

Estas gramáticas persiguen *describir* los fenómenos locales del lenguaje. Se diferencian de otras teorías basadas en restricciones en que no generan estructuras. Su objetivo es representar, a través de restricciones, la información lingüística.

Las gramáticas de propiedades se sirven de la noción de *construcción*, definida como un conjunto de categorías que se relacionan mediante una serie de propiedades.

La caracterización de cualquier entrada consiste en una lista de propiedades satisfechas y violadas. Esta caracterización es posible tanto para inputs gramaticales como agramaticales. Para cada construcción, el formalismo proporciona un valor de gramaticalidad obtenido a partir de la importancia de las restricciones violadas y del valor acumulativo de las violaciones y las propiedades satisfechas.

Este modelo es apropiado para analizar cualquier estructura con independencia de su nivel de gramaticalidad y para determinar el nivel de gramaticalidad de las construcciones. Estas características hacen de estas gramáticas un formalismo útil en la enseñanza/aprendizaje de lenguas, ya que ofrecen al docente herramientas para evaluar a los aprendices:

1. Tolerar distintos niveles de gramaticalidad convierte a estas gramáticas en una herramienta capaz de analizar las construcciones (probablemente no totalmente gramaticales) de los aprendices de segundas lenguas.

2. Proporcionar el valor de gramaticalidad de las producciones permite determinar automáticamente el nivel de adquisición de la lengua meta y, por tanto, evaluar al aprendiz.
3. Explicitar las restricciones satisfechas y violadas en las estructuras producidas hace posible identificar de forma precisa y automática los errores que comete el estudiante. Esto ayuda al profesor a determinar en qué punto del aprendizaje se encuentra el aprendiz.

### 3. COMPLEJIDAD LINGÜÍSTICA

En los últimos años, el interés por la complejidad lingüística ha llevado a cuestionar el dogma de la equi-complejidad, abordando el estudio de la complejidad desde diferentes puntos de vista (Dahl, 2004). Una de las dicotomías más repetidas es la que distingue la complejidad absoluta de la complejidad relativa. La complejidad absoluta se define como una propiedad objetiva del sistema. La complejidad relativa tiene en cuenta a los usuarios del lenguaje y se identifica con la dificultad de procesamiento, aprendizaje o adquisición.

La mayor parte de los trabajos adoptan una perspectiva absoluta y son escasos los que abordan la complejidad desde el punto de vista del usuario. La mayoría de los estudios que se ocupan de la complejidad relativa optan por el aprendizaje de segundas lenguas. Y quienes se centran en el aprendizaje de L2 prestan atención a la lengua meta, sin considerar el peso que la lengua materna del aprendiz tiene en el coste/dificultad para aprender la L2.

A continuación, presentamos un modelo de inferencia gramatical que permite considerar el proceso de adquisición de la L1 para la evaluación de la complejidad lingüística. Introducimos, después, un modelo que permite tener en cuenta la lengua materna de los hablantes a la hora de determinar el coste/dificultad del proceso de aprendizaje de la L2.

#### 3.1 *Inferencia Gramatical*

En todo problema de inferencia gramatical (IG) se distinguen dos elementos (De la Higuera, 2010): un *profesor* que proporciona datos y un *aprendiz* que, a partir de esos datos, tiene que ser capaz de identificar el lenguaje.

En general, los estudios de IG reducen el problema de aprendizaje a la adquisición de la sintaxis y omiten cualquier información semántica. A diferencia de esto, Angluin y Becerra (2010) proponen un modelo que considera la semántica durante el proceso de aprendizaje. En este modelo, profesor y aprendiz interactúan en una serie de situaciones produciendo frases para denotar un objeto. Para evaluar el modelo se utilizó una tarea que consiste en aprender un lenguaje a partir de pares frases-dibujos que denotan figuras geométricas con distintas propiedades. En la evaluación se consideraron diferentes lenguas (inglés, alemán, griego, hebreo, húngaro, mandarín, ruso, español, sueco y turco). En los experimentos realizados se contabilizó el número de interacciones necesarias para alcanzar un buen nivel de actuación en las distintas lenguas. Los resultados distinguieron dos grupos: griego y ruso que necesitan al menos 3400 frases del profesor; y el resto de lenguas que necesitan como máximo 1.000 frases. El modelo muestra que, partiendo de un algoritmo único, no todas las lenguas necesitan el mismo número de interacciones para alcanzar el mismo nivel de rendimiento.

Este modelo permite medir la complejidad lingüística en términos relativos. De hecho, el algoritmo único podría ser equivalente a la capacidad innata que permite a los humanos adquirir un lenguaje. El aprendiz (algoritmo) no tiene conocimientos previos sobre el lenguaje y podría representar al niño que tiene que adquirir una lengua exponiéndose a ella. Contar el número necesario de interacciones para lograr un buen nivel de rendimiento equivale a calcular el coste/dificultad para adquirir un idioma. Demostrar que con el mismo algoritmo no todas las lenguas requieren el mismo número de interacciones puede interpretarse como evidencia para defender que la dificultad para adquirir diferentes idiomas no es la misma y que, por lo tanto, las lenguas difieren en complejidad relativa.

Modelos como este pueden tener consecuencias interesantes en la enseñanza de lenguas. Quienes se dedican a la enseñanza de segundas lenguas se han interesado por el concepto de complejidad como una manera de calcular el nivel de adquisición de la L2. Conocer las diferencias de complejidad de las lenguas puede ayudar a plantear métodos de enseñanza/aprendizaje diferentes para cada lengua. Si las lenguas no son iguales, es probable que el mismo método de enseñanza no sea adecuado para todas ellas.

### 3.2 *Womb Grammars*

Las *womb grammars* (Dahl y Miralles, 2012) son un formalismo que permite generar automáticamente la gramática de lenguas desconocidas (lenguas meta)

a partir de la gramática de lenguas ya descritas (lenguas fuente). Para conseguir inferir la sintaxis de lengua meta, el sistema requiere varios elementos:

1. el lexicón de la lengua meta;
2. un corpus representativo de frases en la lengua meta;
3. la sintaxis de la lengua fuente definida con gramáticas de propiedades;
4. un *parser* basado en restricciones que genere una lista de propiedades satisfechas y violadas.

Se distinguen dos tipos de *womb grammars*, dependiendo de que funcionen con un *parser híbrido* o con un *parser universal*.

En el funcionamiento con *parser híbrido* (Figura 1), se aplica la gramática de propiedades que define la sintaxis de la lengua fuente al léxico y al corpus de la lengua meta. La aplicación del *parser* proporciona una lista de propiedades satisfechas y violadas. Las propiedades satisfechas son restricciones compartidas por ambas lenguas. Las restricciones violadas son propiedades sintácticas que diferencian a la lengua meta de la lengua fuente. Las violaciones que se producen sistemáticamente se entienden como estructuras correctas en la lengua meta. El módulo de reparación de la gramática transforma la gramática de la lengua fuente en la gramática de la lengua meta convirtiendo las restricciones recursivamente violadas en restricciones que deben ser satisfechas. Así, se infiere la gramática de la lengua meta.

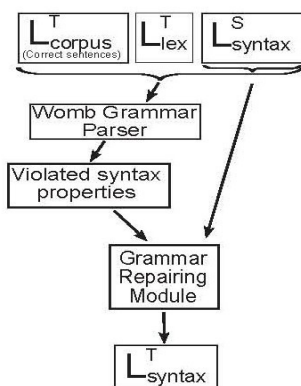


Figura 1. *Womb Grammar: Parser Híbrido* (extraída de Dahl y Miralles, 2012)

Una *womb grammar* con *parser universal* (Figura 2) podría inferir la sintaxis de cualquier lengua a partir de una *gramática universal*. En este caso, el *parser* opera con:

1. el lexicón de la lengua meta;
2. un corpus representativo de frases en la lengua meta;
3. una gramática universal que incluye todas las restricciones que puedan existir en cualquier lengua;
4. un *parser* basado en restricciones que genera una lista de propiedades satisfechas y violadas.

En este caso, aplicado el proceso de *parsing*, lo que obtenemos es una lista de propiedades violadas y una lista de propiedades satisfechas. Para generar la gramática de nuestra lengua, bastará con descartar todas las propiedades violadas e incluir en la gramática todas las propiedades satisfechas.

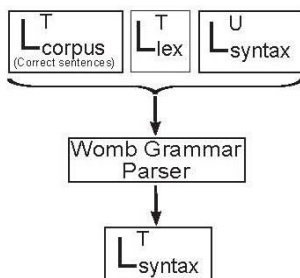


Figura 2. *Womb Grammar: Parser Universal* (extraída de Dahl y Miralles, 2012)

En procesos de enseñanza/aprendizaje de lenguas, el modelo híbrido de las *womb grammars* nos permite medir la distancia entre la lengua materna del aprendiz y la lengua meta, proporcionando una lista de estructuras/reglas/categorías presentes en una lengua y ausentes en la otra. Este funcionamiento permite dos cosas:

1. Calcular la complejidad relativa de una lengua teniendo en cuenta a adultos que adquieren una L2: un número pequeño de diferencias entre las dos lenguas debería coincidir con una complejidad relativa

pequeña. Este modelo calcula la complejidad basándose no exclusivamente en la L2, sino sobre la base de un análisis contrastivo de los sistemas de la L1 y la L2.

2. Realizar automáticamente un análisis contrastivo de las lenguas, proporcionando una lista de similitudes y diferencias. Este análisis puede mejorar la metodología de enseñanza, ya que permite al docente predecir los errores y diseñar materiales didácticos específicos.

El modelo universal de las *womb grammars*, por su parte, puede resultar de gran utilidad a la hora de determinar la interlengua del aprendiz. Lo que hace este formalismo es proporcionar, automáticamente, el conjunto de reglas que utiliza el aprendiz en cada fase del proceso de aprendizaje. Nos ofrece, por tanto, una descripción sistemática que permite determinar la especificidad de la gramática que está adquiriendo el alumno.

#### 4. CONCLUSIONES

El cálculo de la complejidad lingüística a través de modelos formales puede proporcionar al profesional de la enseñanza de lenguas información valiosa respecto a los distintos niveles de dificultad de las lenguas e incluso señalar dónde radican las principales diferencias entre la lengua materna del alumno y la lengua meta. Las diferencias de complejidad podrían justificar el diseño de metodologías de enseñanza específicas para cada lengua.

Determinar el nivel de adquisición del estudiante de segundas lenguas es tarea ineludible para quienes se dedican a la enseñanza de lenguas extranjeras. Disponer de herramientas que, de forma automática, evalúen la competencia lingüística del aprendiz y analicen los errores producidos puede facilitar mucho la labor del docente.

Diseñar modelos que mejoren y faciliten los procesos de enseñanza/aprendizaje de lenguas tiene un impacto social y económico importante. En general, se reconoce la necesidad de fortalecer la enseñanza de segundas lenguas para aumentar la competitividad y el crecimiento económico de los estados. Disponer de modelos computacionales que ayuden al docente o que puedan ser utilizados de forma autónoma por el aprendiz para evaluar sus progresos contribuye, sin duda, a ese fortalecimiento de la enseñanza de lenguas.

## REFERENCIAS BIBLIOGRÁFICAS

- ANGLUIN, D., BECERRA-BONACHE, L. (2010). *A Model of Semantics and Corrections in Language Learning*. Technical Report, Yale University.
- BARALO, M. (2004). La interlengua del hablante no nativo. En J. Sánchez Lobato, y J. Santos Gargallo, I. (Eds.), *Vademécum para la formación de profesores: enseñar español como segunda lengua (L2)/lengua extranjera (LE)* (pp. 369-387). Madrid: SGEL.
- BLACHE, P., BALFOURIER, J.M. (2001). Property grammars: A flexible constraint-based approach to parsing. *Proceedings of Seventh International Workshop on Parsing Technologies*. Beijing: Tsinghua University Press.
- BUSTOS, J.M. (1998). Análisis de errores. *Cuadernos de Filología Hispánica*, 16, 11-40.
- DAHL, Ö. (2004). *The Growth and Maintenance of Linguistic Complexity*. Amsterdam: John Benjamins.
- DAHL, V., MIRALLES, E. (2012). Womb grammars: Constraint solving for grammar induction. En J. Sneyers y T. Frühwirth (Eds.), *Proceedings of the 9th Workshop on Constraint Handling Rules* (pp. 32-40). Roskilde: Roskilde University.
- DE LA HIGUERA, C. (2010). *Grammatical Inference: Learning Automata and Grammars*. Cambridge: Cambridge University Press.
- FISIAK, J. (Ed.) (1981). *Contrastive Linguistics and the Language Teacher*. Oxford: Pergamon.