

<https://doi.org/10.7203/Normas.3.4678>

APRECIACIONES SOBRE LA CUESTIÓN DE LA NORMA EN EL ANÁLISIS DE LA INTERLENGUA*

CONSIDERATIONS ON THE QUESTION OF NORM IN INTERLANGUAGE ANALYSIS

Anna SÁNCHEZ RUFAT

Universidad de Córdoba

FRANCISCO JIMÉNEZ CALDERÓN

Universidad de Extremadura

RESUMEN:

La norma desempeña un papel fundamental en el ámbito de la adquisición de las segundas lenguas (ASL), tanto en el proceso de aprendizaje-enseñanza que se desarrolla en el aula como en la investigación sobre la lengua de los aprendientes. Una lengua solo puede ser descrita a partir de un referente que se erige como el estado de lengua *meta* para el aprendiente, por lo que determina la selección y el tratamiento de las muestras de lengua aportadas en el aula, y, asimismo, se constituye como modelo con el que se contrasta el *output* lingüístico del alumno. En este trabajo pretendemos determinar –tanto desde una perspectiva teórica como aplicada– los aspectos que deben ser considerados a la hora de seleccionar un modelo lingüístico apto para realizar una categorización de los datos sistemática –a partir de un corpus de aprendientes de español dado.

PALABRAS CLAVE: norma lingüística, estándar lingüístico, español como lengua extranjera, interlengua, lingüística de corpus, lengua del aprendiente.

ABSTRACT:

The norm is a centrally important aspect in the field of Second Language Acquisition (SLA), in both foreign/second language teaching and in learner language research. A language can only be described on the basis of some kind of reference point which arises as the target state of language for the learners to arrive at, so that it determines the selection and the processing of the data provided in the classroom, and likewise, it is regarded as the reference point the learners' output is compared with. The aim of this research is to determine –from a theoretical and an applied perspective– the aspects which must be considered when choosing a variety of language which should serve to categorize data systematically, based on a given learner Spanish corpora.

KEY WORDS: linguistic norm, standard variety, Spanish as a foreign language, interlanguage, corpus linguistics, learner language.

* Este trabajo se enmarca en el proyecto de investigación titulado *Modelos y representaciones metateóricas en la historia de la lingüística* (FFI2012-35802), cuya IP es Carmen Galán Rodríguez.

1. INTRODUCCIÓN

La cuestión de la norma lingüística, además de vertebrar numerosos estudios en torno a los usos producidos por los hablantes nativos de una lengua, resulta clave para enfrentar procedimientos básicos en el ámbito de la adquisición de segundas lenguas (en adelante, ASL). En ambos casos, es imprescindible alcanzar una explicación lo suficientemente sólida del concepto que permita un análisis fiable de los usos correspondientes. Habitualmente, el concepto de norma se vincula al de naturalidad, por un lado, y se halla indisolublemente unido al de corrección, por otro. Mientras que la naturalidad mide la representatividad de los datos en función del conocimiento que el hablante tiene de su lengua (Bosque, 2005: XLVIII), la corrección se refiere, desde una perspectiva muy general, a lo que se produce libre de errores o conforme a unas reglas establecidas.

En base a ello, se suele distinguir entre dos orientaciones que la norma puede tomar: por un lado, se habla de una norma prescriptiva, reguladora, que centra las políticas lingüísticas con el conjunto de pautas que definen los usos de una lengua considerados prestigiosos¹; y, por otro, existe una norma descriptiva, referida a los usos sancionados² como comunes o normales por una comunidad de hablantes (Coseriu, 1967). La primera de estas orientaciones se relaciona normalmente con la idea de corrección, y la segunda, con la de naturalidad. Pese a este aparente enfrentamiento, como reconoce también Fernández (1997: 29), la norma reguladora debe estar atenta a lo que es «normal», natural en la comunidad y adecuar a ello sus prescripciones. Así pues, cualquier aproximación al concepto procuraría una integración de ambas nociones –naturalidad y corrección– que, por descontado, suele resultar enormemente compleja. Se trata, en suma, de la búsqueda de un equilibrio necesario en la constitución de un referente lingüístico.

En dicha búsqueda, el concepto de *error* adquiere especial relevancia, pues afecta a toda desviación involuntaria de la norma establecida en una comunidad de hablantes; en relación con las nociones anteriores, puede referirse tanto a lo que no es correcto como a lo que no es natural. El error, además, es particularmente importante en el terreno de la enseñanza-aprendizaje de lenguas extranjeras (LE). Corder (1967) fue el primero en indicar la importancia de los errores por la información que proporcionaban acerca del proceso de adquisición de la lengua, y se refirió a la necesidad de analizar e

¹ El *Diccionario de la Real Academia Española (DRAE)* lo constata: «4. f. Ling. Conjunto de criterios lingüísticos que regulan el uso considerado correcto. 5. f. Ling. Variante lingüística que se considera preferible por ser más culta». La referencia es pertinente por el papel fundamental que desempeña la Real Academia Española en la prescripción de usos lingüísticos.

² Como es usual en los estudios normativos, en este trabajo se emplea *sancionar* con el sentido de ‘Autorizar o aprobar cualquier acto, uso o costumbre’ (*DRAE*, 2001, 20-09-12).

identificar todas las estructuras producidas por el aprendiente, pues el error podía estar encubierto en una estructura superficialmente bien formada pero inadecuada en el contexto dado. Por primera vez, la lengua del aprendiente era el principal foco de atención de los investigadores, y no ya la lengua materna (L1) y la lengua meta (L2) como se venía haciendo en el marco del *análisis contrastivo* –con el objeto de predecir y evitar las transferencias negativas de la L1 en el proceso de adquisición de las estructuras de la L2. Es en ese momento, con la llegada del *análisis de errores*, cuando se hace necesario fijar en los estudios de ASL una norma o modelo lingüístico que permita describir la interlengua (Selinker, 1972)³ y categorizar los datos de la producción no nativa, es decir, determinar con criterio lo que es error y lo que no lo es; una norma que, asimismo, proporcione la posibilidad de contemplar la enorme variabilidad de realizaciones de la lengua –diatópicas, diafásicas y diastráticas–.

En el camino hacia ese modelo lingüístico de referencia adecuado a las pretensiones de la ASL, en el que se busca describir el proceso de adquisición de la lengua y los factores que afectan a los aprendientes en ese proceso, el enfoque del análisis de errores se terminó desvaneciendo prácticamente del todo en los años ochenta⁴ (Ellis, 1994). Atendiendo al objetivo mencionado, y considerando la interlengua como una entidad con valor propio, los corpus de aprendientes⁵ constituyen en la actualidad una adición muy valiosa a las fuentes de datos de ASL (Guo, 2006: 11), pues, al operar con una gran cantidad de datos, permite la obtención de unos resultados generalizables y representativos. Asimismo, el corpus de aprendientes incorpora la ventaja de que parte de datos procedentes de un uso natural de la lengua, y no obtenidos de análisis experimentales o de la introspección, frecuentes en la investigación en ASL (Granger, 2002).

De esta manera, los estudios basados en un análisis de corpus de aprendientes –a diferencia de los fundamentados en un análisis de errores– ofrecen la posibilidad de examinar errores a gran escala y permiten descubrir, además de las dificultades de los

³ El concepto, también denominado *sistema aproximado* (Nemser, 1971) o *competencia transitoria* (Corder, 1967), alude a la lengua de los aprendientes. De aquí en adelante utilizaremos los términos *interlengua* y *lengua de aprendientes* indistintamente, por ser los más extendidos para referirse al sistema lingüístico no nativo.

⁴ Para una reflexión sobre las debilidades de esta metodología, véase Schachter y Celce-Murcia (1977: 441).

⁵ De la definición de corpus aportada por Sinclair (2004) se desprende que actualmente se concibe esta recopilación de textos en soporte electrónico. A finales de los años ochenta y principios de los noventa, dentro del ya reconocido campo de estudio sobre la ASL, surgen los primeros corpus de aprendientes informatizados en inglés como el ICLE, LLC, CLC, SULEC; en español habría que esperar hasta los comienzos del siglo XXI, con el corpus escrito CORANE (2001), y posteriormente el CEDEL2 (2009) –que se encuentra todavía en la fase de compilación de textos– y el corpus oral SPLLOC (2008). La lingüística de corpus dispone de unas herramientas y métodos que ayudan a aportar descripciones mejoradas de la lengua del aprendiente que pueden ser utilizadas en la investigación de la adquisición y enseñanza de las L2 (Granger, 2002: 4).

aprendientes, sus aciertos mediante la introspección del investigador, pero también gracias a la disponibilidad de los contextos situacionales y de uso, que nunca se pierden de vista (Ellis, 1994). No obstante, para que esta labor resulte completa, conviene contar con un corpus de control nativo de iguales características (género textual, temas tratados, características sociales de los participantes, etcétera), que aporta objetividad – más allá de la intuición nativa– a los usos orales o escritos de los hablantes nativos (Nesselhauf, 2004). A partir de este corpus de control, puede establecerse de manera relativamente fiable qué usos lingüísticos no nativos resultan naturales y cuáles no; el procedimiento, que opta más por la vertiente descriptiva de la norma que por la prescriptiva, se basa en que la frecuencia con que se documenta un uso en el corpus de HNN se corresponda con la que presenta el corpus de HN. El criterio de la frecuencia, no obstante, no asegura completamente la naturalidad de un uso, según argumentamos más abajo.

He ahí, en efecto, una de las principales controversias en la investigación basada en el análisis de corpus de aprendientes: la determinación de la norma lingüística –como también señala Kaszubski (2000)–, esto es, los usos lingüísticos con los que los datos de la interlengua son contrastados y juzgados. Según se verá a continuación, Leech (1998) se muestra de acuerdo con la tendencia más reciente, que consiste en operar con el concepto de lengua normativa más descriptivo que prescriptivo, por lo que a partir de un análisis contrastivo de la interlengua sería posible conocer no solo los errores de los aprendientes, sino también los aciertos y los usos que tienden a sobreutilizar e infrautilizar. Por su parte, Quirk (1995) y Widdowson (1998) ponen de manifiesto los riesgos de prescindir de otros estándares en el ámbito de la enseñanza de lenguas, condición también aplicable a las investigaciones en el análisis de la interlengua.

El objetivo de este trabajo, por lo tanto, es proponer la aplicación de un concepto de norma que integre ambos enfoques: uno descriptivo y otro más regulador (prescriptivo). Dicho concepto, además, pretende superar el obstáculo que resulta de la controversia anteriormente descrita, y que dificulta investigaciones como la que desarrollamos actualmente mediante el análisis del corpus escrito de aprendientes informatizado CEDEL2⁶. Se trata, en suma, de la constatación de que disponer de un corpus de control nativo es necesario para analizar la naturalidad en la producción no nativa; y, por otro lado, de poner de manifiesto que el enfoque descriptivo es insuficiente para emitir juicios acerca de la adecuación de los usos lingüísticos.

⁶ El CEDEL2 es el primer corpus de aprendientes de español informatizado que cuenta con un corpus de control de HNN. Está siendo compilado por la Universidad Autónoma de Madrid, y las muestras textuales analizadas corresponden a redacciones elaboradas por personas de nivel educativo universitario. El estudio que actualmente desarrollamos se refiere, en concreto, al uso de verbos frecuentes en un nivel avanzado.

2. LA NORMA DESCRIPTIVA

En la actualidad, lo más frecuente en los estudios sobre la lengua del aprendiente basados en un corpus es disponer de un corpus nativo de control que se concibe como modelo de lengua, pues representa el estado lingüístico ideal (véase, para el inglés, Altenberg y Granger, 2001; Nesselhauf, 2005; Guo, 2006; y Prieto *et al.*, 2009 para el español) que debe ser alcanzado por los hablantes no nativos (en adelante, HNN). En este principio se basa una nueva corriente de estudios en la que se maneja un concepto de lengua más descriptivo que prescriptivo, lo que responde a la conciencia cada vez más extendida en el aula de LE de que la meta en el aprendizaje reside en alcanzar una competencia lo más próxima posible a la del nativo⁷, esto es, lograr la capacidad de llevar a cabo una actuación comunicativa natural.

De este enfoque de análisis contrastivo de interlengua basado en un corpus se desprenden, mediante la comparación entre un corpus de HN (hablantes nativos) y uno de HNN, las diferencias y las semejanzas lingüísticas existentes entre ambos. Leech (1998: XVII) se refiere a este enfoque con las siguientes palabras:

It enables us to investigate the non-native speaking learner's language (in relation to the native speakers') not only from a negative point of view (what did the learner get wrong?) but from a positive one (what did the learner get right?). For the first time it also allows a systematic and detailed study of the learner's linguistic behaviour from the point of view of "overuse" (what linguistic features does the learner use more than a native speaker?) and "underuse" (what features does the learner use less than a native speaker?).

Por primera vez en los estudios de interlengua, al amparo de la comparación de ambos corpus, se obtienen los errores que cometen los HNN, al tiempo que se recogen también los usos correctos; en relación con estos últimos, debe precisarse que la corrección desde un punto de vista gramatical no implica la adecuación de un elemento: las semejanzas en el uso (frecuencia, intencionalidad) apuntan a un dominio por parte de los aprendientes, mientras que las diferencias indican carencias. Según esta perspectiva, si el dominio lingüístico del aprendiente se equipara con el logro de una competencia casi nativa, aquellos rasgos empleados por los nativos y ausentes en los no nativos habrán de ser aprendidos o utilizados –en aquellos casos en los que conozcan las formas y sus usos contextuales pero no la frecuencia–; y viceversa: aquellos usos correctos formalmente que son frecuentes entre los no nativos y, en cambio, no aparecen en el corpus nativo deberán ser, si no erradicados, usados con menor asiduidad.

Que un corpus nativo de control se conciba como norma descriptiva o modelo de lengua para los HNN ha supuesto sin duda alguna un importante avance en los estudios de interlengua y en la posterior aportación de sus hallazgos al aula: el foco ahora está

⁷ Se trata de una aspiración no exenta de obstáculos, según se verá *infra*.

puesto en el uso real que los hablantes nativos hacen de la lengua, por lo que las intuiciones del investigador acerca de la naturalidad o corrección de los usos lingüísticos de los HNN pueden ser contrastadas y, por tanto, confirmadas o rechazadas. Asimismo, los rasgos y procesos –presentes en la producción no nativa– permitidos por el sistema lingüístico pero que no han sido sancionados o no forman parte de la variante lingüística considerada preferente⁸ por la mayoría de los miembros que integran la comunidad de hablantes de una lengua, se consideran desviaciones de la norma descriptiva concebida como referente lingüístico.

Obsérvense las siguientes muestras obtenidas del corpus CEDEL2⁹ referidas a la combinatoria léxica de los predicados *dar de* y *dar a*:

1. Ella va a dar de cenar a los niños
2. Luis daba a beber a los caballos

En el diccionario combinatorio *Redes* (2005: 642) aparece información al respecto que puede sintetizarse de la siguiente manera: 1) *dar de* + verbos que denotan ingestión: *comer/ beber/ cenar/ merendar/ desayunar/ mamar*; 2) *dar a* + verbos que denotan ingestión: *comer/ beber/ ?cenar/ ?mamar/ ?merendar/ ?desayunar*¹⁰. Estas construcciones aportadas por *Redes* ponen de manifiesto que el sistema tiene dos predicados con igual significado –*dar de* y *dar a*–, cuyo potencial combinatorio es, en principio, el mismo, a saber: seleccionan sustantivos que denotan ingestión. Sin embargo, si realizamos una búsqueda en un corpus nativo de referencia más extenso (CREA), obtenemos los siguientes resultados: *dar de mamar*: 28 resultados / *dar a mamar*: 1; *dar de comer*: 158 / *dar a comer*: 2; *dar de beber*: 21 / *dar a beber*: 7 (4 de los resultados pertenecen a un mismo hablante); *dar de cenar*: 4 / *dar a cenar*: 0.

Así, observamos que *dar a comer, a cenar, a beber y a mamar* son todas ellas opciones del sistema del español que o bien no son sancionadas por los HN, pese a que el sistema lingüístico lo permite, o son concebidas como opciones no preferentes –nótese que *dar a cenar* ni siquiera aparece en el corpus y *dar a mamar* una sola vez– frente a *dar de comer, de cenar, de beber y de mamar*, que son las coapariciones preferentes y, por tanto, naturales para los oídos nativos. Que los docentes y autores de materiales conozcan estos resultados permitirá que tanto el *input* como el *output* lingüístico de los aprendientes se filtre con esa norma descriptiva, dado que ahora sí se cuenta con usos lingüísticos reales, y no con la intuición nativa, que no siempre se corresponde con el

⁸ Por *preferente* entendemos la opción más frecuente, pero no exclusiva. Para una reflexión sobre la interpretación de la preferencia, véase Bosque (2001: 29-31).

⁹ Véase la nota 6.

¹⁰ Los infinitivos precedidos por el signo de interrogación no aparecen en *Redes*, pero ello no implica que la combinación no exista, pues en este diccionario no se pretende proporcionar paradigmas completos, sino definir y ejemplificar paradigmas representativos de las combinaciones léxicas que se agrupan y se analizan en las entradas del diccionario (Bosque, 2005: LXV).

uso real que los nativos hacen de su lengua; de este modo, volviendo al ejemplo anterior, en un intento de aproximar su producción a la del nativo, el aprendiente seleccionará el predicado *dar de*, aunque pueda tener almacenado como conocimiento receptivo el predicado *dar a* con algunos sustantivos del paradigma, como *comer* y *beber*. Pese a estas ventajas mencionadas, este principio es cuestionable: su aplicación a rajatabla puede llevar a errores.

2.1. Debilidades de la norma descriptiva

Que una determinada construcción o uso lingüístico abunde entre los HN y que, al mismo tiempo, sea muy escasa entre los HNN –y viceversa– impide hablar de una competencia similar entre ambos grupos. Ahora bien, las semejanzas no implican necesariamente dominio lingüístico, ni la presencia –puntual o reiterada– de un determinado fenómeno lingüístico en el corpus de HN implica necesariamente que su uso sea modélico. Los datos indican únicamente eso, frecuencia, y este es un concepto estadístico, es decir, está en función del número de apariciones de un determinado rasgo en los textos. Frecuencia y naturalidad suelen darse con simultaneidad, pero no coinciden necesariamente siempre: un determinado uso puede estar documentado pocas veces pero resultar muy natural para oídos nativos (Bosque, 2005: XLVIII), y algo frecuente puede ser interpretado como natural o no por los nativos, y hasta puede incluir errores, como se recoge a continuación.

La lengua nativa no está libre ni de marcas¹¹ –véanse los ejemplos [9] y [10] *infra*– ni de errores o, dicho de otro modo –según apuntábamos más arriba–, desviaciones involuntarias de la norma establecida en una comunidad de hablantes. Por ello, la producción nativa no puede erigirse, en un estudio de interlengua, como único referente a ciegas, sin matizaciones¹², lo que implica que el estudioso, a nuestro parecer, debe ser capaz de reconocer en su estudio dichas marcas y errores.

¹¹ Como es sabido, el concepto de *marca* ha evolucionado mucho desde su origen en el seno de la Escuela de Praga. Aquí nos interesa la perspectiva según la cual los elementos menos marcados se consideran más naturales, más frecuentes que los más marcados (en el sentido en que lo emplean, precisamente en el ámbito de la ASL, Larsen-Freeman y Long, 1994).

¹² En Sánchez Rufat (en prensa) se realiza un análisis de corpus contrastivo –nativo y no nativo– sobre el uso de la coma en oraciones subordinadas antepuestas. En este estudio se parte precisamente de esta asunción propia de los recientes trabajos en ASL (Guo, 2006: 5): que el corpus nativo ha de funcionar como referente o modelo. Contrario a las expectativas, el análisis muestra resultados sorprendentes: los nativos adolecen de un mayor desconocimiento del uso adecuado de la coma en este contexto sintáctico, y esa diferencia es significativa –tal y como se constata en el análisis–. Esto demuestra la necesidad sentida durante dicha investigación de recurrir a una norma reguladora de los contextos sintácticos concretos en los que el uso de la coma resulta adecuado, y, al mismo tiempo, prueba que no bastó establecer como única norma un corpus de control escrito de similares características.

Por lo que respecta, en primer lugar, al error, repárese en los siguientes ejemplos extraídos del corpus de control del CEDEL2¹³:

3. Se *dá* por muerta
4. Hay veces que las parejas gay *les* dan más amor a un niño

Muestras lingüísticas con este tipo de deficiencia ortográfica o morfosintáctica (de concordancia, en este caso) ponen de manifiesto que no toda la producción nativa que conforma el corpus de control está exenta de problemas. Aunque no sean frecuentes en el corpus, pues el porcentaje de aparición es bajo –por lo que podrían considerarse errores más bien aislados, individuales (o tipográficos) o no representativos de la producción nativa en su totalidad–, estos datos integrarían la descripción de la lengua nativa del grupo de control analizado, por lo que la producción no nativa estaría siendo juzgada con respecto a ellos.

Los hablantes nativos conocedores de la variante lingüística que se considera preferible por ser más culta¹⁴ saben que estos usos no forman parte de ese referente lingüístico. Pero si se establece como criterio para categorizar un dato de los aprendientes como aceptable natural el de su aparición en al menos cinco ocasiones en un corpus nativo de control (como fija Neeselhauf en su estudio sobre las colocaciones, 2005), usos similares a los de (3) y (4) podrían ser etiquetados de esta manera, como aceptables naturales.

Con esta misma línea argumentativa puede abordarse, por poner un ejemplo, el *queísmo*¹⁵, fenómeno presente y frecuente en ambos corpus. De los usos no nativos de *darse cuenta* en el corpus CEDEL2, se presenta este rasgo en un 30 % de los casos, frente al 18 % en los nativos. Obsérvense los siguientes ejemplos de HN ([5] y [6]) y de HNN ([7] y [8]):

5. Me he dado cuenta que lo construyo con cada acción que hago
6. De repente, nos dimos cuenta que no era posible viajar
7. Y me di cuenta que los italianos son tan aficionados
8. También me dí cuenta que quiero trabajar con latinos

¹³ Las cursivas de los ejemplos son nuestras.

¹⁴ Véase la nota 1.

¹⁵ El *DRAE* lo define de la manera siguiente: «m. Gram. Empleo indebido de la conjunción *que* en lugar de la secuencia *de que*; p. ej. **Me da la sensación que no han venido*».

Aunque la frecuencia pueda ser un indicador de dominio en la producción no nativa, especialmente cuando la frecuencia es amplia (Guo, 2006: 179), no constituye una garantía de dominio. De hecho, existen razones suficientes para que las muestras correspondientes a ese 18 % de queísmo no formen parte del referente lingüístico al que el alumno se ha de aproximar. Un estudioso de la interlengua, aunque opere con un modelo de lengua más descriptivo, puede optar por referirse a estos datos como usos no prescriptivos. Esto es así, no solo por el aliento de focos con vocación prescriptiva como la Real Academia Española, sino también por la «conciencia normativa» o el apego más o menos riguroso a lo lingüísticamente correcto –que suele identificarse, como ya se ha dicho, con la variante más culta– que la mayoría de los hablantes posee. En el caso del queísmo, aunque sea un fenómeno que pueda estar extendiéndose, aparece en el *DRAE* como «indebido» (véase la nota 15), pues, en el intento de equilibrar lo natural con lo correcto, no se considera lo suficientemente generalizado y aceptado en la población hispanohablante. Por su parte, muchos de los nativos que incurren en este uso no tendrían inconveniente en aceptarlo como algo inadecuado tras una explicación satisfactoria (pruebas de sustitución pronominal, por ejemplo), en razón de la mencionada «conciencia normativa» que los acerca a valores como la corrección o el prestigio lingüístico; ello es aún más esperable en los sujetos cuyas muestras integran el corpus de control manejado, en tanto que se encuentran cursando estudios superiores. He ahí el peligro de considerar sin cuestionamientos estos usos nativos como adecuados: aunque el objetivo del aprendiente sea lograr una competencia lo más próxima posible a la de un hablante nativo, debe tenerse en cuenta que ese hablante nativo dispone de su concepto de norma, en el que, de algún modo u otro, la prescripción está presente.

En relación con las marcas, si el dominio del aprendiente es juzgado en función de la proximidad de su producción con respecto al referente lingüístico tomado, hay que comprobar si el corpus de control representa fielmente la variedad de lengua seleccionada como modelo lingüístico, o si, por el contrario, incluye marcas o rasgos marcados que lo desvinculan de ese modelo. Porque, en efecto, otro de los problemas que presenta el objetivo de alcanzar una competencia lo más próxima posible a la del nativo lo constituye el hecho de que la lengua que se aprende contenga numerosos estándares. En tal caso, la norma descriptiva tendría que vincularse a uno de esos estándares o, en el ámbito del español, a un estándar panhispánico no practicado, en realidad, en ninguna área nativa concreta. Obsérvense los siguientes ejemplos:

9. Le doy un *puntaje*
10. Trato de alejarme pero *me da cosita*

Al tomar como norma descriptiva una variedad concreta, surge la posibilidad de contemplar en el corpus las variaciones diatópicas, diafásicas y diastrácticas que se

alejen de ella¹⁶. Así, ante el sustantivo *puntaje*¹⁷ que aparece en el ejemplo (9), el investigador, si fija como referente lingüístico el habla formal de los hablantes cultos del Cono Sur de Latinoamérica, ha de etiquetarlo como natural, dado que su uso está extendido en Argentina, Chile o Uruguay –en estas zonas, por tanto, a este sustantivo le corresponde un uso no marcado. Sin embargo, en España o México esta voz no se reconoce como perteneciente a la variedad extendida entre los hablantes cultos¹⁸, por lo que, en el caso de que la variedad de referencia seleccionada fuera la del español peninsular o mexicano, estaríamos ante un claro caso de sustantivo marcado desde una perspectiva diatópica. De acuerdo con ello, los hallazgos de este tipo de estudios de lingüística de corpus son aplicables a las aulas de LE. El aprendiente, si tiene esa posibilidad, selecciona una variedad lingüística de acuerdo con sus intereses y, en función de la variedad que elija, unos elementos y no otros serán marcados. Así, y siguiendo con el mismo ejemplo, si el HNN está aprendiendo la variante peninsular, el anterior sustantivo no formará parte del referente lingüístico, pues no es una voz que deba adquirirse para asemejarse a la producción nativa de esta zona geográfica; y, si de alguna forma, durante su formación se topa con él, necesitará saber que es un elemento marcado geográficamente.

De la misma manera, en (10) tanto el investigador como el aprendiente etiquetarán –en el lexicón, en el caso de este, y en la interlengua, en el caso de aquella– esta secuencia con la marca diafásica correspondiente, ya que *darle (a uno) cosa (algo)* es una expresión marcada en la situación comunicativa a la que pertenecen las muestras, que son redacciones de clase. Se trata, en efecto, de un uso coloquial presente en una situación que exige, si no una variedad formal, al menos neutra. Así, sustantivos como *vergüenza* o *apuro* resultan más precisos y apropiados en este tipo de texto; y, por otro lado, el diminutivo aporta una intensidad afectiva igualmente inesperada, pues estaríamos ante un rasgo de estilo muy poco frecuente.

¹⁶ Al igual que existen diferentes variedades lingüísticas insertas en el corpus CEDEL2, en el aula de LE el alumno está expuesto a un *input* integrado por diferentes variedades: por un lado, el profesor maneja una variedad diatópica más o menos acusada, y, al mismo tiempo, se acoge a la variación diafásica adecuada a cada situación comunicativa; y, por otro lado, los audios y vídeos presentes en los materiales utilizados en el aula suelen presentar, cada vez con más frecuencia, las mismas variaciones, pues se busca desarrollar y ampliar la competencia receptiva del estudiante al tiempo que se pretende la coherencia –la adscripción a una sola variedad– en la competencia productiva.

¹⁷ *DRAE* s. v. *puntaje*: «América m. Arg., Chile, El Salv., Hond. y Ur. Conjunto de puntos obtenidos en algún tipo de prueba».

¹⁸ Téngase en cuenta la diferencia entre *entender* y *reconocer* un uso. Aquí aplicamos *reconocer* cuando el hablante identifica un rasgo como perteneciente a la norma que hayamos seleccionado, independientemente de que dicho hablante también lo use o de que simplemente sea capaz de interpretarlo por el contexto aportado.

Como acabamos de ver, un corpus de control nativo no está libre de contener todas estas desviaciones (fallo ortográfico, queísmo, dialectalismo¹⁹, coloquialismo) del propio modelo que, de acuerdo con el concepto de norma descriptiva, representa. Dichas desviaciones ponen de manifiesto que una lengua debe analizarse y enseñarse sobre la base de algún tipo de norma prescriptiva. No cabe duda de que, cuando se lleva a cabo un análisis de la interlengua, disponer de un corpus de control nativo de características similares al de HNN (muestras que pertenezcan al mismo género discursivo, coincidencia de temas sobre los que versan las muestras, participantes con iguales características sociales, como la edad o la formación) es muy beneficioso. Permite saber qué es lo que los nativos producen típicamente en una situación comunicativa determinada, y el análisis contrastivo de la producción no nativa revela si hay determinados rasgos o procesos lingüísticos que los aprendientes tienden a sobreutilizar o infrautilizar en ese mismo contexto. Por ejemplo, un análisis del verbo *dar* en el corpus CEDEL2 indica que los aprendientes de nivel avanzado sobreutilizan las estructuras ditransitivas e infrautilizan las estructurales pronominales del verbo (como *darse*, con el sentido de *suced*er una cosa), así como las formadas con un complemento de régimen (como *dar por* + infinitivo; *dar con* + algo o alguien; *dar por* + participio). En cuanto a los sentidos del verbo, sobreutilizan los de *proporcionar* (*dar cariño*) y *producir(se)* (*dar sed*), y, en cambio, infrautilizan el de *ejecutar* (*dar un salto*). Son resultados de un análisis contrastivo que, además de ofrecer aplicaciones al proceso de aprendizaje de español como lengua extranjera, contribuyen al conocimiento de la interlengua de un grupo particular de aprendientes²⁰.

A pesar de ello, para poder describir la interlengua de un grupo de aprendientes, a los resultados de esta comparación, que son interpretados únicamente en clave de frecuencia –y ya se ha señalado que una amplia frecuencia no refleja necesariamente dominio–, hay que añadir los que se obtengan de un análisis que determine la adecuación de las muestras lingüísticas al sistema lingüístico del español y al contexto situacional en el que aparecen, es decir, a la variedad del español seleccionada como modelo; y, en relación con esto, un corpus nativo también puede presentar problemas (errores aislados o no tan aislados, como es el caso del queísmo) y ciertas particularidades (marcas) de distinto orden. El enfoque aportado por Leech (1998), sintetizado en el fragmento que reproducíamos más arriba, requería, por tanto, ser matizado y completado.

¹⁹ El español se puede concebir como un complejo dialectal (Borrego, 1999), por lo que empleamos aquí el término *dialecto* para referirnos a cada una de las variedades diatópicas del español.

²⁰ Los resultados de este análisis del verbo *dar* han sido presentados en el XIV Coloquio Internacional de Lingüística Íbero-Románica, Montpellier (30-05-2013), mediante una comunicación titulada «El uso del verbo *dar* en la producción escrita de estudiantes nativos y no nativos: un estudio contrastivo».

3. NORMA Y ESTÁNDAR

De lo expuesto anteriormente, nos interesa retomar, por un lado, la idea de que, en un estudio sobre la lengua de aprendientes basado en un corpus, disponer de un corpus de referencia nativo es, por sí solo, insuficiente para emitir juicios lingüísticos; y, por otro lado, se desprende de los ejemplos tratados que tanto la descripción de una lengua como el desarrollo de sus posibilidades de aplicación –como constituye la enseñanza de español como lengua extranjera– se impregnan del concepto de *norma* (Demonte, 2003: 1). En el establecimiento de esa norma, de esa variedad de referencia, debe acudirse al enfoque prescriptivo para solucionar problemas como los descritos en relación con el manejo de un corpus de control. Y dicha prescripción, en razón de las consideraciones de los especialistas y de la «conciencia normativa» de los hablantes a la que antes aludíamos, suele vincularse, como ya se ha mencionado, a variantes que se consideran preferibles por ser más cultas.

Así las cosas, en el momento de seleccionar esa variante lingüística preferible entran en conflicto las variedades que existen en la lengua. En un estudio como el que llevamos a cabo con el corpus CEDEL2, se está operando con una lengua, el español, que es oficial en veintidós países, por lo que plantearse qué variedad de lengua ha de erigirse como norma para contrastar y analizar los datos es una cuestión compleja de por sí. Más aún –como veremos más adelante– si se tiene en cuenta la enorme heterogeneidad respecto de la procedencia de los participantes del corpus. Son, así, veintidós las academias de la lengua que existen, cada una de ellas con una norma caracterizada por un rasgo propio de prestigio que representa lo que es correcto. Cada una de estas normas se establece con carácter sociolingüístico frente a la norma ideal de referencia (Borrego, 2001), esto es, lo que Demonte (2003: 4) denomina «el estándar en estado puro». Las primeras reflejan el origen de los usuarios, constituyen la variable de prestigio en un ámbito determinado; la segunda, a la que de ahora en adelante nos referiremos como *estándar*, es una koiné (Benincá, 1993: 248), una variedad común a un conjunto de dialectos en la que se descartan las peculiaridades y se seleccionan piezas léxicas y morfológicas transparentes²¹ y de consenso; esta definición refleja por sí misma la dificultad que se plantea a la hora de tener que seleccionar por consenso la opción que resulta más transparente o más prestigiosa.

Para la fijación del español estándar, ya hace algunos años que se está experimentado un giro hacia una idea más vaga e incluyente (Borrego, 1991); en otras palabras, se ha trazado un panorama más panhispánico. El prestigio ya no se identifica con un dialecto determinado, esto es, «no aspira ya a ir asociado a rasgos típicos del castellano peninsular como la pronunciación de la *c* y *z* como interdentales, de la *s* como apicoalveolar o el uso del leísmo de persona» (Hernández Alonso, 1996: 197,

²¹ La noción *transparente* que aparece en los trabajos de Benincá (1993) o Crystal (1995) resulta un tanto más superficial que la que nosotros manejamos. Aportamos una definición más precisa *infra*.

apud Demonte, 2003: 7). Si a este cambio de actitud se suman las características de un corpus como el CEDEL2, en el que los informantes no nativos aprenden el español tanto en EE. UU. como en el Reino Unido y los nativos proceden de muchos dialectos del español (desde el español de México al de Argentina pasando por el español peninsular), resulta lógico y esperable que las muestras de lengua no sean juzgadas de acuerdo con una norma sociolingüística concreta, y que, por lo tanto, se tomen los dos estándares (americano y peninsular)²² como el referente con el que las muestras, tanto de los HNN como de los HN, serán contrastadas. Este referente es el que debiera ser seleccionado tanto en un estudio de interlengua de las características descritas más arriba como en el aula de español (ya sea en un contexto de LE o de L2). Si bien es cierto que la mayoría de los docentes adopta como variedad para enseñar la propia (resulta hasta cierto punto inevitable no hacerlo), se intenta –como señala Demonte (2003: 7) en consonancia con la anterior explicación de Benincá (1993)– que dicha variedad esté formada por un léxico y una sintaxis no específica, y que las peculiaridades fónicas no se manifiesten de forma llamativa, aunque persistan rasgos tanto fonéticos como prosódicos que identifican la zona de la que se procede; de hecho, es frecuente que estas matizaciones se apliquen también en la interacción con hablantes de otras variedades del español. Según eso, estas características comprenden el material lingüístico más próximo a un estándar puro. Si a la enseñanza de este estándar se une el beneficio de exponer a los alumnos a la mayor cantidad de variantes posible (o normas sociolingüísticas), se logra que estos adquieran, como también observa Vázquez (2008), una competencia receptiva amplia y una producción unívoca no impuesta, resultado de la elección del aprendiente según factores de motivación o necesidades.

Así, en el desarrollo de un estudio de interlengua puede tomarse como referente la unión de los estándares peninsular y americano, que, aunque no son del todo homogéneos, se pueden unificar en el análisis y constituir un referente lingüístico de carácter panhispánico. Ahora bien, a efectos prácticos, tomar como modelo un estándar conlleva una serie de complicaciones y dificultades. Las lenguas estándar son abstractas, y no están descritas²³; los rasgos y procesos de una variedad estándar no configuran un sistema, un todo exhaustivo y homogéneo, sino que surgen (Crystal, 1995) por contraste con los rasgos y procesos considerados regionales, rurales,

²² La cuestión acerca de si deben aceptarse uno o dos estándares es recurrente en el mundo hispanohablante (Demonte, 2003: 6). No es el objetivo de este trabajo intentar aclarar esta cuestión. No obstante, debe tenerse en cuenta que el estándar americano, si es que existe, representa a veintiún países, veintiuna normas de las que se extraen los rasgos y procesos que configuran el estándar; mientras que el estándar peninsular se corresponde con un solo país, por lo que el esfuerzo homogeneizador para configurarlo es mucho menor. En este trabajo asociamos la variedad estándar con la mencionada configuración de carácter panhispánico.

²³ Se definen más por lo que no son que por lo que son, lo cual se debe a que las lenguas estándar son objetos que por definición están siempre incompletos (Demonte, 2003: 5), en proceso continuo de configuración y pactando consigo mismos: «son susceptible de cambios que dependen más de la voluntad de los usuarios que de propiedades objetivas, y constituyen entidades heterogéneas (sociales, convencionales, políticas, lingüísticas) tanto en su origen como en sus límites y contenido».

marginales, inapropiados o incorrectos. Para concretar, por tanto, ese estándar panhispánico se ha de establecer una serie de criterios para que el análisis pueda realizarse de manera sistemática.

En primer lugar, si –como mantiene Crystal (1995)– poder ser entendido por el mayor número posible de hablantes ha de ser un criterio de estandaridad para el referente lingüístico escogido, un ejercicio de introspección por parte del investigador basta para juzgar los datos sin marcas de especificidad, transparentes, como pertenecientes al estándar panhispánico –y, por tanto, categorizarlos como adecuados. La noción *transparente* se relaciona con la diferencia entre *entender* y *reconocer* que apuntábamos más arriba (ver la nota 18). No señalamos como transparente el uso o proceso que pueda ser interpretado fácilmente en un contexto dado, sino aquel que los hablantes reconocen como perteneciente a una zona suficientemente representativa del español, independientemente de que sea utilizado en su propia variedad. Así, consideramos insuficiente el criterio de Crystal (1995) que se refiere a la comprensión mutua para establecer la estandaridad: volviendo al ejemplo (9), cualquier hablante del español puede entender fácilmente el significado del sustantivo *puntaje*, pero no reconocerlo como un uso suficientemente representativo de uno de los dos estándares principales. El término, por tanto, sería no transparente, en razón de su pertenencia a una norma sociolingüística concreta y, por tanto, restringida a ciertas zonas geográficas. Si, no obstante, se duda de la transparencia de un uso a causa de las restricciones, limitaciones y a la parcialidad atribuible al desconocimiento propio, debe recurrirse a otras fuentes para poder categorizar los datos de forma sistemática.

Es cierto que, por las razones ya esgrimidas, no existen gramáticas ni diccionarios dedicados únicamente a ese estándar general que incluyan todos los rasgos y procesos panhispánicos (estos se describen en parte en algunas gramáticas normativas). Sí es cierto que, en lo que respecta al sistema gramatical, la *Nueva gramática de la lengua española* (2009; en adelante, *NGLE*) puede considerarse como obra de referencia para juzgar los datos de un corpus desde una visión panhispánica. En su introducción, se describe como una obra panhispánica, la primera gramática académica que refleja todas las variedades del español, pues constituye el producto de una acción conjunta de las veintidós academias; implica, por tanto, la articulación de un consenso que fija la norma común para todos los hispanohablantes, armonizando la unidad del idioma con la fecunda diversidad en que se realiza. La obra, en definitiva, admite recoger una muestra del español de todas las áreas lingüísticas con sus variantes geográficas y sociales, con lo que reflejaría así la unidad y la diversidad de la lengua española. Aunque resulta discutible que tal pretensión, acaso demasiado ambiciosa, haya podido lograrse por completo, la *NGLE*, por la amplitud de sus contenidos y por su vocación panhispánica, se convierte en una herramienta muy útil para la adscripción de los datos al estándar, pues permite adscribir al estándar panhispánico todos aquellos rasgos o procesos que no aparezcan marcados por su pertenencia a una o más variedades.

Tómese, por ejemplo, el componente sintáctico, el más impermeable a los cambios²⁴. En él, el estándar (sintáctico) se corresponde en su mayoría con la sintaxis normativa de todos los dialectos del español; en otras palabras, las estructuras suelen estar presentes en todas las variedades, y la única diferencia reside en su valor en la escala de disponibilidad. El siguiente ejemplo ilustra la manera en que pueden procesarse las estructuras sintácticas en un análisis de interlengua. Obsérvese la siguiente muestra extraída del corpus de HNN del CEDEL2:

11. Me da gusto saber que

A partir de un primer ejercicio de introspección, consideramos que, mientras que en el estándar americano nos parece natural y frecuente la combinación *dar gusto* tanto en estructuras transitivas (*da gusto nadar en el mar*) como en impersonales (*así da gusto*) y en ditransitivas (ejemplo [11]), en el estándar peninsular, en cambio, el valor en el grado de disponibilidad es el mismo en cuanto a las estructuras transitivas e impersonales, pero en las ditransitivas se reduce a contextos más restringidos, en ocasiones con cierta connotación sexual. Por lo tanto, en el español peninsular la construcción ditransitiva sería –y hablamos siempre desde la perspectiva de la introspección– la estructura preferente únicamente en este contexto, aunque su uso, como comprobamos después, no llega a ser del todo exclusivo de dicho contexto. En cualquier caso, si nuestras intuiciones se confirmasen, los tres usos de la estructura deberían adscribirse al estándar panhispánico, puesto que se documentan en al menos uno de los dos estándares principales.

Para verificar si nuestras intuiciones acerca del uso de esta expresión son correctas, hemos de comprobar el uso que verdaderamente hacen los nativos de esta estructura. El corpus de control del CEDEL2 no es lo suficientemente amplio como para albergar la estructura, pero con una búsqueda de la expresión (*me da gusto*) en un corpus mucho más extenso como el CREA podemos confirmar nuestras intuiciones: se obtienen 17 resultados, 15 de ellos de diferentes dialectos americanos, y 2 de ellos se corresponden, en principio, con el dialecto peninsular. De estos, uno se refiere al contexto antes mencionado: *A mí me da gusto, mi madre me da gusto, es una cerda*; y el otro, en cambio, no se refiere a este ámbito: *Creo que me da gusto también escribirte (...) Porque lo hago por decisión propia, porque quiero*. Ahora bien, hay que tener en cuenta que este ejemplo está extraído de una obra del escritor español Víctor Alba, quien estuvo exiliado en México durante más de veinte años. De igual modo, el ejemplo recogido en *Redes* para esta combinación procede también de México: *qué gusto me*

²⁴ Fue fijado a finales de la Edad Media; las pocas variaciones que persisten (Demonte, 2003: 11) tienen por lo general una larga historia y representan estrategias alternativas posibles que se sienten todas ellas como normales.

daría que los señores tomaran como ejemplo a este amante esposo. Dado que la información contenida en los diccionarios combinatorios²⁵ y en los corpus de referencia no es exhaustiva, pues pueden no contener usos muy utilizados por los usuarios de una lengua o incluir otros en un número tan bajo de apariciones que no se puedan hacer interpretaciones sobre su aceptabilidad (Nesselhauf, 2005: 49), y dados los riesgos, ya expuestos, que entraña prescindir de una norma reguladora, ha de recurrirse a una herramienta como la *NGLE* para determinar finalmente la estandaridad de los datos. En ella se presentan ejemplos de la expresión *dar gusto* tanto en estructuras transitivas como intransitivas indistintamente, sin ningún tipo de especificación geográfica, por lo que sus usos transitivo, ditransitivo o impersonal deben, en definitiva, interpretarse en el análisis como pertenecientes al estándar panhispánico, en tanto que los tres se extienden por, al menos, una de las orillas del Atlántico. En definitiva, las comprobaciones a través de un corpus extenso como el CREA y de un diccionario combinatorio como REDES, más la aplicación de la herramienta que supone la *NGLE*, confirman las intuiciones iniciales.

En lo que se refiere al vocabulario estándar, diccionarios como el *Diccionario de la lengua española (DRAE)*, el *Diccionario panhispánico de dudas (DPD)* o el *Diccionario de coloquialismos y términos dialectales del español* resultan muy útiles para fijarlo. Este vocabulario se presenta en formas léxicas sentidas como transparentes para la mayoría de los usuarios del español, y las obras mencionadas contienen información acerca de la estandaridad de algunas voces y construcciones, al tiempo que recogen también estructuras y piezas léxicas marcadas geográficamente y diastráticamente, que llegan a estar incluso indicadas como marginales, inapropiadas o incorrectas. El siguiente ejemplo ilustra este proceso de etiquetación de datos léxicos como estándar o no estándar. A juzgar por la información obtenida del *DRAE*, tanto *coche* como *carro* forman parte del estándar, pues cada pieza pertenece al estándar del español hablado en cada orilla del Atlántico: *coche* en el estándar peninsular y *carro*²⁶ en el estándar americano; la mayoría de hispanohablantes deberían reconocer ambos lemas como pertenecientes cada uno de ellos a una zona lo suficientemente representativa del español, independientemente de cuál sea el lema utilizado en su propia variedad o dialecto. Diferente es el caso de voces como *puntaje*, que forma parte de la norma culta argentina, chilena, salvadoreña, hondureña y uruguaya. El *DRAE* y el *Diccionario de coloquialismos y términos dialectales del español* ayudan en este caso a categorizar estos usos de manera sistemática. Así, mientras que en el anterior ejemplo *carro* es etiquetado como perteneciente al estándar panhispánico, *puntaje* se etiquetaría como restringido a determinadas zonas.

²⁵ Véase la nota 10.

²⁶ De una búsqueda del lema *carro* en el *DRAE* se obtiene: «9. m. Am. coche (l. vehículo automóvil)».

Dejando a un lado el criterio estrictamente diatópico, un terreno como la combinatoria léxica se muestra muy susceptible a la variación diafásica y a la variación diastrática. A la primera ya nos hemos referido más arriba a través de la expresión *me da cosita*. A la segunda podemos aproximarnos a través del grado de selección léxica: una gran especialización semántica suele asociarse a un nivel sociocultural alto (Montoro, 2010). Obsérvese el siguiente caso:

12. Aprendí que el museo fue establecido antes de su muerte y después Miró le *dio muchas obras*

Al uso del verbo *dar* en este contexto no puede otorgársele una marcación neutra²⁷, sino de cierta pobreza léxica, pues su grado de especialización semántica se halla por debajo de lo que un hablante nativo consideraría natural. Un verbo como *donar*, sin embargo, sí cumpliría esa función en la situación descrita y podría ser marcado de forma neutra. El estándar manejado, por tanto, debe atender también a las diferencias diafásicas y diastráticas contenidas en cada uno de los estándares que lo componen.

Asimismo, hay variaciones en el uso individual de ciertas combinaciones por parte de los nativos. Tomemos, por ejemplo, las combinaciones con verbos frecuentes²⁸. Cuando se pide a nativos –incluso de la misma variedad diatópica y con semejante destreza lingüística– que juzguen una determinada combinación, a menudo no la encuentran ni completamente aceptable ni completamente inaceptable (Mel'cûk, 1996: 171, Stengers, 2009). El concepto de error, tan unido al de norma, es, en este ámbito, más difícil de captar (Nesselhauf, 2005: 39) y, por lo tanto, es más complejo establecer una tipología con los dominios de cada categoría claramente delimitados. Debido a esta variabilidad, tanto en el tipo de variación (diatópica, diafásica, diastrática) a la que puede estar sujeta una combinación léxica como en el grado de aceptación de ciertos usos, se puede establecer una triple categorización de los datos para describir este aspecto de la interlengua. En primer lugar, pueden distinguirse los usos (formas y significados) del verbo que se sobreutilizan o se infrautilizan de los que se usan con igual frecuencia con la que lo hacen los nativos; para ello, como ya se ha indicado, es suficiente disponer de un corpus de control nativo, el cual nos permite calcular frecuencias. En segundo lugar, hay que clasificar esos datos de acuerdo con su aceptabilidad. Para ello, no es suficiente contar con un corpus de control ni con uno más extenso como el CREA, sino que es necesario aplicar el estándar que hemos

²⁷ A un elemento se le aplica una marcación neutra cuando se usa en situaciones que de por sí no requieren de un mayor grado de especialización semántica para que resulten naturales a oídos nativos (Montoro, 2010: 257).

²⁸ Véase la nota 6.

seleccionado para considerar los usos aceptables o inaceptables. Según el grado de aceptabilidad, se puede establecer una escala formada por dos categorías: aceptable-natural y aceptable-restringido. Los usos clasificados como naturales no difieren en nada del estándar, esto es, no presentan ninguna desviación ortográfica, semántica, gramática, combinatoria, diatópica, diastrática ni diafásica; en cambio, los usos restringidos son muestras de usos correctos desde una perspectiva formal, pero considerados no preferentes por lo hablantes nativos²⁹ (como los casos *de dar a vistos* más arriba). En tercer lugar, el término inaceptabilidad y otros términos relacionados (error, fallo, incorrección) toman como único sentido en un estudio de este tipo aquellas formas o usos que difieren del estándar en cuanto a los aspectos ortográfico, semántico, gramatical, combinatorio, diatópico, diastrático y diafásico. Dado que el análisis tiene como objeto de estudio la interlengua, no se deben inferir afirmaciones del tipo «los errores son malos» o «es un error, por lo que necesariamente debe ser corregido o erradicado en este nivel de dominio».

4. CONCLUSIONES

Teniendo en cuenta todas las consideraciones anteriores, para la descripción de la interlengua de un grupo de aprendientes resulta de enorme utilidad disponer de un corpus de control nativo de iguales o muy similares características, que permite, al contrastar las frecuencias de los usos de uno y otro corpus, determinar qué elementos sobreutilizan o infrautilizan los aprendientes. No obstante, puede ocurrir que un determinado uso esté poco documentado en el corpus de control, pero resulte muy natural para el hablante nativo, y también es posible que un uso frecuente sea interpretado por un alto porcentaje de nativos como no natural o incluso, mediante el filtro de la «conciencia normativa» que todo hablante posee, como erróneo. Para superar estos obstáculos, es necesaria la aplicación de una norma reguladora que precise el grado de adecuación de los usos no nativos, en tanto que el contraste con el corpus de control, como se ha dicho, no siempre resulta suficiente. Piénsese, por ejemplo, en el uso del *queísmo*, aludido más arriba: aunque su presencia en un corpus de control sea relativamente amplia, tanto los focos especializados como un buen número de hablantes lo señalarían como no adecuado.

Para un corpus como el que hemos venido mencionando, el CEDEL2, en el que tanto los HN como los HNN manejan referentes diatópicos distintos, la unión de los estándares peninsular y americano en un estándar panhispánico constituye un referente lingüístico operativo para categorizar los datos de la interlengua. Para ello, sugerimos, una vez establecida una primera categorización basada en los índices de frecuencia, una doble categorización: el primer paso consiste en clasificar los datos de acuerdo con su

²⁹ Véase la nota 8.

aceptabilidad, con lo que se distingue entre secuencias aceptables, que, a su vez, se subdividen en estructuras naturales y restringidas, e inaceptables; y el segundo paso se centra en los datos inaceptables y consiste en localizar la fuente de la inaceptabilidad. El estándar propuesto, por lo tanto, contendrá los usos considerados aceptables en al menos una de las dos esferas que lo componen. Además, la determinación de un estándar permite filtrar los usos desde las perspectivas diatópica, diafásica y diastrática. Piénsese en los ejemplos aportados aquí: *puntaje*, desde el punto de vista diatópico, no pertenece al estándar porque se restringe a ciertas zonas de América —es decir, el término puede ser entendido pero no es reconocido por un hablante como propio de la totalidad del estándar americano—; *dar (a alguien) cosa (algo)*, desde el punto de vista diafásico, no pertenece al estándar propuesto porque este toma exclusivamente una variedad formal o, al menos, de marcación neutra; y, en fin, *dar obras*, desde el punto de vista diastrático, no pertenece al estándar porque el contexto lingüístico exige un mayor grado de especialización semántica en el verbo seleccionado. En los casos en los que el investigador albergue dudas sobre la estandarización de algún elemento y sienta la necesidad de contrastar sus intuiciones, puede recurrir a herramientas como un corpus más amplio (CREA), diccionarios apropiados (*Redes, Diccionario panhispánico de dudas...*) o la *Nueva gramática de la lengua española*, mediante un proceso similar al que hemos expuesto a propósito de las estructuras ditransitivas con la expresión *dar gusto*.

REFERENCIAS BIBLIOGRÁFICAS

ALTENBERG, Bengt y Silvine GRANGER (2001): «The grammatical and lexical patterning of MAKE in native and non-native student writing», *Applied Linguistics*, 22, 2, 173-195.

BENINCÀ, Paola (1993): «Sintassi», en Sobrero, Alberto A., ed., *Introduzione al italiano contemporaneo. Le strutture*, Roma-Bari, Laterza, 247-288.

BORREGO NIETO, Julio (1999): «El español de Castilla León. ¿Modelo lingüístico o complejo dialectal?», en Álvarez Tejedor, Antonio, coord., *La lengua española patrimonio de todos*, Burgos, Caja de Burgos, 13-40.

BOSQUE MUÑOZ, Ignacio (2001): «Sobre el concepto de colocación y sus límites», *LEA*, 23, 1, 8-35.

BOSQUE MUÑOZ, Ignacio, dir. (2005): *Redes. Diccionario combinatorio del español contemporáneo*, Madrid, SM.

CORDER, Stephen Pit (1967): «The significance of learners' errors», *IRAL*, 5, 161-170.

COSERIU, Eugenio (1967): *Teoría del lenguaje y lingüística general*, Madrid, Gredos.

CRYSTAL, David (1995): *The Cambridge Encyclopedia of the English Language*. Cambridge, Cambridge University Press.

DEMONTE, Violeta (2003): «Lengua estándar, norma y normas en la difusión actual de la lengua española», *Circunstancia* [en línea]: <<http://digital.csic.es/bitstream/10261/13074/1/Documento1.pdf>>. [Consulta: 15-06-2013].

ELLIS, Rod (1994): *The Study of Second Language Acquisition*, Oxford, Oxford University Press.

FERNÁNDEZ, Sonsoles (1997): *Interlengua y análisis de errores*, Madrid, Edelsa.

FITCH, Roxana (2011): *Diccionario de coloquialismos y términos dialectales del español*, Madrid, Arco Libros.

GRANGER, Sylviane (2002): «A bird's-eye view of learner corpus research», en Granger, Sylviane *et alii*, eds., *Computer Corpora, Second Language Acquisition and Foreign Language Teaching*, Amsterdam, Benjamins, 3-33.

GRANGER, Sylviane *et alii*, eds. (2002): *Computer Corpora, Second Language Acquisition and Foreign Language Teaching*, Amsterdam, Benjamins.

GUO, Xiaotian (2006): *Verbs in the Written English of Chinese Learners: A Corpus-based Comparison between Non-native Speakers and Native Speakers*, tesis doctoral, Universidad de Birmingham, *The linguistics journal* [en línea]: <http://www.linguistics-journal.com/thesis_Guo.pdf>. [Consulta: 20-06-2013].

KASZUBSKI, Przemek (2000): *Selected aspects of lexicon, phraseology and style in the writing of Polish advanced learners of English: a contrastive, corpus-based approach*, tesis doctoral, Universidad de Adam Mickiewicz [en línea]: <<http://www.staff.amu.edu.pl/~przemka/rsearch.html#PhD>>. [Consulta: 21-06-2013].

LARSEN-FREEMAN, Diane y Michael H. LONG (1994): *Introducción al estudio de la adquisición de segundas lenguas*, Madrid, Gredos.

LEECH, Geoffrey (1998): «Preface», en Granger, Sylviane, ed., *Learner English on Computer*, London, Longman, XIV-XX.

LOZANO, Cristóbal (2009): «CEDEL2: Corpus Escrito del Español L2», en Bretones Callejas, Carmen M. *et alii*, eds., *Applied Linguistics Now: Understanding Language and Mind*, Almería, Universidad de Almería, 197-212.

MEL'ČUK, Igor (1996): «Lexical Functions in Lexicography and NLP» en Wannet, L., ed., *Lexical Functions: A Tool for the Description of Lexical relations in a Lexicon*, Amsterdam, John Benjamins, 37-102.

MONTORO DEL ARCO, Esteban T. (2010): «Norma y uso de las colocaciones léxicas en español», en Aleza Izquierdo, Milagros, coord., *Norma y usos correctos en el español actual*, Valencia, Tirant lo Blanch.

NEMSER, William (1971): «Approximative systems of foreign language learner», *IRAL*, 9, 2, 115-123.

NESSSELHAUF, Nadja (2004): «Learner corpora and their potential for language teaching», en Sinclair, John, ed., *How to use corpora in language teaching*, Amsterdam, Benjamins, 125-152.

NESSSELHAUF, Nadja (2005): *Collocations in a Learner Corpus*, Amsterdam, Benjamins.

PRIETO GONZÁLEZ, Sabela *et alii* (2009): «Córpora y enseñanza de lenguas: se buscan colocaciones», en Cantos Gómez, Pascual y Aquilino Sánchez Pérez, eds., *A survey on corpus-based research*, Murcia, AELINCO, 336-373.

QUIRK, Randolph (1995): *Grammatical and lexical variance in English*, London, Longman.

REAL ACADEMIA ESPAÑOLA (2001): *Diccionario de la lengua española*, Madrid, Espasa.

REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA). *Corpus de referencia del español actual* [en línea]: <<http://www.rae.es>>. [Consulta: 20-06-2013].

REAL ACADEMIA ESPAÑOLA: *Diccionario panhispánico de dudas* [en línea]: <<http://www.rae.es/dpd/>>. [Consulta: 26-06-2013].

REAL ACADEMIA ESPAÑOLA y ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (2009): *Nueva gramática de la lengua española*, Madrid, Espasa.

SÁNCHEZ RUFAT, Anna (en prensa): «La coma en oraciones subordinadas antepuestas: un estudio de corpus contrastivo», en Montoro del Arco, Esteban T., coord., *Ortografía e Imagen Personal*, Granada, Universidad de Granada.

SCHACHTER, Jacquelyn y Marianne CELCE-MURCIA (1977): «Some reservations concerning error analysis», *TESOL*, 11, 4, 441-451.

SELINKER, Larry (1972): «Interlanguage», *IRAL*, 10, 2, 209-231.

SINCLAIR, John, ed. (2004): *How to use corpora in language teaching*, Amsterdam, Benjamins.

SOBRERO, Alberto A., ed. (1993): *Introduzione al italiano contemporaneo. Le strutture*, Roma-Bari, Laterza.

STENGERS, Helene. (2009): *The Idiom Principle put to the test: An exercise in Applied Comparative linguistics*, Tesis doctoral inédita, Bruselas, Universidad Vrije de Bruselas.

VÁZQUEZ, Graciela (2008): «¿Qué español enseñar? Preguntas frecuentes», teleconferencia pronunciada en las III Jornadas de Español como Lengua Extranjera, I Congreso Internacional de Enseñanza e Investigación en ELSE, en la Universidad Nacional de Córdoba, Argentina [en línea]: <http://www.lenguas.unc.edu.ar/elsecongreso/teleconferencia_vazquez.pdf>. [Consulta: 24-06-2013].

WIDDOWSON, Henry (1998): «Skills, abilities and contexts of reality», *Annual Review of Applied Linguistics*, 18, 323-333.