

Plenary

DATA-DRIVEN LEARNING Y EL APRENDIZAJE DE IDIOMAS

Dr. Pascual Pérez Paredes

Universidad de Murcia
pascualf@um.es

Pascual Pérez-Paredes es Profesor Titular de la Universidad de Murcia y, desde 1994, Traductor Oficial-Intérprete Jurado para la lengua inglesa por el Ministerio de Asuntos Exteriores. Sus principales intereses investigadores son la recopilación y la utilización de corpus lingüísticos y las tecnologías de la información y la comunicación en el proceso de aprendizaje de una lengua extranjera. Dirige en la actualidad el proyecto SACODEYL – MINERVA (PROMOTION of OPEN and DISTANCE LEARNING, INFORMATION and COMMUNICATION TECHNOLOGIES IN THE FIELD of EDUCATION) y es miembro del Grupo de Investigación Lingüística Aplicada Computacional, Enseñanza de Lenguas y Lexicografía (LACELL).

Data-Driven Learning (DDL) como propuesta metodológica para el aprendizaje de idiomas cumple sus primeros quince años de existencia. En este trabajo se repasan algunas de sus premisas fundamentales y se abordan cuestiones de interés para los profesionales de la investigación aplicada y la enseñanza de lenguas extranjeras.

DDL: FUNDAMENTOS

Hace 15 años de la publicación del artículo de Tim Johns (1991) titulado "Should you be persuaded - two samples of data-driven learning materials", trabajo considerado como el precursor de un tipo de aprendizaje de idiomas que utiliza el lenguaje utilizado por los participantes en el contexto comunicativo como punto de partida para el análisis y la (auto-) instrucción en la lengua extranjera. Esta propuesta metodológica fue bautizada por el propio Johns como *Data Driven Learning*, término de difícil traducción al castellano y que no voy a traducir en este foro dada la idiosincrasia docente de sus lectores, por un lado, y mi deseo de no contribuir a incrementar con un nuevo sintagma nominal a la ya de por sí difícil jungla terminológica de la lingüística aplicada. Por ello, mantendré el término en el original inglés y me referiré a él con frecuencia mediante el acrónimo DDL, utilizado ampliamente en la literatura especializada.

El enfoque que aquí nos ocupa hunde sus raíces en el debate lingüístico que durante la segunda mitad del siglo XX cuestionó la dicotomía que

planteaban el estructuralismo y la lingüística chomskyana, con sus respectivos énfasis en los ejes sintagmático-paradigmático y el binomio competencia-actuación. Este hecho por sí solo ya justifica la atención que atrajo ulteriormente la lingüística de índole comunicativo-situacional, que culminó con el trabajo de Sinclair (1991) en el que se asentaron las bases de la metodología de la lingüística del corpus y de lo que en el futuro desembocaría en la formulación de las nuevas unidades de análisis gramatical (Tognini-Bognelli, 2001).

DDL parte de una visión funcional del lenguaje humano, más centrada en el análisis de los elementos que inciden en la comunicación en cualquiera de sus manifestaciones y, por lo tanto, más preocupada por analizar los factores socioculturales y situacionales que afectan y determinan la comunicación, en general, y su plasmación en textos, en particular. Este nuevo horizonte en el campo de la lingüística reconoce, por añadidura, la dimensión cognitiva del aprendiz como un elemento de primer orden en el proceso de adquisición y aprendizaje de una lengua. Esto es, se incorporan al debate

lingüístico avances que se produjeron en la psicología cognitiva desde la década de los 60 del siglo XX como por ejemplo la idea de aprendizaje por esquemas (Barlow, 1996).

De especial interés para el tema que nos ocupa es la ruptura con la preeminencia de las unidades de análisis gramatical y sintáctico propias tanto del estructuralismo como de la gramática generativa. Si bien éstas son de gran interés y utilidad para el lingüista en el desarrollo de su trabajo, así como para el propio aprendizaje de lenguas en su proceso de aprendizaje, no es menos cierto que las categorías *palabra*, *sintagma* u *oración*, por citar las principales, determinan una percepción discente de la realidad lingüístico comunicativa que no es suficiente para el desarrollo de la nueva competencia comunicativa que propugna DDL. Es necesario, sin embargo, que apuntemos que muchas de las aplicaciones de la lingüística del corpus a la enseñanza de idiomas no participan necesariamente de este componente cognitivo. Coxhead (2002) y Pérez-Paredes (2003), a modo de ejemplo de lo anterior, utilizan los métodos de la lingüística del corpus para la selección de léxico relevante para la enseñanza del inglés para fines específicos. En trabajos como éstos, la lingüística del corpus ofrece una metodología para el tratamiento de los datos lingüísticos, generalmente basada en la frecuencia y en la concurrencia de elementos discretos, que carece de un marco pedagógico propio y diferenciado. Este uso lo describe Murison-Bowie (1996) como la versión reducida¹ de la metodología propia de la lingüística del corpus con fines descriptivos.

DDL, por el contrario va más allá, utiliza los datos lingüísticos de manera característica para conducir al estudiante hacia un dominio de experiencias de aprendizaje que tienen como común denominador una concepción *profesional* del discente. Me atrevo a llamar así a esta concepción ya que, usando las metáforas de Bernardini (2000), el estudiante pasa a convertirse en un investigador y en un viajero;

"The main tenet of this methodology is that it provides opportunities for the learner to develop inductive discovery strategies, or to learn how to learn, along with opportunities

to increase her/his own competence [...] learners can exercise their inferring and generalising skills at the level of subtlety allowed by their current capacity, without making it necessary for teachers to manipulate input in order to simplify it" (Bernardini, 2000: 135-6).

En informática, se habla de software que es *menu-driven* cuando el usuario lo maneja mediante la selección de opciones en un menú en vez de teclear las órdenes o comandos. En realidad, la mayoría del software que usamos a diario es de este tipo, ya que la interfaz gráfica del mismo facilita la utilización de este tipo de interacción con el ordenador. Utilizando la analogía informática, cabría preguntarse qué mueve el proceso de aprendizaje de un idioma extranjero: ¿es el profesor, será la metodología, el aprendiz quizá?

DDL sitúa los datos lingüísticos en una posición privilegiada ya que son precisamente el componente metodológico que determina las experiencias de aprendizaje de los estudiantes de una lengua extranjera. Este componente no es tan visible en los métodos de enseñanza de lenguas que aparecen con más frecuencia en la literatura especializada. Es más, es interesante observar que algunos de los métodos más extendidos como CLT (*Communicative Language Teaching*) o CBT (*Content Based Teaching*) no utilizan la palabra *aprendizaje* en su denominación. Si los datos lingüísticos son capaces de motivar *per se* la génesis de un nuevo enfoque metodológico para el aprendizaje de lenguas, cabría preguntarse si los datos que actualmente maneja la lingüística del corpus son, de algún modo, diferentes de los datos que manejaban hace cuarenta años, por citar a un eminente autor de la época, L.G. Alexander.

Las nuevas unidades de análisis referidas anteriormente son más sensibles al texto en el que aparecen, por un lado, y al contexto comunicativo que motiva a estos últimos, por otro. Esto determina que desde la teoría lingüística se proponga una nueva forma de contemplar el fenómeno lingüístico, una nueva forma de entender cómo se conforma la realidad textual. En este sentido, la lingüística

del corpus y, por extensión, DDL nos permiten ciertamente otear un nuevo horizonte en lo que a *input* se refiere. Los trabajos desarrollados en los últimos años se han centrado en tres grandes áreas. Por un lado, encontramos esfuerzos encaminados hacia la identificación de patrones gramaticales que permitan establecer una nueva dimensión de análisis lingüístico en la que léxico y gramática no se hallen en terrenos estancos. Hunston y Francis (2000), sin duda con la experiencia previa² desarrollada en el proyecto COBUILD, proponen la idea de patrón como la unidad de análisis y de estudio idónea para contextualizar el léxico de un idioma en una interfaz más integrada en la idea tradicional de gramática.

En segundo lugar, nos encontramos con un plano de análisis de mayor abstracción. El trabajo de Hoey (2005) sitúa el léxico en fronteras que superan las tradicionales de la lexicografía, otorgando a éste protagonismo frente a la visión de la gramática de un idioma como una estructura apriorística que determina la elección del léxico a utilizar por un hablante. El trabajo referido parte de la premisa contraria. Las palabras, ni siquiera los lemas, son las unidades que determinan fenómenos como la colocación, y no al revés. Dicho de otro modo, aquello que es frecuente y natural en un idioma parte de la predilección semántica que el léxico, nuestro léxico, nuestro propio idiolecto, tiene en determinados contextos semánticos o prosodias y en determinadas situaciones comunicativas.

Por último, quisiera destacar un conjunto de esfuerzos de naturaleza aplicada que están haciendo posible la implementación de DDL. Scott y Tribble (2006), haciendo uso de software y medidas estadísticas *ad hoc*, han diseñado escenarios docentes y discentes en los que aplicar métodos propios de DDL, trascendiendo de una manera muy efectiva la utilización de líneas de concordancia y profundizando en la relación entre textos y la noción de género. Por desgracia, el software que utilizan a tal efecto no se distribuye gratuitamente. Renouf (2002), por el contrario, ha venido trabajando en los últimos años en el perfeccionamiento de un interesantísimo sistema³ abierto que utiliza la información

disponible en Internet como corpus potencial para el aprendizaje y la enseñanza de idiomas:

"A simple search on WebCorp generates a list of examples called concordances from a selection of websites. These examples can then be sorted into a list where your search term is highlighted in red. The concordances are listed in alphabetical order according to the word that occurs to the left or right of your word. This gives you an instant picture of any grammatical patterns or collocation. WebCorp also provides a link to websites the words were taken from⁴".

Durante años, el sitio web de Tim Johns ha servido como catalizador de experiencias pedagógicas vinculadas a DDL⁵. A modo de ejemplo, el tipo de material bautizado como *remedial grammar* asumía lo siguiente sobre los aprendices de inglés y francés como lengua extranjera en la Universidad de Birmingham:

"As engineers, the British students should be accustomed to the idea of research into data as a fundamental method of learning about the world, while many of the French students are already attending EISU classes on grammar, vocabulary etc. based on monolingual concordance output and a research-based methodology⁶".

DDL: APORTACIONES Y RETOS

DDL adopta de forma integradora esta nueva visión y propone una pedagogía del descubrimiento (Bernardini, 2000) en la que el aprendiz ha de asumir un papel activo y central. La pedagogía del descubrimiento se asienta sobre la premisa de que el aprendiz está en la mejor de las posiciones posibles para reestructurar su propio conocimiento del idioma que está aprendiendo. A tal efecto, las líneas de concordancia han sido tradicionalmente el vehículo elegido por los docentes interesados en DDL. El siguiente ejemplo⁷ ilustra las diferencias de uso de determinantes en inglés. Para la resolución de esta tarea, el aprendiz deberá atender al significado de los dos sintagmas nominales propuestos en los diferentes contextos de uso:

What is the difference between the Victorian society (2) and Victorian society (5)?

Look also at citation 3. Here are some more citations showing this pattern. What do they have in common? Can you invent some more to describe the society of today?

1. heyite, a sworn enemy of the acquisitive society, with a strong trade union association and
2. cument the rise and rise of the affluent society - provided, of course, one takes account of
3. ustice as a key ingredient of the better society which it seeks to bring about. "If we are t
4. it seeing the formation of the classless society that he so eagerly desires.
5. e situation could change if the consumer society were to collapse. Frantic consumerism is on
6. emain two final requirements of the good society. It must have an effective working economy.
7. . Ever wilder swipes against the lawless society will promise better times for the core Cons
8. Marxism. His most famous work, *The Open Society and Its Enemies*, published in 1945, was reg
9. s mind. "I was born after the permissive society was inaugurated," he says. "I was 16 when t
10. neficiaries of the something-for-nothing society, put something back into society in the int
1. a speech of the president of _____ Royal Society, delivered in 1978: "Ominously, voices have
2. ent change in the taste of _____ English society in a way which is probably without parallel
3. ion of science to industry. _____ modern society may collapse unless the atmosphere of scien
4. nkings "the architect of _____ permissive society" when launching her autobiography. An inqui
5. urs "sociological" explanations of _____ society, perhaps because when most editors and seni
6. e widespread complacency in _____ German society: the current generation of young people ha
7. oy the fabric of _____ modern industrial society, they will prevent an enemy from being stud
8. Keel, "What's holding up _____ cashless society?" asked *New Scientist* on its front cover, s
9. d with the functioning of _____ consumer society, there are no safe havens. Cities are the p
10. In June 1888, a meeting of _____ Fabian Society concerned with the lot of female labour hea
11. time when large tracts of _____ Italian society were deeply corrupt. Earlier, Di Pietro int
12. tion of how much "_____ post-industrial society hates children. Career-wise and status-wise

Figura 1: Líneas de concordancia para el análisis de determinantes en inglés.

Bernardini (2000:139) sostiene que estas experiencias se caracterizan por promover (1) una concepción del aprendizaje de idiomas inductiva e independiente; (2) una concepción probabilística del lenguaje y (3) una concepción de la pedagogía de idiomas que soluciona la dicotomía entre la atención a la forma y al significado. Aunque los tres aspectos mencionados son de interés para el docente de idiomas, es el primero de ellos el que presenta mayores y más visibles implicaciones en las actividades de aula. Tomando como referente el doble papel del aprendiz de lenguas bien como investigador (el discente resuelve tareas de distinta complejidad propuestas por el docente), bien como viajero (el discente aborda por sí mismo aquellas tareas que mejor cubren sus intereses o necesidades comunicativas), es evidente que la experiencia docente se organiza en torno a una nueva taxonomía de competencias. Así, el discente deberá ser capaz de:

- Desarrollar y poner en práctica estrategias de búsqueda de patrones lingüísticos en un corpus: utilización de nodos de diferente complejidad, interpretación de las líneas de concordancia (lectura vertical acumulativa), etc.
- Desarrollar y poner en práctica estrategias para refinar las búsquedas en un corpus lingüístico: determinar la cantidad de output necesaria para llegar a conclusiones de relevancia.

- Desarrollar y poner en práctica estrategias para clasificar los resultados de las búsquedas en un corpus lingüístico.
- Desarrollar y poner en práctica estrategias para determinar la relevancia y la adecuación de los resultados de las búsquedas en un corpus lingüístico a un posible patrón o patrones lingüísticos.
- Desarrollar y poner en práctica estrategias para determinar la influencia del contexto en los resultados de las búsquedas en un corpus lingüístico a un posible patrón o patrones lingüísticos.
- Desarrollar y poner en práctica estrategias para determinar la influencia del contexto en los resultados de las búsquedas en un corpus lingüístico a un posible patrón o patrones lingüísticos.
- Desarrollar y poner en práctica estrategias para determinar la prosodia semántica en los resultados de las búsquedas en un corpus lingüístico a un posible patrón o patrones lingüísticos.

Todo ello, no cabe duda, supone reorganizar el aprendizaje de los estudiante y orientarlo hacia actividades de naturaleza inductiva para desarrollar la concienciación lingüística y fortalecer la autonomía de los aprendices. Los retos de DDL son, pues, muchos y de difícil resolución en el corto plazo.

Algunos de estos retos están relacionados con los niveles a los que se dirige la propuesta. Hasta la fecha, desconozco la existencia de propuestas basadas en DDL para aprendices de niveles A1 o A2, centrándose la mayoría de las propuestas actuales en los niveles universitarios (Bernardini, 2004, Pérez-Paredes, 2004a; Pérez-Paredes, 2004b). DDL, por otro lado, requiere de una infraestructura mínima que posibilite la utilización de corpus en línea o almacenados en un servidor; esto es, precisa de una política de distribución en red que no todos los centros de enseñanza pueden ofrecer. En tercer lugar, como rama joven de la lingüística, la lingüística del corpus todavía no ha alcanzado un estatus de reconocimiento amplio. Así, no es inusual encontrar profesorado que nunca ha recibido formación en la materia, lo cual impide la implementación de DDL. Cuarto, por motivos que desbordan los fines y la orientación de este trabajo, los proveedores de contenidos educativos hasta la fecha utilizan, si acaso, los corpus en su versión débil, esto es, como fuente de información para la confección de sus materiales de enseñanza. Está por aparecer el texto o recurso que apueste decididamente por este enfoque inductivo. En quinto lugar, destacaría, paradójicamente, la inercia de la metodología comunicativa como un obstáculo par la difusión de DDL. El debate actual sobre la idoneidad de la enseñanza de lenguas

centrada en la forma o en el contenido es heredero de la aversión del enfoque comunicativo al estudio de la forma lingüística. Ellis (2005), tras una revisión bibliográfica, concluye que ambos enfoques son necesarios en la pedagogía de idiomas actual, por lo que se hace necesario integrar ambos en las propuestas metodológicas de aula.

El último reto que apuntaré es el de mayor calado y, posiblemente, el que ocupará el debate profesional y científico durante los próximos años. El debate al que me refiero está relacionado con la propia metodología de trabajo del aprendiz en DDL y con la idoneidad de los procesos cognitivos que lo articulan. En este debate las aplicaciones pedagógicas de la lingüística del corpus se juegan su difusión y aplicación en la pedagogía de lenguas. Si en el futuro queremos que DDL sea más visible y que esté más presente en el aula, no sería aceptable que estas aplicaciones fueran patrimonio exclusivo bien de un reducido grupo de estudiantes con extraordinarias capacidades analíticas, bien de estudiantes de último año de carrera. En el desarrollo de este debate precisaremos descripciones que, proviniendo del campo de la psicología aplicada al aprendizaje, nos aporten datos sobre la idoneidad de los procesos cognitivos asociados a estas prácticas de aprendizaje inductivo.

Notas

¹ *Weak claim*

² Por ejemplo, Collins COBUILD Grammar Patterns 2: Nouns and Adjectives (1998), (Eds.) Gill Francis, Susan Hunston y Elizabeth Manning. HarperCollins Publishers.

³ <http://www.webcorp.org.uk/>

⁴ <http://www.macmillandictionary.com/MED-Magazine/may2005/30-Corpora-Tips.htm>

⁵ <http://www.eisu.bham.ac.uk/johnstf/timconc.htm>

⁶ <http://www.eisu.bham.ac.uk/johnstf/lingua.htm>

⁷ <http://www.eisu.bham.ac.uk/johnstf/lingua.htm>

REFERENCIAS

- Barlow, M. 1996. "Corpora for Theory and Practice". *International Journal of Corpus Linguistics* 1/1: 1-37.
- Bernardini, S. 2000. *Competence, capacity, corpora. A study in corpus-aided language learning*. Bologna: CLUEB.
- Bernardini, S. 2004. "Corpus-aided language pedagogy for translator education". *Translation in Undergraduate Degree Programmes. Teaching and Learning by Doing Corpus Analysis*. Ed. K. Malmkjaer. Amsterdam: John Benjamins, 97-112.
- Ellis, R. 2005. "Principles of instructed language learning". *System* 33: 209-224.
- Coxhead, A. 2002. "The Academic Word List: A Corpus-based Word List for Academic Purposes". *Teaching and Learning by Doing Corpus Analysis*. Eds. B. Kettemann y G. Marko. Amsterdam y Nueva York: Rodopi. 73-89.
- Granger, S., y S. Petch-Tyson, eds. 2002. *Extending the scope of corpus-based research*. Amsterdam: Rodopi.
- Hoey, M. 2005. *Lexical Priming*. London: Routledge.
- Hunston, S., y G. Francis. 2000. *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
- Johns, T. 1991. "Should you be persuaded - two samples of data-driven learning materials". *Classroom Concordancing*, monográfico de *ELR Journal* 4: 1-13.
- Kettemann, B. y G. Marko, eds. 2002. *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi.
- Malmkjaer, K., ed. 2004. *Translation in Undergraduate Degree Programmes*. Amsterdam: John Benjamins.
- Murison-Bowie, S. 1996. "Linguistic corpora and language teaching". *Annual Review of Applied Linguistics* 16: 182-199.
- Pérez-Paredes, P. 2003. "Small corpora as assisting tools in the teaching of English news language: A preliminary tokens-based examination of Michael Swan's Practical English Usage news language wordlist". *ESP WORLD* 3/6. [Documento disponible en http://www.esp-world.info/articles_6/pascual.htm]
- Pérez-Paredes, P. 2004a. "Network: Learner oral corpora and network - based language teaching: Scope and foundations". *Corpus, concordance and collocation*. Ed. J. McH. Sinclair. Amsterdam: John Benjamins. 249-268.
- Pérez-Paredes, P. 2004b. "Lightnings of Intuition: A Protocol for University Students' Oral Discourse Appreciation". *BELLS* 12. [Documento disponible en <http://www.publicacions.ub.es/revistes/bells12/PDF/art12.pdf>]
- Renouf, A. 2002. "WebCorp: providing a renewable data source for corpus linguists". *Extending the scope of corpus-based research*. Eds. S. Granger y S. Petch-Tyson. Nueva York: Rodopi. 39-58.
- Scott, M., y C. Tribble. 2006. *Textual Patterns: Key words and corpus analysis in language education*. John Benjamins.
- Sinclair, J. 1991. *Corpus, concordance and collocation*. Oxford University Press.
- Sinclair, J., ed. 2004. *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins.
- Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. John Benjamins.