

DE MAURO, T.; MANCINI, F.; VEDOVELLI, M.; VOGHERA, M.: *Lessico di frequenza dell'italiano parlato*. Milano, Etaslibri, 1993, 542 pp. + 2 diskettes.

Pura GUIL

Aunque la cercanía temporal, como uno de los factores que determinan la oportunidad de una recensión, no pueda ser aducida en este caso, ya que su publicación coincidirá aproximadamente con el segundo aniversario del *Lessico di frequenza dell'italiano parlato* (LIP) —primavera de 1993—, en cambio sí que resulta de aplicación otra, y no la menor, de sus posibles justificaciones: la importancia y trascendencia de la obra reseñada.

Dada la peculiar historia lingüística italiana —cuyas líneas maestras traza De Mauro, junto con la descripción de la situación actual, en el primer capítulo del libro— sólo desde hace muy pocos años, en Italia, se está *hablando* italiano de forma difundida y habitual, no limitada a Toscana o a Roma, ni a ocasiones marcadas formalmente. Pero este italiano oral se aleja del italiano descrito por las gramáticas, que tienen fundamentalmente por modelo el uso literario escrito. Las dificultades propias del análisis de textos orales se incrementan en el caso del estudio del uso hablado del italiano, con la no fácil diferenciación de esta variedad diamésica —con sus posibles subvariedades— respecto no sólo al uso escrito, sino también respecto a otras variedades diastráticas y diafásicas del repertorio italiano, algunos de cuyos rasgos distintivos entreveran los textos orales. Así pues, si se tiene en cuenta su valor como instrumento de verificación de meras expectativas o hipótesis, mediante la aportación de consistentes datos cuantitativos, no resulta exagerado afirmar que el LIP ha de marcar un antes y un después en uno de los sectores de la investigación lingüística que más atención ha recibido en los últimos años, ya que la lengua oral añade a su interés intrínseco el de constituir una variedad dinámica en la que se evidencian con mayor claridad que en la escrita las tendencias evolutivas en curso en el italiano contemporáneo.

Sobre la base de un *corpus* de italiano oral de 500.000 palabras —extensión considerada necesaria y suficiente para el fin propuesto— y por medio de su elaboración y lematización automáticas, se obtiene un total de 15.641 lemas, presentados en varias listas de frecuencia que ocupan dos tercios del volumen, destinándose el primer tercio —integrado por seis capítulos— a la presentación de las hipótesis de partida, objetivos perseguidos, criterios aplicados, primeros análisis del material lingüístico obtenido y evaluación final de los resultados.

De gran acierto hay sin duda que calificar la innovadora decisión adoptada por los autores de hacer también público el propio *corpus* de base, recogido en dos *diskettes* que se incluyen en el volumen. Con ello —muy en línea con su postura de explicitar diáfananamente los criterios aplicados— propician el control de la labor efectuada, y más aún, no sólo permiten así la contextualización de los datos, sino que también brindan un amplio y representativo *corpus* de *parlato* seleccionado, registrado, transcrito y memorizado en soporte electrónico de fácil manejo, es decir, ya listo («precocinado») para ulteriores análisis que gozarán, además, de la ventaja de poder ser cotejados con los datos globales ya cuantificados.

Con el fin de obtener un *corpus* representativo desde el punto de vista diatópico, la recogida de datos, que en vista de la situación lingüística italiana no resulta adecuado circunscribir a una sola área, se ha realizado en cuatro ciudades —Milán, Florencia, Roma y Nápoles— consideradas centros irradiadores del italiano *standard*, cuya elección justifica detalladamente De Mauro, quien informa también de la selección exclusiva de producciones orales orientadas en modo dominante hacia el *standard*, si bien registrando como tales los dialectalismos que pudieran presentar.

Pero no sólo se ha tenido en cuenta la variación geolingüística, sino también la variación tipológico-textual con el fin de facilitar el cálculo de la dispersión y ofrecer datos tanto sobre la frecuencia como sobre el uso. Asumiendo como lengua oral toda producción fónica espontánea, se atribuye a la conversación el carácter de sistema primario de modelización, diseñándose a partir de él un *continuum* de usos graduado en cinco puntos a lo largo de la siguiente escala:

- «1. scambio bidirezionale faccia a faccia con presa di parola libera (conversazione in tutte le sue possibili forme);
2. scambio bidirezionale non faccia a faccia con presa di parola libera (conversazioni telefoniche);
3. scambio bidirezionale faccia a faccia con presa di parola non libera (dibattiti, interviste, interrogazioni ecc.);
4. scambio unidirezionale in presenza di destinatario/i (lezioni, conferenze, omelie, comizi ecc.);
5. scambio unidirezionale o bidirezionale a distanza (trasmissioni radiofoniche e televisive)» (p. 35).

Esta tipología une a su propio valor de sistematización original, la ventaja de ofrecer la posibilidad de confrontar las frecuencias generales y las de cada uno de los cinco géneros textuales, pudiendo establecer así su incidencia relativa en la configuración global de la lengua oral. Sin embargo, seguramente no habría estado de más subdividir el tipo final, contabilizando separadamente las transmisiones radiofónicas y las televisivas, para poder calibrar así el impacto que el soporte de la imagen tenga en estas últimas.

En coherencia con los criterios citados, la recogida de datos se ha efectuado en modo cuantitativamente homogéneo (unas 100.000 palabras por cada tipo textual, a razón de 25.000 por cada una de las cuatro ciudades elegidas). Las correspondientes grabaciones han tenido lugar entre noviembre de 1990 y julio de 1992: los datos manejados son por tanto de una extraordinaria actualidad. Muy sugestiva resulta la información relativa a la menor fluencia (número medio de palabras por minuto) observada en Milán, que ha hecho necesario un mayor tiempo de grabación, respecto a —en orden creciente— Roma, Nápoles y Florencia. Con el fin de preservar la espontaneidad de la lengua usada, se han excluido técnicas estimuladoras, como las entrevistas, tratando en la medida de lo posible de realizar grabaciones ocultas.

Respecto a la extracción sociolingüística de los hablantes, que se ha procurado fuera suficientemente representativa de la situación italiana, la oportuna informa-

ción se ha introducido en una base de datos, pero sólo se ofrece de forma global, es de suponer que por el notable incremento de trabajo que habría ocasionado su indicación —sin embargo, utilísimas— en los distintos ficheros del *corpus*.

Dada la amplitud del *corpus*, la ardua labor de transcripción, cuyos procedimientos se detallan, ha corrido a cargo de cuatro colaboradores, lo que sin duda puede explicar alguna falta de homogeneidad en la aplicación de las convenciones —como ellos mismos reconocen—, que puede percibirse, al examinar los textos de los *diskettes*, básicamente en las indicaciones de datos contextuales o de fenómenos paralingüísticos, pero que no afecta a las «palabras» transcritas y, por tanto, tampoco al leuario.

Otro de los grandes aciertos del LIP radica en haber realizado la categorización gramatical atendiendo a la polifuncionalidad de las palabras, registrando como lemas diversos aquellas que asumen funciones gramaticales diversas, con lo que se incrementa enormemente el alcance informativo de las listas de frecuencia. De la elaboración y lematización automática del material lingüístico da cuenta Federico Mancini, ingeniero de la IBM Semea, aportando información técnica detallada del proceso y sistema adoptado. Corresponde a Miriam Voghera señalar los criterios analíticos seguidos en la asignación de cada ocurrencia del *corpus* a una determinada clase gramatical y los problemas que un tal análisis plantea en la lengua oral. Se han subdividido algunas de las clases tradicionales y añadido otras distinciones en aras de una mayor especificidad, incluyendo dialectalismos y exotismos. Merece la pena resaltar, por su valor de novedad y notable interés, la elaboración, dentro de la clase «interjecciones», de una lista de 24 fonosímbolos y sus variantes, con indicación del acto lingüístico que manifiestan o función discursiva que desempeñan. Igualmente, por primera vez en un léxico de esta índole, han sido registradas y clasificadas las expresiones polirremáticas, esto es, aquellas cuyo significado global no es directamente deducible a partir del de cada uno de sus miembros, labor de la que podrán, sin duda, obtener un gran rendimiento los estudios de lingüística en general y de forma muy directa aquellos de lingüística aplicada centrados en la enseñanza del italiano.

Por su importancia, nos hacemos aquí eco del anuncio realizado por De Mauro en la obra que estamos examinando, quien, a propósito de las expresiones polirremáticas, manifiesta que serán también tenidas en cuenta en el *Grande dizionario italiano dell'uso* (GRADIT) que está elaborando, con la intervención en calidad de asesores de Annibale Elia, Giulio Lepschy y Edoardo Sanguineti, y que espera concluir antes del fin de 1996 para su publicación por UTET y Paravia.

Desde una perspectiva española, a partir de la experiencia en nuestro propio idioma, se trata sin duda de una excelente noticia, pues a nadie que haya utilizado el de María Moliner se le ocultan las ventajas que puede proporcionar la consulta de un diccionario «de uso» del italiano, máxime que dada su autoría es de esperar un nivel de calidad, cuanto menos, similar. Y decimos *cuanto menos*, porque es de suponer que contará con el subsidio de mayores medios que los artesanales que tuvo a su disposición D^a María Moliner para elaborar el suyo, magnífico (teniendo probablemente que atender, además, a «sus labores»).

Volviendo de nuevo al LIP, aunque presentados como simples esbozos, mucha más alta consideración alcanzan los concienzudos análisis del material lingüístico obtenido, inaugurando así la trayectoria de futuros trabajos investigadores.

A partir de la asunción, comúnmente aceptada, de que el italiano contemporáneo se caracteriza por la reducción y simplificación de las estructuras sintácticas, manifestada principalmente en los usos orales, se centra Miriam Voghera en el examen de dos fenómenos emblemáticos, múltiples veces señalados: la reducción del número y la variedad de los conectivos y la simplificación de los paradigmas verbales. A través de un análisis minucioso de los datos recabados del LIP, demuestra su distinta incidencia en los varios géneros textuales hablados, tendencialmente mayor en los dialógicos y menor en los prevalentemente monológicos, atribuible a la presión ejercida por las condiciones de enunciación en la selección de los medios verbales; con ello se confirma lo oportuno de la subdivisión tipológica propuesta dentro de los usos orales. La confrontación de tales datos con los correspondientes al uso escrito, obtenidos del *Lessico di frequenza della lingua italiana contemporanea* (LIF) (U. Bortolini, C. Tagliavini, A. Zampolli, Milano, Garzanti-IBM, 1972), le permite distinguir entre usos propios del *parlato* (señales de articulación textual, expansión del presente de indicativo) y usos frecuentes en el *parlato* pero ampliamente compartidos por el italiano escrito (reducción del número y variedad de las conjunciones y disminución del uso del futuro simple y del presente de subjuntivo).

Como broche de su trabajo, señala Voghera: «Solo attraverso una comparazione sistematica tra dati raccolti da materiale scritto e parlato sarà dunque possibile cogliere le vere differenze tra le due norme di realizzazione e individuare con maggiore chiarezza le tendenze evolutive in atto nell'italiano contemporaneo, che si possono definire tali proprio per la loro diffusione in tutti gli usi della lingua» (p. 111). Conclusión ésta, suscribible sin reserva alguna, que pone en guardia respecto a apresuradas generalizaciones, fruto más de la sensibilidad del usuario, inmerso en la lengua, que del trabajo metódico y riguroso del estudioso de la lengua.

Dice renunciar Massimo Vedovelli a un examen exhaustivo de los datos del LIP, fuera de lugar en el contexto de una introducción, y propone una primera lectura susceptible de confirmar hipótesis anteriormente formuladas sobre rasgos distintivos de la lengua oral por contraposición a la escrita, mecanismos propios de la gestión de la interacción oral y características de la lengua italiana contemporánea determinables a partir de su uso oral.

Si no exhaustivos, sí ciertamente numerosos son los datos que maneja y los aspectos y matizaciones que evidencia Vedovelli, tanto en su lectura 'interna' del LIP como en relación con otras listas de frecuencia del italiano: distinta capacidad de cobertura textual que confirma la relativa pobreza léxica de la lengua oral, primeras posiciones de la lista de frecuencias copadas por palabras gramaticales y breves, reestructuración de las posiciones relativas en el rango de uso que afecta principalmente a verbos y adverbios con desesemantización parcial correlacionada con un aumento de su valor funcional e indicativa de un distinto recurso a las varias categorías gramaticales, contribución del mecanismo textual de repetición en la estructura de la lista de

frecuencias, ausencia de específicos campos semánticos privilegiados respecto al escrito, y tantos otros rasgos que no es posible recoger aquí.

Parece, sin embargo, oportuno reseñar una cuestión, sobre todo por el impacto que pudiera causar en el usuario desprevenido. Puede llamar la atención, al consultar las listas del LIP, comprobar que ciertos vocablos muy comunes alcanzan una baja frecuencia (p.ej. los términos de parentesco) o incluso ni siquiera aparecen (p. ej. *coniglio, elettricità, pettinare, salame, tossire*). Se trata de un conocido fenómeno que aún no ha recibido una sistematización definitiva y que lleva a la necesidad de distinguir entre vocabulario de 'alta frecuencia' y vocabulario de 'alta disponibilidad', integrado este último por «quell'area di lessico legata a aree di esperienza altamente familiari eppure caratterizzata da bassa frequenza e da forte dipendenza dal contesto» (p. 140). Como muestra Vedovelli —cuyos trabajos en este ámbito son notorios—, la confrontación del LIP con otras listas de frecuencia puede arrojar luz sobre esta problemática, que afecta a la estructuración del léxico de una lengua, ayudando a determinar los componentes del léxico de alta disponibilidad y su estabilidad relativa.

A propósito de las cifras que arroja el LIP sobre dialectalismos (1,33% del *corpus* total) señala De Mauro su mayor incidencia en el tipo textual 'conversaciones telefónicas' (3,86%), «luogo di massima rilassatezza del parlato», y, en cuanto a ciudades, en Nápoles (2,57%) frente a Milán (0,32%) que «in fatto di adesione allo standard, supera la stessa Firenze». Mucho más restringida es la presencia de exotismos: 1.474 ocurrencias, que representan el 0,30% del total. De ellas, 1.049 corresponden al inglés —alcanzando *okay* nada menos que 242 ocurrencias— y al latín 244. Resulta manifiesto, por tanto, que «il lessico dell'italiano parlato è per il 98,4% fatto di parole di riconosciuto e consolidato uso nazionale comune» (p. 150), cifra que se ve aumentada si se añaden los dialectalismos, con lo que debe admitirse que la presencia innegable de exotismos tiene sin embargo una mínima significación estadística.

Se encarga también Tullio De Mauro del capítulo conclusivo, en el que, a la vista de los resultados, evalúa el grado de confirmación alcanzado por las hipótesis de partida, pasando revista a múltiples aspectos, desde los de pura elaboración material a los de orden más estrictamente lingüístico, general o histórico.

Mayor riqueza de formas morfosintácticamente ambiguas que la lengua escrita, diferente recurso a las distintas categorías gramaticales, selección entre sinónimos, léxico menos rico de lemas diversos pero de una alta frecuencia de uso frente a la necesidad de la lengua escrita para alcanzar un similar nivel informativo de un repertorio léxico más variado y por ende con baja frecuencia (lo que induce a De Mauro, con su estilo característico, a afirmar: «come i veri signori il parlato può prendersi il lusso di vestir panni più dimessi», p. 155), son algunas de las características del *parlato* —mejor «parlati», vista su diversificación interna—, aludidas en las páginas anteriores de la Introducción, que ahora De Mauro recoge y comenta como reflejos léxico-estadísticos de las dos modalidades de uso y norma determinadas por la diversidad semiótica entre escritura y oralidad.

Dichas constataciones, en el marco de una Italia en la que el italiano ha entrado en los últimos decenios en el ámbito del *parlato*, reservado secularmente a los dialectos, permiten confirmar, según De Mauro, el proceso en curso de cambio en el uso y la norma de la lengua italiana en su conjunto. Una lengua italiana aún sensible a la variación diatópica pero en la que se han impuesto las tendencias unitarias, convergiendo hacia unos usos orales comunes que se manifiestan en todas las áreas del léxico, incluida, significativamente, la relativa a la cotidianidad.

Concluye así la primera parte del LIP que si como introducción cumple ampliamente con su función de explicitar y justificar los innovadores criterios aplicados, como estudio razonado y reflexivo constituye una valiosa aportación al análisis del uso oral de la lengua en general y del italiano en particular.

La segunda parte del volumen está integrada por cinco listas de frecuencia.

La lista A (pp.171-400) recoge las informaciones relativas a los 7.213 lemas con al menos tres ocurrencias, que en el *corpus* aparecen en 20.388 formas, incluidas las dialectales. Además de la categoría gramatical y el rango de uso, para cada lema y para cada una de sus correspondientes formas se indica el número de ocurrencias registrado en cada uno de los cinco tipos textuales considerados, el total de ocurrencias en el *corpus* y el uso.

La lista B (pp.401-435) presenta los 8.428 lemas que han alcanzado menos de tres ocurrencias. Las informaciones que se ofrecen son: categoría gramatical, número de ocurrencias en cada tipo textual y rango de uso.

La totalidad de los lemas, 15.641, aparecen ordenados por uso, y dentro de él alfabéticamente, en la lista C (pp. 436-530). En este caso, junto con la indicación de categoría gramatical, ocurrencias totales, uso y rango, se señala el número de ocurrencias registradas en cada una de las cuatro ciudades.

Las ocurrencias totales de los fonosímbolos se recogen en la lista D (p. 531) y las de las 1.933 expresiones polirremáticas presentes en el *corpus*, junto con la categoría gramatical asociada, se reseñan en la Lista E (pp. 532-540).

Esta distribución del material facilita enormemente su utilización y la hace más fructífera en diversas aplicaciones. Del gran acopio de datos así clasificados se puede obtener un cúmulo de informaciones valiosas referidas a elementos singulares, a áreas léxicas o a categorías morfosintácticas, como por ejemplo, delimitar elementos asociados a la gestión de la comunicación oral, evidenciar tendencias de selección entre sinónimos, detectar homogeneidades o no de uso según la dimensión geográfica y/o según el género de discurso, determinar fenómenos interpretables en clave funcional y semántica, etc. etc.

En suma, como documento objetivo de una realidad lingüística viva y compleja, el LIP constituye un instrumento de consulta imprescindible que, sin duda, redundará en beneficio de la labor investigadora y/o docente de los estudiosos de la lengua italiana, dentro y fuera de Italia.