

Algunos aspectos teóricos y metodológicos de la investigación del aprendizaje léxico y gramatical del español como L2⁵²

Javier Aguado-Orea y Marta Baralo

Universidad Antonio de Nebrija

Resumen

Este trabajo incluye dos secciones bien diferenciadas para dar una visión general del campo de la investigación aplicada a los procesos de aprendizaje en L2. La primera sección trata de dar cuenta de los principales desafíos metodológicos que plantea el estudio de los procesos cognitivos que intervienen en el aprendizaje léxico y gramatical del español como segunda lengua, ~~en~~ aportando algunos ejemplos del contexto escolar, aunque dando por sentado que el aprendizaje de una lengua va más allá de esos aprendizajes. En este marco se trata el problema de la representatividad de los datos lingüísticos, la evaluación del conocimiento y las distintas formas de analizar los datos. En la segunda parte, se abordan directamente una serie de posibles explicaciones acerca de las propiedades de estos procesos. En concreto, se revisan dos posiciones recientes basadas en el enfoque cognitivo del aprendizaje de las lenguas: el aprendizaje implícito y los modelos conexionistas de aprendizaje.

Palabras clave: metodología de investigación, lingüística aplicada, L2, aprendizaje de lenguas, procesos de aprendizaje, representatividad, modelos conexionistas.

Abstract: Cognitive processes that take part in the acquisition of Spanish as a second language at school: new methodologies and new perspectives This paper is based on two well-differentiated sections intended to provide a full picture of the research filed applied to L2 learning processes. The first section is aimed at giving account of the most important methodological challenges faced by the study of the cognitive processes that take part in the acquisition of Spanish as a second language at school. Issues regarding representativeness of linguistic data, comprehension testing and different ways to perform analyses of data. The second section directly covers a series of possible explanations regarding the features of these processes. In particular, two recent proposals based on the cognitive approach to language learning are reviewed: implicit learning and connectionist models of learning.

Keywords: research methods, applied linguistics, L2, language learning, learning processes, representativity, connectionist models.

Revista de Educación, 343 xxx

⁵² Como en otros casos, la paginación reproduce la de la edición original. Este artículo debe citarse como: Aguado-Orea, J. y Baralo, M. (2007). Aspectos teóricos y metodológicos de la investigación sobre el aprendizaje léxico y gramatical del español como L2. *Revista de Educación*, 343, 113-132

Introducción

Este artículo tiene como objetivo llamar la atención sobre algunas alternativas a un modo de concebir el procesamiento lingüístico, así como proponer reflexiones acerca de la propia metodología aplicada en su estudio científico. Dentro del procesamiento lingüístico pueden estudiarse distintos procesos cognitivos, que incluyen diferentes formas de aprendizaje (por ejemplo, memorístico o basado en reglas) y diferentes cadenas de actividades lingüísticas (por ejemplo, las de la comprensión de un texto oral). Tal y como indica Skehan (1998, p. 2), «el estudio de la adquisición de segundas lenguas (L2) ha recibido excesiva atención por parte de la lingüística y la sociolingüística y en cambio, hasta hace poco, apenas ha recibido interés por parte de la psicología cognitiva contemporánea»⁵³. De ser cierta, esta situación sería cuando menos bastante llamativa, ya que el estudio del aprendizaje de cualquier función cognitiva debería parecer campo legítimo de la psicología.

Esta situación podría deberse a que es prácticamente imposible hablar de procesos cognitivos de aprendizaje de una lengua sin haber adoptado previamente un posicionamiento teórico determinado. Haciendo un esfuerzo de simplificación considerable, se podrían contraponer dos visiones distintas del aprendizaje. Una visión que hereda la corriente de pensamiento más *racionalista*, según la cual la labor principal del aprendiz consiste en rellenar de palabras un espacio mental preparado de forma innata para contener una serie de reglas específicas de combinación de elementos léxicos. Un buen ejemplo de esta perspectiva es el libro de divulgación de Pinker (1999), que lleva un título que difícilmente podría ser más ilustrativo de la idea: *Palabras y Reglas: los ingredientes del lenguaje*. En un artículo de título idéntico, publicado en la revista *Lingua*, Pinker (1998, p. 240) concluye que «la evidencia, por tanto, apoya la hipótesis de que el diseño del lenguaje humano comprende dos mecanismos mentales: la memoria, encargada de los signos arbitrarios que subyacen a las palabras, y la computación simbólica, encargada del uso infinito que subyace a los medios finitos de la gramática». El artículo de Pinker contiene críticas frontales a un enfoque que es heredero de una tradición de vocación más *empirista* y que se resiste a asumir que las reglas simbólicas que usamos para describir las propiedades gramaticales de las lenguas tengan alguna realidad psicológica para los hablantes. Según esta posición contraria a la concepción basada en reglas, es el propio uso de la lengua el que configura las propiedades estructurales, y no al contrario. En concreto, Bybee y MacClelland (2005, p. 382) indican que «la experiencia de los usuarios con la lengua va conformando las representaciones cognitivas, que se construyen mediante la aplicación de principios generales de la cognición humana a los elementos lingüísticos». En otras palabras, no se excluye en este enfoque que las reglas gramaticales sean la mejor forma de describir las propiedades de la lengua; lo que se rechaza es el valor cognitivo que tienen estas reglas gramaticales para explicar la competencia lingüística de los hablantes.

⁽⁵³⁾ Todas las citas de textos originales en inglés se han traducido directamente por los autores de este artículo.

También hay que tener en cuenta que, para la posición más racionalista, el proceso de aprendizaje queda prácticamente reducido a la ejecución de un mecanismo de llenado de la memoria. Los principios que gobiernan la lengua no se aprenden; a lo sumo, se encargan de guiar el proceso de *adquisición*. Pero no olvidemos que el objetivo de este artículo está en valorar los procesos cognitivos de aprendizaje. Por tanto, el simple hecho de hablar de procesos incluye ya un cierto posicionamiento hacia una de las dos perspectivas: la más empirista. Centrar la atención en los procesos implica ir más allá de la concepción del conocimiento basado en contenidos simbólicos, o en mecanismos basados en reglas o principios predeterminados de forma innata. Implica, en cambio, tratar de averiguar cuál es la naturaleza de parte del conocimiento que hace posible la producción y comprensión de construcciones lingüísticas en distintos momentos de la trayectoria vital de una persona. En concreto, nos referimos en este trabajo a los procesos cognitivos que conllevan al aprendizaje de la gramática y el léxico sustentan el sistema formal que permite la codificación y la interpretación de los significados que se comunican

Una segunda cuestión que es preciso tener en cuenta en este debate es que la lingüística aplicada no cuenta con un enfoque metodológico que haya sido el resultado del desarrollo de la propia actuación científica, sino que se trata de una disciplina que tiende a adoptar metodologías desarrolladas inicialmente en otras disciplinas. Esto hace que, por un lado, conviva en la actualidad una considerable variedad de técnicas de investigación. Sin embargo, por otro lado, el hecho de que estas técnicas se hayan validado en disciplinas con objetivos no siempre similares plantea algunas dificultades, que se mencionarán a continuación. Es cierto que siguen siendo tres las principales herramientas de estudio: la entrevista, el cuestionario y el corpus (López Morales, 1994), y que, dada la importancia de estos tres instrumentos de recogida de datos, nos detendremos muy brevemente más adelante en ellos para tratar de explicar sus implicaciones para el aula (epígrafe 1.4). También se mencionarán brevemente otras técnicas menos habituales en el campo (epígrafe 1.5). Sin embargo, nuestro objetivo aquí no es el de recopilar técnicas de análisis (ya existen buenos manuales al respecto, como Brown, 1988; Nunan, 1992; o Seliger y Shohamy, 1989, por citar tres ejemplos), sino el de llamar la atención sobre algunos problemas fundamentales a los que se tienen que enfrentar todas las técnicas de investigación. Así, nos gustaría que en el quehacer investigador del lingüista aplicado se reflexionara sobre las ideas de *representatividad* (epígrafe 1.1.), de *evaluación de la comprensión lingüística* (1.2.) y del *análisis de datos* (1.3).

Metodología de investigación aplicada al estudio del aprendizaje de lenguas en contextos escolares

Como se ha mencionado más arriba, la pretensión de este artículo es la de indicar una serie de aspectos que conviene tener en cuenta antes de aplicar una técnica de investigación específica. Pero, antes de seguir, conviene definir brevemente algunos conceptos que se van a utilizar aquí, ya que se asumirá en todo momento que son entendidos por el lector. Cuando hablamos de *datos*, nos referimos a aquellos segmentos incompletos de la realidad que tienen que ser explicados; por *teorías* aludimos a las explicaciones acerca del modo en que ocurren las cosas normalmente; y por *hipótesis* entendemos aquellas explicaciones provisionales (que aún no hemos sometido a prueba) acerca del modo en que ocurren los datos.

La representatividad

El científico se ve abocado a satisfacer dos objetivos: dar con un fenómeno empírico relevante y satisfacer la representatividad de este fenómeno. Para detectar un *fenómeno empírico*, tendrá que hacer visible algún aspecto de la realidad que pueda ponerse en relación con una hipótesis determinada, de tal forma que la apoye o la refute. Incluso cuando sólo esté tratando de describir la realidad *a posteriori*, deberá concentrar su interés en determinados datos que vayan constituyéndose en fenómenos relevantes. Cuanto más claro y robusto sea el *fenómeno* encontrado, más claramente quedará apoyada o refutada la hipótesis de partida y/o más satisfactoria será la descripción de la realidad.

Por ejemplo, imaginemos que un profesor de secundaria quiere comprobar si una actitud negativa de los alumnos inmigrantes de su aula hacia la cultura española hace más difícil la adquisición de la lengua. A primera vista, dar con este fenómeno empírico (la relación entre actitudes negativas y aprendizaje) se antoja una tarea difícil. Olvidémonos por un momento de las dificultades que se encontrarán para evaluar las actitudes de sus alumnos. Fijémonos, en cambio, en la complejidad de dos ideas importantes de esta hipótesis: cultura y lengua. Para encontrar un fenómeno empírico interesante, el investigador tendrá que alcanzar primero una definición operativa de la idea de cultura española, pero también deberá indicar qué aspectos de la lengua son interesantes. El éxito, por tanto, no dependerá sólo de su capacidad para evaluar las actitudes (es decir, de aplicar la técnica de evaluación de actitudes más adecuada), sino que su habilidad para definir los aspectos esenciales y a la vez operativos de la cultura y la lengua será determinante en el éxito de la investigación. Ahora bien, aunque este profesor de secundaria cuente ya con definiciones satisfactorias, esto no garantizará que pueda abarcar todos los aspectos culturales y lingüísticos que quepan en esas definiciones. Dentro de la cultura cabrán elementos gastronómicos, musicales, artísticos, etc., mientras que dentro de la lengua se podrán tener en cuenta propiedades fonológicas, sintácticas, pragmáticas, etc. Imaginemos que el profesor decide centrar su estudio en la actitud de sus alumnos inmigrantes hacia tres o cuatro platos típicos de la gastronomía española, por un lado, y en el

reconocimiento de vocabulario de alimentos por otro lado. Si encuentra una relación significativa, es decir, observa que los alumnos con una actitud más negativa hacia esos tres o cuatro platos típicos reconocen menos palabras de alimentos, habrá conseguido, en términos estrictos, detectar un fenómeno empírico interesante. Sin embargo, se trataría de una relación más bien anecdótica para la hipótesis que se planteaba inicialmente (posible relación entre actitud hacia la cultura y adquisición de la lengua). Lógicamente, ni la gastronomía es el rasgo principal de una cultura ni la cantidad de vocabulario específico acerca de la comida lo es de una lengua, por interesantes que puedan ser ambos aspectos en sí mismos; es decir, no entendidos como elementos que representan la cultura y lengua españolas. En definitiva, el profesor tendría que elaborar un constructo válido, que haya surgido directamente de las hipótesis de partida, o simplemente tratar de plantear una hipótesis menos ambiciosa. En concreto, podría limitarse a tratar de encontrar relaciones entre la actitud hacia la gastronomía de un país y el aprendizaje del vocabulario de alimentos de la lengua de ese país.

Es precisamente en esta relación entre la amplitud de la hipótesis de partida y la validez del fenómeno donde cobra especial interés la idea de *representatividad*. Acabamos de ver que la hipótesis puede ser más o menos amplia, y que, cuanto menos amplia sea, más validez tendrá el fenómeno empírico encontrado. Pero, lógicamente, también es cierto que, cuanto menos ambiciosa sea la hipótesis, menor repercusión teórica tendrá. También hemos pasado demasiado deprisa sobre el hecho de que se estén haciendo inferencias acerca de, al menos, dos factores fundamentales: la *población de aprendices* de una lengua específica y las *propiedades de un constructo* determinado. En el ejemplo anterior, se asume que el profesor de secundaria había comprobado la relación entre actitudes y vocabulario con una muestra determinada de aprendices de español. Lógicamente, si esta muestra hubiera sido de dos o tres alumnos inmigrantes, sería más difícil asumir que la misma relación entre actitud y aprendizaje se da en cualquier otro alumno inmigrante de secundaria que seleccionemos al azar (o sin, ir más lejos, de un alumno del aula contigua). Si la muestra es de diez personas, y encontramos una relación clara entre actitudes y vocabulario en todas ellas, la probabilidad de que una undécima persona también mostrara esa relación empieza a ser muy elevada. Así, la pregunta que suele hacerse el investigador que se enfrenta por primera vez a un estudio de lingüística aplicada es: *¿qué tamaño ha de tener mi muestra para que represente a toda la población?* Lamentablemente, no hay un número mágico (sorprende comprobar que en más de un manual se considera que una muestra de más de treinta personas ya deba considerarse grande, tal y como ocurre en Carrasco y Calderero Hernández (2000, p. 40). Es más, el mejor modo de responder a esa pregunta suele consistir en formular otra pregunta por respuesta: *¿qué propiedades de la población quiere el investigador que represente su muestra?* Cuantas más propiedades de la población nos interesen, más grande (a la vez que variada) habrá de ser la muestra de participantes. Por ejemplo, si queremos saber si el efecto de las actitudes ocurre igualmente (o no) con chicos y chicas, nuestra muestra deberá contener un número suficiente de alumnos de ambos sexos. Pero seguramente también queramos incluir alumnos de distintas edades, orígenes geográficos, etc.; estas propiedades de la población que queremos que represente nuestra muestra son las que hacen que el

tamaño crezca (existe un buen manual de técnicas de muestreo en el capítulo cuarto de López Morales, 1994). Aun así, no podremos estar seguros de si nuestra muestra era capaz de representar la población hasta después de haber llevado a cabo el estudio y hacer una inferencia estadística determinada. Es más, nunca podremos estar *totalmente* seguros de ello. El análisis de datos estadístico sólo nos indicará qué probabilidad de equivocarnos asumimos al decir que la muestra representa a la población.

Pero hemos mencionado más arriba que la representatividad no sólo era una cuestión de muestras y poblaciones de personas, sino de realidades empíricas: de constructos científicos. Es seguramente éste un buen momento para aclarar a qué hace referencia la idea de *constructo*. Brown (1988, p. 8) compara esta idea con la de *variable* del siguiente modo: mientras que una variable es simplemente lo que podemos observar o cuantificar, el constructo es la auténtica característica que tratamos de evaluar mediante el uso de las variables; en general, el constructo será normalmente la propiedad cognitiva de los seres humanos que estamos tratando de estudiar. Del mismo modo que la muestra aumentará con la cantidad de propiedades de la población que queramos (o debamos) representar, los datos empíricos que vayamos a recoger también aumentarán con la complejidad del constructo teórico que queramos someter a prueba o deseemos explorar. Volviendo al ejemplo anterior, no es lo mismo representar todas las propiedades de la cultura española que tratar de representar sólo su gastronomía. Pero, aun así, si elegimos tres o cuatro platos típicos, tal y como se ha indicado en el ejemplo anterior, ¿estaremos representando todas las variedades regionales de la gastronomía española? Evidentemente, no. El mismo problema existe si pensamos en el conocimiento de nombres de alimentos que queremos muestrear. Podríamos acudir a cualquier diccionario de español y seguramente nos asombraríamos de la cantidad de palabras que habría que tener en cuenta.

En suma, el investigador en lingüística aplicada necesita herramientas que le permitan detectar fenómenos empíricos de una realidad compleja, no siempre fácil de definir (los conceptos de cultura, actitudes y lengua se han usado aquí como ejemplos). Por tanto, ha de hacer el esfuerzo inicial por reducir la complejidad del constructo lingüístico. Una buena forma de validar las herramientas de investigación consiste en evaluar la capacidad de sus datos para representar las propiedades centrales de esos conceptos.

Evaluación de la comprensión

Prácticamente, ningún tipo de competencia lingüística puede representarse adecuadamente sin tener en cuenta que el hablante ha de comprender la lengua. A pesar de ser algo obvio, no es común observar estudios que evalúen la comprensión de un modo satisfactorio. Es más, hay una cierta tendencia a asumir que el hablante que usa una estructura lingüística determinada ya la comprende. Esta inferencia podría ser cierta, pero es importante

recordar la idea de representatividad que se ha discutido en el epígrafe anterior. Podríamos pensar que la competencia lingüística para llevar a cabo una operación gramatical determinada es un equivalente a la población de aprendices en los ejemplos anteriores. Es decir, para estar seguros del grado de competencia lingüística de un hablante, necesitaríamos una muestra de emisiones suficientemente rica y variada para que represente, por ejemplo, las propiedades gramaticales centrales (siempre teniendo presente que la competencia lingüística es mucho más que el conocimiento y uso de las propiedades gramaticales).

Por ejemplo, si queremos saber si un hablante ya es totalmente competente en la comprensión del subjuntivo, no podríamos conformarnos con una muestra de habla en la que aparezcan dos o tres verbos conjugados adecuadamente en subjuntivo. En cambio, necesitaríamos un conjunto de emisiones que permitan asegurar con una probabilidad de error «convincentemente» baja que las propiedades de comprensión del subjuntivo se han muestreado adecuadamente. Por ejemplo, podría ocurrir que un aprendiz de español use la expresión indicada en (1) y en (2), pero no sea aún capaz de usar el subjuntivo con otros verbos (es decir, no sea productivo en el uso del subjuntivo), como indican los ejemplos en (3) y (4).

- (1) *quiero que venga él*
- (2) *quiero que tenga un regalo*
- (3) **quiero que compro yo*
- (4) **puedo venga yo*

Si tenemos además en cuenta que muchas veces no estaremos interesados en evaluar el resultado del aprendizaje, sino que queremos hacer inferencias acerca de las propiedades del proceso de aprendizaje, la representabilidad es aún más complicada. En definitiva, si nos preocupan los procesos de aprendizaje asociados a la comprensión del lenguaje, convendrá estar preparados para rechazar aquellas técnicas de investigación que no vayan a permitir hacer inferencias fiables.

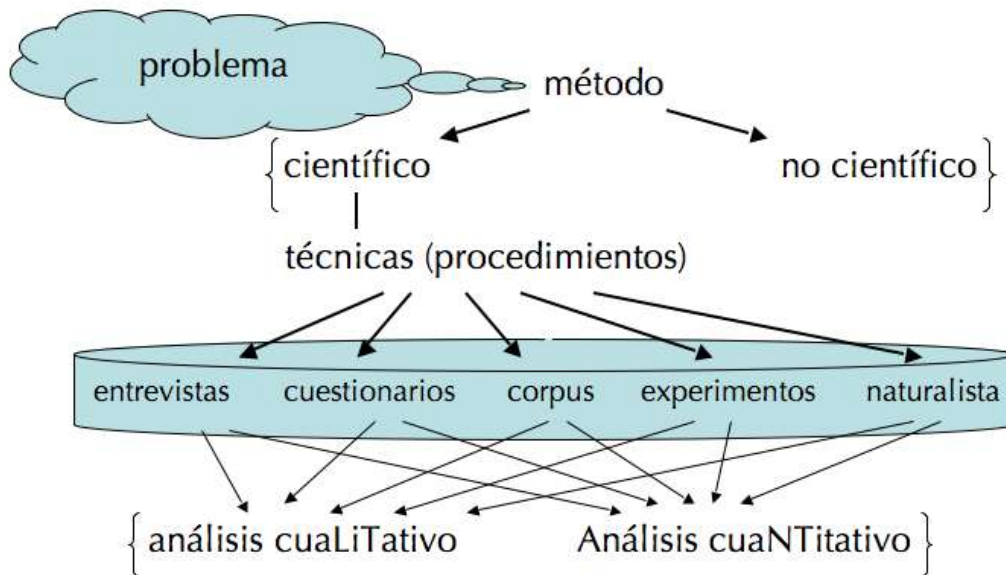
Análisis de datos

Seliger y Shohamy (1998, p. 113) agrupan, con astucia, aquellas «perspectivas de la investigación que intentan describir los fenómenos de segundas lenguas tal y como ocurren de forma natural, como la investigación cualitativa, descriptiva y correlacional y multivariada». Es decir, la observación de los fenómenos desde el punto de vista natural no excluye técnicas que exijan un elevado conocimiento matemático (las técnicas multivariadas). De hecho, para ellos, la diferencia entre la metodología cualitativa y cuantitativa no es tanto una opción del investigador, como un continuo en el que no suele quedar más remedio que situarse. Desde luego, no es algo que tenga que ver simplemente con el modo en que analizamos los datos lingüísticos. Nunan (1992) también opina que la distinción entre ambas metodologías lleva a simplificaciones excesivas, si bien reconoce que es real, dado que está basada en concepciones distintas de la naturaleza del conocimiento que se está analizando. Precisamente, la realidad es que, tal y como señalan Larsen-Freeman y Long (1991), muchos investigadores

tienden a tomar posiciones rígidas con respecto a una u otra metodología, normalmente previas a la propia realización de la investigación. Es decir, un investigador puede tender a posicionarse como cualitativista o cuantitativista, y a mirar con desdén la producción científica del otro bando.

Es cierto que la disposición del investigador a intervenir sobre los datos no siempre es la misma. El investigador puede adoptar una posición más pasiva frente a la realidad lingüística o, por el contrario, puede tratar de reducirla, o incluso modificarla a su antojo. Nuestra visión, en cambio, consiste en que el investigador deberá tener en cuenta si el análisis que va a hacer de los datos va a llevar a cabo operaciones cuantitativas o no (una diferenciación similar puede consultarse en McMillan y Schumacher, 2005). Es muy difícil imaginar operaciones cuantitativas que no impliquen la reducción de la información. Por tanto, en aquellos casos en que la naturaleza de los datos no recomiende su reducción (básicamente, porque cada uno de esos datos es demasiado importante en sí mismo), el análisis tendería a ser cualitativo. Esta visión se resume gráficamente en la Figura I.

FIGURA I. Distinción entre análisis cualitativo y cuantitativo



Cualquier problema relativo a la enseñanza y aprendizaje de las lenguas se puede resolver mediante la utilización del método científico o no. Si se opta por seguir el método científico, este problema deberá ser operacionalizado de tal forma, que pueda analizarse mediante alguna de las técnicas de investigación conocidas (entrevistas, cuestionarios, etc.). Dependiendo del valor de los datos recogidos, éstos se agruparán mediante un análisis cuantitativo o se detallarán mediante un análisis cualitativo. Ahora bien, el análisis cualitativo de los datos lingüísticos entraña una dificultad enorme. Cuando se analiza cualquiera de las oraciones indicadas en los ejemplos (1) a (4) –más arriba– de un modo cualitativo, se corre el riesgo de exagerar el papel de la representación basada en reglas. Por ejemplo, se podría asumir a partir de los ejemplos (1) y (2) que el hablante está aplicando reglas de combinación gramatical de morfemas a partir de las propiedades sintácticas de la

oración. Tal análisis cualitativo podría concluir que estas reglas se han adquirido de modo totalmente satisfactorio. Al hacer esto, resulta difícil sacar a la luz los procesos de aprendizaje que pudieran subyacer a momentos en los que el conocimiento es aún parcial. El análisis cualitativo de los errores indicados en (3) y (4) sí que aportaría información más detallada acerca del proceso de aprendizaje. En definitiva, no sería fácil conciliar el análisis de los errores con el de las expresiones correctas.

En resumen, se han revisado hasta aquí tres aspectos fundamentales para evaluar las técnicas de investigación utilizadas en lingüística aplicada. A continuación se mencionan y describen muy brevemente algunas de las técnicas más habituales.

Técnicas habituales de investigación en lingüística aplicada a la enseñanza de español en el aula

La entrevista viene utilizándose desde los inicios de la investigación en ciencias sociales como una técnica que proporciona mucha información rica de una o pocas personas. Como en todos los demás instrumentos de investigación, el objetivo (qué queremos saber) será el que determine las características de la entrevista: duración, número de personas, tipo de preguntas, etc. Es importante tener en cuenta que normalmente la entrevista se situará en algún punto de dos ejes: estructurada frente a abierta y *a priori* frente a *a posteriori*.

La distinción entre el enfoque *a priori* y *a posteriori* es ya clásica en metodología de la investigación. Hace referencia a la perspectiva que adoptemos acerca del establecimiento de hipótesis: más exploratoria (*a posteriori*) o de comprobación de hipótesis previas (*a priori*). Por ejemplo, podríamos utilizar el uso que hace del lenguaje el alumno para evaluar su competencia en la utilización de la morfología verbal. Por tanto, el contenido de las respuestas a nuestras preguntas sólo sería relevante en tanto que sirve para evaluar el uso de la morfología verbal. El otro eje que tenemos que considerar es también conocido: hasta qué punto el guión de nuestra entrevista es cerrado o, por el contrario, estamos dispuestos a incluir preguntas nuevas en función de las respuestas del entrevistado. Como en el eje anterior, no se trata de dos categorías excluyentes entre sí, sino de un continuo en el que se situará nuestra técnica. Además del conciso manual de Gillham (2000), puede consultarse el capítulo quinto de López Morales (1994) para descubrir más detalles acerca de esta técnica.

Una alternativa menos costosa para obtener mucha información en poco tiempo, si bien normalmente menos rica que en el caso de la entrevista, es el uso de cuestionarios. Dörnyei (2003) ofrece un excelente manual para construir, administrar y analizar encuestas en lingüística aplicada a la enseñanza de lenguas. La intención de este artículo es que el encuestador tenga en cuenta los tres criterios que se han mencionado más arriba: representatividad, posibilidad de estudiar la comprensión y el modo en que se van a analizar los datos. En el caso concreto que nos ocupa, resulta especialmente difícil imaginar situaciones en las que los cuestionarios puedan

sacar a la luz procesos cognitivos de aprendizaje de lenguas, ya que están demasiado ligados a la evaluación de contenidos de conocimiento.

La recogida de corpus (tanto orales como escritos) es también una técnica ya clásica. Nos gustaría llamar aquí la atención especialmente sobre el criterio de representatividad. Aunque parezca una obviedad, es necesario recordar una y otra vez que el corpus no es el resultado de *toda* la competencia lingüística del hablante, sino una muestra de esa competencia. Lo que buscamos es que el corpus represente sobre todo dos cosas: las propiedades de la lengua que nos interesan y la capacidad/habilidad del hablante para producir y/o comprender esas propiedades. Así, suele olvidarse que determinadas palabras, o incluso propiedades gramaticales, aparecen con más frecuencia que otras. Por tanto, no podemos esperar que nuestro corpus represente por igual todas las posibles palabras o propiedades gramaticales del español.

Una posible solución al problema de la representatividad en la recogida y análisis de corpus consiste en tratar de provocar en el hablante el uso de aquellas características lingüísticas que más nos interesen. El capítulo séptimo de Nunan (1992), *Elicitation techniques*, propone una introducción breve a algunas de estas estrategias. En definitiva, antes de recoger el corpus, el investigador tendrá que poner en un extremo de la balanza la naturalidad y espontaneidad del texto que se va a recoger y en el otro, la cantidad de información específica que quiere encontrar.

No hay espacio suficiente aquí para dar cuenta de otras técnicas no menos importantes en el estudio de la enseñanza/aprendizaje de lenguas, ya que se impone la necesidad de presentar técnicas menos habituales, pero más apropiadas para la evaluación de los procesos de aprendizaje. En cualquier caso, conviene mencionar al menos dos manuales orientados directamente al aula. Lasagabaster y Sierra (2004) recopilan una serie de estudios de gran interés acerca de las técnicas de observación en el aula. Otro manual con una temática similar es el de Schachter y Gass (1996).

Técnicas más recientes

Un modo bastante sencillo de evitar la dependencia de la producción lingüística con cuestionarios sencillos consiste en pedirle al aprendiz que realice juicios de gramaticalidad acerca de una serie de oraciones o de cualquier otro tipo de construcción lingüística. La tarea consiste en que el participante diga si le *suena* bien o mal la oración que se le presenta. De esta forma, podemos evaluar el conocimiento gramatical de los aprendices sin necesidad de recoger una muestra de producción lingüística. Mandell (1999) ha encontrado resultados satisfactorios al evaluar la validez lingüística de esta técnica, triangulándola con otras herramientas de análisis. Dekeyser (2005) revisa algunos aspectos interesantes mediante el uso de esta técnica. Montrul (1991, 2001) ha combinado la técnica con el uso de imágenes para obtener información acerca de la interpretación de determinadas escenas mediante verbos causativos. Sin embargo, esta herramienta tiene al menos dos problemas.

Uno de ellos es que los ítems los elabora el propio investigador. Por tanto, es muy difícil averiguar hasta qué punto el conocimiento gramatical es productivo (si bien es cierto que se podrían introducir expresiones nunca oídas, por ejemplo, usando «verbos» nuevos). Otra limitación consiste en que su utilización está bastante vinculada a la indagación del conocimiento gramatical (aunque, con la elaboración de construcciones ingeniosas se podría evaluar también el conocimiento léxico e incluso la segmentación fonética). Por supuesto, la presentación de estos textos puede hacerse de forma oral o por escrito. Lo ideal, en cualquier caso, es comparar los juicios de gramaticalidad de los hablantes de la L2 con los de los hablantes de la L1.

Una alternativa para evaluar las propiedades del conocimiento léxico y gramatical consiste en medir el tiempo que tarda en reaccionar verbalmente el participante ante la presentación de un objeto o una escena. La instrucción que recibe es la de nombrar el objeto que aparece o explicar qué está ocurriendo. De este modo, es posible diferenciar entre palabras y construcciones de gran accesibilidad (aquéllas con tiempo de reacción muy pequeño) y aquéllas a las que es más difícil acceder (tiempo de reacción mayor).

Por último, una técnica de escasa utilización en lingüística aplicada, pero que debería tener unas posibilidades considerables es la simulación de modelos de aprendizaje mediante el ordenador. La idea consiste en someter a prueba hipótesis acerca del funcionamiento de procesos específicos de aprendizaje mediante el ordenador. Los modelos que se emplean para simular estos procesos asumen unos supuestos tanto de conocimiento inicial del lenguaje como de mecanismos de analogía muy básicos. En concreto, los modelos conexionistas (Rumelhart & McClelland, 1986) tratan de explicar el aprendizaje gramatical mediante redes neuronales artificiales, compuestas únicamente por conjuntos de elementos idénticos que tienen la capacidad de establecer nexos entre sí.

Procesos de aprendizaje

El empleo de nuevas metodologías de investigación (lingüística de corpus, análisis computacionales) está generando nuevas posiciones teóricas (por ejemplo, modelos de aprendizaje conexionistas frente a modelos de aprendizaje basados en símbolos) que, a su vez, permiten hacerse nuevas preguntas acerca de los procesos de adquisición de lenguas. En esta sección, no se van a tratar las metodologías de investigación, sino las teorías que han surgido de su utilización.

Procesos comunes y diferencias de aprendizaje de L1 y L2

En líneas generales, se pueden señalar tres diferencias principales entre los procesos de aprendizaje de una primera lengua (L1) y los que tienen lugar durante el aprendizaje de otra lengua distinta (L2) en momentos posteriores del desarrollo. En primer lugar, por obvio que parezca a primera vista, hay que insistir en el hecho de que un aprendiz ya ha adquirido una lengua previa mientras que el otro aún está construyendo su primer sistema de representación lingüística.

Pero, además, el niño normal está inmerso en un *contexto de apoyo social*, de atención y de cuidados que no recibe igualmente el adulto. Tampoco es lógico pensar que el niño que se vea privado de ese contexto de apoyo tendrá enormes dificultades para adquirir su primera lengua, pero no es menos arriesgado pensar que ese sistema de apoyo social (o al menos de cuidados maternos) es irreproducible durante el aprendizaje de la L2.

Una tercera diferencia entre los dos procesos de aprendizaje (el de la L1 y la L2) tiene que ver con el grado de especificidad del lenguaje. Es lógico pensar que el aprendiz de una L1 aún no conoce la lengua de los adultos de su comunidad, pero no es demasiado arriesgado pensar tampoco que desconozca muchos de los aspectos esenciales del mundo que le rodea. Al contrario, el aprendiz de una L2 no sólo ha desarrollado ya un sistema lingüístico inicial, sino que ha adquirido un grado de *conocimiento general del mundo* lo suficientemente elevado como para sobrevivir en una sociedad compleja.

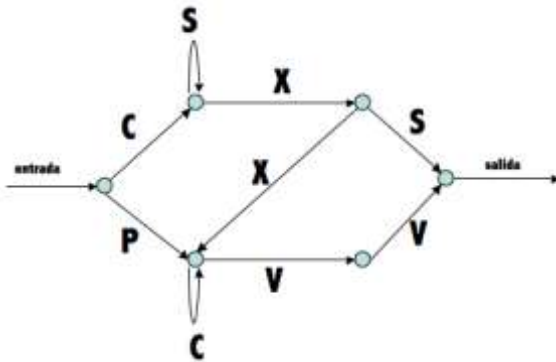
Esta tercera diferencia se puede poner en relación directa con la distinción teórica mencionada al comienzo de este artículo. Hemos visto cómo Bybee y McClelland (2005) defendían la explicación de los procesos de aprendizaje de lenguas a partir de otros procesos cognitivos más básicos. Esta idea no es nueva. De hecho, Piaget ya defendía una posición similar en la primera mitad del siglo pasado. Algunos investigadores actuales (por ejemplo, MacWhinney, 2007) retoman este tipo de posiciones constructivistas y llaman la atención sobre los procesos de aprendizaje que sí podrían compartir mecanismos esenciales en el caso de la L1 y la L2. La asunción de la existencia de una competencia común subyacente al aprendizaje de las competencias lingüísticas en L1 y en las lenguas subsiguientes, que adquiera un individuo a lo largo de su vida, son propuestas que ya se vienen defendiendo, en las dos últimas décadas, en la investigación sobre el aprendizaje de lenguas en diferentes contextos, y que se enmarcan en el concepto de plurilingüismo del Marco de Referencia europeo (epígrafe 1.3).

El papel del aprendizaje implícito

Winer y Reber (1994, p. 117) definen el aprendizaje implícito como la «habilidad humana para derivar información acerca del mundo de un modo inconsciente, no reflexivo». Vivimos en un mundo tecnológicamente complejo, por lo que todos los días nos vemos abocados a aprender a utilizar herramientas para las que no contamos con manuales de formación ni con instrucción explícita. La idea del aprendizaje implícito es que ni siquiera somos conscientes de que estamos aprendiendo a utilizar esas herramientas. En un estudio histórico,

aunque no muy conocido en lingüística aplicada, Reber (1967) trató de comprobar si las personas éramos capaces de aprender gramáticas artificiales sin instrucción explícita. Para ello, diseñó un experimento en el que se pedía a los participantes que memorizaran listas de letras sin sentido aparente. Aunque ellos no lo sabían, el orden en que estaban organizadas estas listas de letras obedecía a unas reglas determinadas, como las que se representan en la Figura III –simplificación propia de Cleeremans et al. (1998) y Winter y Reber (1994).

FIGURA III. Ejemplo de gramática artificial



Emisiones posibles de esta gramática artificial serían *CSXS*, *PCCVV* o *CSSXXCVV*. Sin embargo, emisiones como *VS*, *PCXV* o *CSSXVV* no serían gramaticalmente correctas. Tras la fase de memorización, se pedía a los participantes que dijeran si otras cadenas de letras (la mayoría de ellas no vistas hasta entonces) podían haber sido creadas por las mismas reglas que la que habían memorizado o no. Curiosamente, los participantes aciertan a juzgar si la cadena es correcta o no por encima de lo que se esperaría de ellos si contestaran al azar. Otra serie de experimentos más recientes han utilizado cadenas de sonidos sin sentido para observar fenómenos de aprendizaje similares en bebés de nueve meses (Aslin, Saffran y Newport, 1998).

Cleeremans, Destrebecqz y Boyer (1998) tratan de indagar qué es lo que hace posible que los aprendices se comporten por encima de lo que se esperaría si respondieran al azar en este tipo de estudios. En primer lugar, las personas podrían detectar las reglas con contenido simbólico que originan el experimento. Es decir, siguiendo la estructura representada en la Figura I, los aprendices podrían darse cuenta de que, necesariamente, la cadena de palabras debería empezar por *C* o por *P*, que si empieza por *P* sólo puede ir seguida por una *S*, etc. Esto es lo que ellos denominan el enfoque «abstraccionista» (en la línea de Pinker). En segundo lugar, Cleeremans et al. (1998) proponen como explicación una posible sensibilidad a unidades de ítems que se repiten bastante y que, por tanto, son más fáciles de recordar. Esta sensibilidad sería lo que se suele denominar enfoque basado en el *chunking* (Ellis, 2003). Las personas detectamos unidades pequeñas que aparecen de forma consistente. La gramática artificial de la Figura III produciría, por ejemplo, secuencias en las que *XX* y *VV* son bastante habituales. Sobre estas repeticiones habituales, se podrían detectar cadenas mayores, como *SXX* y *CVV*. Por último, un procesador de la información que detecte la probabilidad de qué letra aparezca en una posición específica de la secuencia, debería tener un gran éxito para decidir si una nueva cadena de caracteres es correcta

o no. Esta sensibilidad estadística es la que tratan de resolver los modelos de aprendizaje basados en redes neuronales, que se describen a continuación.

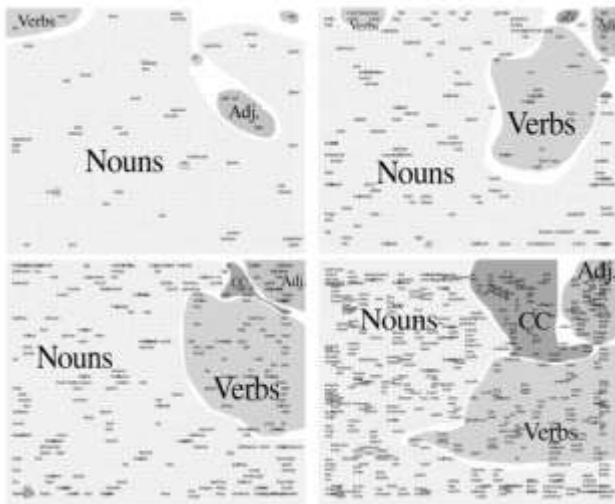
En resumen, los resultados de los estudios basados en la detección inconsciente (o implícita) de patrones de regularidad en ítems producidos por gramáticas artificiales son interesantes precisamente porque se había despojado al lenguaje de muchas de sus propiedades esenciales: sociopragmáticas, semánticas, fonológicas, etc. De hecho, se podría argumentar que no es un estudio sobre el lenguaje (ni siquiera sobre la gramática), sino de la capacidad de los humanos para detectar regularidades en la organización inconscientemente.

Modelos conexionistas

La agenda investigadora de los equipos de investigación en modelos conexionistas no para de crecer. De hecho, va mucho más allá de los estudios basados en el aprendizaje de las lenguas. Aquí sólo hay espacio para presentar algunas propuestas recientes del aprendizaje de segundas lenguas. Un monográfico ya clásico es el de Elman y sus colegas (Elman et al., 1996). Durante las cuatrocientas páginas del libro, los autores tratan de explicar los procesos de cambio cognitivo con los mecanismos de representación mental más sencillos que se puedan imaginar: las redes neuronales artificiales. El objetivo es precisamente eludir cualquier hipótesis que incluya presupuestos fuertes acerca de la especificidad de principios innatos para la adquisición de funciones cognitivas superiores, sin excluir el lenguaje. *La idea que subyace es que, en vez de aprenderse reglas con contenido simbólico, lo que se detectan son regularidades en los datos, a los que el sistema cognitivo es sensible.* Eso sí, para estos autores no hace falta asumir que el sistema cognitivo tenga una estructura compleja. En las propias palabras de Elman et al. (1996, p. 50), los elementos responsables del procesamiento de la información no son más que *nodos* o *unidades* que ponen en relación grupos de neuronas artificiales. Los dos únicos supuestos son que las neuronas pueden tener distintos niveles de activación y que pueden estar enlazadas o no. Ante una serie de *inputs* que tienen unas propiedades específicas, la organización de las neuronas se modifica y la activación global de la red de neuronas se traduce en una respuesta determinada. La «experiencia» de la red con distintos tipos de *input* hace que la organización de la red cambie: nuevos nodos y distinta activación de neuronas. A medida que aumenta el entrenamiento, la plasticidad de la red disminuye.

MacWhinney (1997) ha utilizado este tipo de modelos para simular distintos fenómenos observados en el estudio de la adquisición de L1 y L2. La Figura IV ilustra cuatro momentos de un mapa auto-organizado (un tipo de red).

FIGURA IV. Ejemplo de mapa auto-organizativo (MacWhinney, 2007)



La figura indica la organización del conocimiento tras haber aprendido 50 palabras (recuadro superior-izquierdo), 100 palabras (superior-derecho), 250 (inferior-izquierdo) y 500 (inferior-derecho). Es posible observar que durante los primeros momentos existe mucha plasticidad. Por ejemplo, los verbos se mueven de un lado para otro durante las primeras cien palabras. También es posible observar cómo las palabras de tipo cerrado (por ejemplo, preposiciones) van apareciendo mucho más tarde que las de tipo abierto (nombres y verbos), pero, tras la adquisición de las 500 palabras, ya se ha establecido una zona bastante estable con ellas, y es más importante que la de los adjetivos.

Además de este fenómeno léxico, permite simular propiedades silábicas. Por ejemplo, el entrenamiento en sílabas de una lengua determinada (el japonés) hace que no haya flexibilidad suficiente para producir sílabas substancialmente distintas (del inglés) tras un período de entrenamiento de la red prolongado. Sin embargo, el sistema silábico es más flexible cuando el entrenamiento ha sido menor. Este resultado se ha utilizado para cuestionar la idea de período crítico: no es necesario contar con un programa de maduración innato para perder plasticidad tras una exposición prolongada a una L1, incluso cuando la arquitectura cognitiva básica y de aprendizaje no incluye la representación de reglas con contenido simbólico. Existen muchos otros fenómenos de este tipo que los modelos conexionistas pueden simular sin recurrir a reglas explícitas, pero no hay espacio aquí para revisarlos (remitimos al lector al propio artículo de MacWhinney, 2007).

Conclusiones

Este artículo se abría indicando la confrontación teórica que existe al tratar de explicar los procesos de aprendizaje de lenguas. Por un lado, se indica la existencia de hipótesis de explicación que surgen de una tradición más racionalista, dispuesta a asumir especificidad innata para explicar la adquisición de las lenguas,

pero también para explicar su uso. Por otro lado, se ha comentado el intento de ofrecer explicaciones alternativas desde posiciones que emergen de una tradición más empirista.

Pero este artículo se ha preocupado también por presentar tres cuestiones metodológicas que afectan a la investigación en lingüística aplicada en general, y a la de los procesos de aprendizaje en particular. Se ha señalado la necesidad de que el fenómeno observado gracias a los datos lingüísticos represente adecuadamente el constructo teórico que se quiere someter a prueba o que se pretende describir. También se ha mencionado que una buena técnica de investigación tiene que permitir hacer inferencias acerca de los procesos de comprensión, no sólo de actuación lingüística. Cuando el constructo teórico consiste en la explicación de procesos de aprendizaje, la representatividad es aún más problemática, y exige el empleo de metodologías más ingeniosas que las tradicionales técnicas de entrevistas o cuestionarios. Además, se han discutido brevemente los problemas asociados a la utilización de análisis cualitativos y cuantitativos.

Entre las técnicas recientes de investigación, se ha descrito la utilización de juicios de gramaticalidad. También se han descrito algunos experimentos que permiten vislumbrar las sorprendentes capacidades humanas de detección de regularidades en la lengua. Por último, se han introducido brevemente las técnicas de simulación de modelos conexionistas de aprendizaje basados en redes neuronales artificiales y sus implicaciones acerca del debate de partida.

Por supuesto, aún queda mucho camino por recorrer antes de decidir si la balanza se inclina más hacia los modelos que adoptan reglas de contenido simbólico abstracto para explicar los procesos de aprendizaje de lenguas, o por el contrario se inclina más hacia los modelos que prefieren pensar que esas reglas sólo tienen valor descriptivo y tienen poco que ver con lo que nos pasa a los humanos cuando aprendemos tanto una L1 como una L2. Porque en todo caso aprender lengua es un proceso muy complejo que dura toda la vida, y que por ir mucho más allá de la morfología, la sintaxis o la gramática, requiere mantener el esfuerzo por refinar los modelos y métodos de investigación.

Referencias bibliográficas

- BROWN, J. D. (1988): *Understanding Research in Second Language Learning: A Teacher's Guide to Statistics and Research*. Cambridge, Cambridge University Press.
- CARRASCO, J. B.; CALDERERO HERNÁNDEZ, J. F. (2000): *Aprendo a Investigar en Educación*. Madrid, Rialp.
- CLEEREMANS, A.; DESTREBECQZ, A.; BOYER, M. (1998): «Implicit learning: News from the front», en *Trends in Cognitive Sciences*, 2, pp. 406-416.
- DEKEYSER, R. (2005): «What makes second-language grammar difficult? A review of issues», en *Language Learning*, 55, pp. 1-25.
- ELLIS, N. (2003): «Constructions, Chunking and Connectionism», en C. J. DOUGHTY; M. H. LONG (eds.). *The Handbook of Second Language Acquisition*. Oxford, Blackwell
- ELMAN, J. L.; BATES, E. A.; JOHNSON, M.; KARMILOFF-SMITH, A.; PARISI, D.; PLUNKETT, K. (1996): *RETHINKING*

- INNATENESS: A CONNECTIONIST PERSPECTIVE ON DEVELOPMENT. CAMBRIDGE, M A, MIT Press.
- GILLHAM, B. (2000): *The Research Interview*. London, Continuum.
- KOHONEN, T. (1990): «The Self-Organizing Map», en *Proceedings of the IEEE*, 78, pp. 1464-1480.
- LARSEN-FREEMAN, D.; LONG, M. H. (1991): *An Introduction to Second Language Acquisition Research*. Edinburgh, Longman.
- LASAGABASTER, D.; SIERRA, J. M. (eds.) (2004): *La observación como instrumento para la mejora de la enseñanza-aprendizaje de lenguas*. Barcelona, ICE Universitat de Barcelona-Horsori.
- LÓPEZ MORALES, H. (1994): *Métodos de Investigación Lingüística*. Salamanca, Colegio de España.
- MANDELL, P. B. (1999): «On the Reliability of Grammaticality Judgement Tests in Second Language Acquisition Research», en *Second Language Research*, 15, pp. 73-99.
- MACWHINNEY, B. (2007): «A Unified Model», en N. ELLIS; P. ROBINSON: *Handbook of Cognitive Linguistics and Second Language Acquisition*. Mahwah, NJ, LEA.
- MONTRULL, S. (1999): «Causative errores with unaccusative verbs in L2 Spanish», en *Second Language Research*, 15, pp. 191-219.
- (2001): «Causatives and Transitivity in L2 English», en *Language Learning*, 5,1 pp. 51-106.
- NUNAN, D. (1992): *Research Methods in Language Learning*. Cambridge, Cambridge University Press.
- PINKER, S. (1998): «Words and rules», en *Lingua*, 106, pp. 219-242.
- (1999): *Words and Rules. The Ingredients of Language*. New York, HarperCollins.
- RUMELHART, D. E.; McClelland, J. L. (1986): «On Learning the Past Tenses of English Verbs», en *Volume 2 of Rumelhart, McClelland, and the PDP Research Group*, p. pp. 216-271
- SCHACHTER, J.; GASS, S. (eds.) (1996): *Second Language Classroom Research. Issues and Opportunities*. Mahwah, NJ: LEA.
- SELIGER, H. W.; SHOHAMY, E. (1989): *Second Language Research Methods*. Oxford, Oxford University Press.
- SKEHAN, P. (1998): *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- WINTER, B.; REBER, A. S. (1994): «Implicit Learning and the Acquisition of Natural Languages», en N. ELLIS. (ed.), *Implicit and explicit learning of languages*. London, Academic Press.