

RESCATE DE LOS ARCHIVOS SONOROS DEL CENTRO DE LINGÜÍSTICA HISPÁNICA JUAN M. LOPE BLANCH

JULIO CÉSAR SERRANO

El Centro de Lingüística Hispánica Juan M. Lope Blanch (CLH) cuenta con un importante acervo fonográfico que se construyó a partir del trabajo conjunto de investigadores y becarios del Centro en tres grandes proyectos, todos coordinados por Juan M. Lope Blanch: “La norma culta del español de la ciudad de México” (1967-1970), “El habla popular de la ciudad de México” (1972-1975) y “El español del Suroeste de los Estados Unidos” (1985-1986).¹ Mil doscientos sesenta y ocho entrevistas con más de mil trescientos informantes (en muchas de estas participaba más de un informante) fueron grabadas en cintas analógicas que representaban lo mejor en cuanto a grabación magnetofónica en campo en su momento. Si sumamos únicamente los acervos de los proyectos sobre norma culta y habla popular contamos con ochocientas treinta y ocho entrevistas con hablantes de español de la ciudad de México, lo que constituye a la fecha la base de datos de lengua hablada más grande sobre una variante específica del español.² El valor de estos *corpora* para

¹ Estos proyectos han dado lugar a publicaciones importantes (todas en la UNAM) como *El habla de la ciudad de México: materiales para su estudio* (Lope Blanch 1971); *El habla popular de la ciudad de México* (Lope Blanch 1976); *Léxico del habla culta de México* (Lope Blanch [coord.] 1978); *El español hablado en el Suroeste de los Estados Unidos: materiales para su estudio* (Lope Blanch 1990); *El estudio del español hablado culto: historia de un proyecto* (Lope Blanch 1986), por mencionar sólo algunas.

² Macroproyectos como el “Proyecto para el Estudio Sociolingüístico del Español de España y América” (PRESEEA, cf. Moreno Fernández 2003) toman en cuenta corpus

la investigación lingüística y en distintos campos de las humanidades es incalculable y deben ponerse al día valiéndonos de los recursos tecnológicos e informáticos con que contamos actualmente.

Importancia de la digitalización de archivos sonoros

El actual proceso de digitalización, que se inició en junio de 2007, contempla la transferencia del acervo del CLH a formato .wav, lo que posibilita (cuando la calidad de la cinta original lo permite) realizar estudios fonéticos detallados, incluso a través de programas de cómputo para el análisis acústico.

Entre las ventajas que conlleva contar con un acervo sonoro digital se encuentran las siguientes:

- Los dispositivos magnetofónicos analógicos como la cinta de carrete (en el que se realizaron las grabaciones originales) tienen, en buenas condiciones de almacenamiento, una vida útil aproximada de 40 años; un archivo digital puede tener una vida bastante más prolongada sobre todo si consideramos que los archivos digitales son copias fieles de los originales y que pueden almacenarse en muy diversos dispositivos de memoria (discos rígidos, DVD, CD, incluso Internet).
- El almacenamiento de la información de audio digital ocupa muchísimo menos espacio en las fonotecas que las cintas analógicas (de hecho, en un solo disco rígido de gran capacidad puede almacenarse todo el acervo sonoro del CLH, que se encuentra actualmente en 908 cintas de carrete abierto).
- Con una grabación digital adecuada, los materiales del IIFL son susceptibles de analizarse con los instrumentos informáticos más actuales para el análisis acústico del habla, como PRAAT (Boersma y Weenink 2008), algo que hemos podido constatar en trabajos recientes (Serrano 2008).

importantes en numerosas ciudades del mundo hispánico; se plantea un mínimo de encuestas que va de 54 hasta 108 por ciudad, dependiendo de su densidad poblacional.

- Los materiales analógicos pueden perder calidad al copiarse en otros formatos analógicos: esto no sucede cuando se realiza una copia digital de otro archivo digital. Además, la copia y la reproducción de este tipo de archivos es mucho más sencilla y rápida que la de los materiales analógicos. Por último, la tecnología para la reproducción y copia digitales tiene la ventaja de estar (relativamente) al alcance y el manejo de todos.
- Los materiales de audio digital pueden compartirse fácilmente a través de medios electrónicos como las redes de cómputo locales o internet.

Con todo lo anterior no debe entenderse que no debemos preservar los formatos analógicos o que la tecnología digital no presente sus limitaciones: debemos considerar que cuando se digitaliza una señal de audio (esto es, cuando es trasladada a un código binario) no tenemos la gama completa de frecuencias sonoras, sino una muestra de todas estas frecuencias, con límites precisos; la diferencia para quien escucha se encuentra en dos aspectos básicos: una mayor potencia y menos ruido en la señal digital que en la analógica; sin embargo, la riqueza sonora sigue siendo mayor en esta última. También es cierto que la tecnología digital avanza rápidamente y seguramente llegará el momento en que los formatos de soporte que usamos actualmente, como el *Compact Disc*, serán sustituidos por otros más pequeños y con mayor capacidad de almacenamiento. Además, es muy probable que la tecnología para la grabación digital también se modifique y que tengamos tipos de archivo que deberán en su momento "actualizarse" en otros formatos.

Características de los archivos sonoros originales del IIFL

Todas las entrevistas están grabadas en cinta magnetofónica analógica de un cuarto de pulgada de ancho (6.35 mm) en carretes de 5 pulgadas de diámetro, de media hora de duración por cada lado a una velocidad de $3\frac{3}{4}$ "/seg., en formato monoaural. Tanto el entrevistador como el infor-

mente se escuchan en un solo canal, aunque en la mayoría de las entrevistas el micrófono se colocó más cerca del informante.³ Con estas características, las grabaciones registran de manera eficiente las frecuencias que van de los 20 a los 15,000 Hz, rango más que suficiente para el registro de la voz humana. Todas las grabaciones presentan el ruido de fondo de altas frecuencias (o *hiss*) que es característico de las grabaciones analógicas. Este ruido es eliminado por el proceso de digitalización, con el que no se capturan frecuencias tan altas.

Respecto a la duración de las entrevistas, doscientas noventa y siete (el 57%) del corpus del proyecto "Norma culta" tiene una duración de una hora, y el resto son entrevistas de 30 minutos. Los proyectos subsiguientes ("Habla popular de la ciudad de México" y "Español del Suroeste de los Estados Unidos") cuentan mayormente con entrevistas de una hora de duración.

Cuadro 1. Acervo sonoro del Centro de Lingüística Hispánica

	Número de entrevistas	Número de informantes	Número de cintas
Norma culta	518	598	412
Habla popular	220	220	146
Suroeste de EUA	430	430	350
TOTALES	1168	1248	908

El muestreo de Norma culta estaba organizado de acuerdo con los siguientes criterios sociodemográficos: hombres y mujeres en un 50%; tres grupos de edad: jóvenes (25-35 años), 30%; adultos (36-55 años), 45%; mayores (56 o más), 25%. El muestreo para Habla popular sigue muy de cerca estos parámetros: hombres y mujeres distribuidos por igual en un 50%; grupos de edad: jóvenes 18-34 años, adultos 35-54 años, mayores de 55 años. Para la muestra sobre el español del Suroeste de los Estados Unidos el criterio utilizado fue que el informante hubiese nacido en la localidad y formara parte de la tercera generación de inmigrantes del lugar.

³ En el Apéndice se muestran fotografías del equipo utilizado.

Detalles técnicos del proceso de digitalización

Cada cinta original es reproducida en una grabadora-reproductora que envía la señal monoaural, en tiempo real, a una tarjeta de sonido instalada en una computadora; la conexión se realiza con cables DIN-RCA. El programa utilizado es WaveStudio de la compañía Creative. Los archivos digitales tienen una frecuencia de muestreo de 44.1 Khz, y la codificación es PCM con una profundidad de 16 bits, que son estándares internacionales para la documentación lingüística.

Los archivos digitales reciben el nombre de la cinta de la que se obtienen, seguido del nombre del informante y la fecha de la grabación. De cada grabación se crea una copia de respaldo que se almacena en un disco rígido externo.

El tiempo estimado para copiar cada entrevista es del doble o triple de ésta, puesto que debe hacerse una escucha de la grabación, monitorearla constantemente en el tiempo real, y hacer una última audición para revisión y anotar los detalles más sobresalientes de la encuesta: si existen ruidos de ambiente, ruidos ocasionados por los aparatos utilizados, intervención de otros informantes, características del discurso (si es fluido o poco natural), distancia del informante respecto al micrófono, etc. Debe aclararse que esta copia se realiza sin ninguna manipulación del archivo sonoro original. Se trata de obtener la grabación tal como fue realizada y no perder información acústica en el proceso de digitalización. En general, las cintas trabajadas hasta el momento tienen una excelente calidad y no requieren de un procedimiento especial de "limpieza" de la señal para hacer análisis fonético detallado. En todo caso, los programas de cómputo de tratamiento de audio pueden modificar la señal de los archivos que interesen, pero este trabajo deberá realizarse de acuerdo con las necesidades específicas de las investigaciones que vayan surgiendo.

Otros procesos

Son varios los procesos paralelos a la digitalización en sí. Se lleva una bitácora del laboratorio donde se anotan los equipos utilizados y las redes de

conexión de los aparatos, el programa de cómputo y los parámetros de digitalización, entre otros detalles. También se añade la información sobre cada cinta y su contenido en la base de datos diseñada para el acervo.⁴ Ésta consta de 36 campos y contiene información valiosa sobre el informante (sexo, edad, estudios, ocupación), sus experiencias en el extranjero, idiomas, tiempo de radicar en la ciudad, etc., además de datos sobre sus padres y cónyuge, lo que permite trazar un perfil sociolingüístico muy detallado de cada uno. La base de datos permite hacer búsquedas en cualquiera de los campos, lo que resulta muy valioso para los investigadores. Por ejemplo, si una investigación se realiza por tipo discursivo pueda seleccionarse en el campo "Tipo de entrevista" la opción que más interese: diálogo dirigido, conversación entre el entrevistador y el informante, conferencia, monólogo, grabación secreta, etcétera.

Aprovechamiento de los acervos sonoros para la investigación

El valor de los *corpora* de lengua oral del CIH radica en ciertas características de su diseño y recolección. Primero, que fueron pensados para fines propiamente lingüísticos y por lo tanto se buscó la mejor calidad de grabación posible. El equipo utilizado permitió la obtención de grabaciones muy bien logradas a pesar de ser hechas en el campo, esto es, en el lugar de trabajo o la casa de los entrevistados. El tiempo de cada entrevista tiene un mínimo de 30 minutos y se extienden hasta una hora, suficiente para realizar análisis de gran variedad de fenómenos fonéticos o morfosintácticos. Finalmente, las grabaciones para los proyectos sobre Norma culta y Habla popular fueron recogidas hace 35 o 40 años y reflejan fielmente la realidad sociolingüística de la capital mexicana en dos estratos socioculturales bien diferenciados.

Con estas grabaciones pueden realizarse estudios de sociolingüística urbana del español capitalino en *tiempo real*, esto es, se puede trabajar con muestras de habla de la misma comunidad pero levantadas en épocas dis-

⁴ El ingeniero Julio Pérez (departamento de cómputo del IIFL) creó esta base de datos en Lotus Approach.

tintas (cf. Labov 1994; Trudgill 1988; Bailey 2002). Un trabajo reciente es el de Serrano (2008), donde se analiza la asibilación de las vibrantes (*r*, *rr*) en dos muestras: la primera obtenida de las entrevistas de Norma culta y Habla popular (1967-1972) y la segunda de encuestas del proyecto "Cambio y variación en la ciudad de México" (recogidas en El Colegio de México entre 1997 y 2002). Asimismo, el autor proyecta estudiar la evolución del debilitamiento vocálico (cf. Lope Blanch 1983; Serrano 2006) y un estudio sobre disponibilidad léxica en ambas muestras y determinar cómo ha evolucionado el repertorio léxico en esta variedad; la evolución sociolingüística de la entonación circunfleja (Martín 2004) también puede abordarse con este corpus. Por otra parte, las encuestas sobre el español del Suroeste de los Estados Unidos ofrecen una veta importantísima para estudiar una multiplicidad de procesos asociados con el bilingüismo (*code-switching*, desplazamiento del español, préstamos fónicos y léxicos, actitudes comunitarias ante el bilingüismo inglés-español, etc.). Un nuevo levantamiento de datos en las ciudades donde se realizaron las encuestas originales podría arrojar luz sobre los procesos de desplazamiento y retención del español en esta región estadounidense. Incluso podría explorarse el impacto del flujo migratorio de los últimos 20 años en esas ciudades.

Estos materiales tienen una importancia lingüística que rebasa obviamente el nivel fonológico. Existen en español muchas variables susceptibles de analizarse en tiempo real que se ubican en distintos niveles de análisis lingüístico. Posibles investigaciones que podrían abordarse son: la evolución del sentido epistémico de la construcción *deber de* a uno de tipo deóntico (Vázquez Laslop, 2001); el cambio (al parecer consolidado) de la concordancia de persona en el verbo *haber* impersonal (*hubo personas* contra *hubieron personas*); la evolución de ciertos ítems léxicos como *qué onda* y todas sus derivaciones fraseológicas; el rastreo de expresiones en (posible) desuso como *a lo macho* o *¿ves?*, entre otros.

Por supuesto, otras disciplinas sociales como la sociología, la antropología o la historia pueden valerse de estos archivos sonoros como fuentes de información para la historia oral, la microhistoria, el estudio del cambio social, de los efectos económicos y sociales de la urbanización y la migración, etcétera.

Bibliografía

- BAILEY, Guy (2002). "Real and apparent time", en *The Handbook of Variation and Change*. J. K. Chambers, Peter Trudgill, y Natalie Schilling-Estes (eds.). Oxford, Blackwell, pp. 312-332.
- BOERSMA, Paul, y David WEENINK (2008). *PRAAT. Doing Phonetics by Computer*. Universidad de Ámsterdam. <http://www.fon.hum.uva.nl/praat/>.
- LABOV, William (1994). *Principles of Linguistic Change*, vol. I: *Internal Factors*. Oxford, Blackwell.
- LOPE BLANCH, Juan M. (1983). "En torno a las vocales caedizas del español mexicano", en *Estudios sobre el español de México*. México, UNAM.
- MARTÍN BUTRAGUEÑO, Pedro (2004). "Configuraciones circunflejas en la entonación del español mexicano", *Revista de Filología Española*, 84, pp. 347-373.
- MORENO FERNÁNDEZ, FRANCISCO (2003). Metodología del "Proyecto para el estudio sociolingüístico del español de España y de América" (PRESEEA). Versión revisada en octubre de 2003. Descargable en internet en: <http://www.linguas.net/default.aspx?alias=www.linguas.net/portalpreseea>
- SERRANO, Julio César (2006). "En torno a las vocales caedizas del español mexicano: una aproximación sociolingüística", en *Los líderes lingüísticos*. Pedro Martín Butragueño (ed.). México, El Colegio de México, pp. 37-59.
- (2008). "Vibrantes asibiladas en español de la ciudad de México (1964-1972)", en *Patrones fónicos y variación*. Pedro Martín Butragueño y Esther Herrera Z. (eds.). México, El Colegio de México.
- TRUDGILL, Peter (1988). "Norwich revisited: recent changes in an English urban dialect", *English World-Wide*, 9 (1), pp. 33-49.
- VÁZQUEZ LASLOP, María Eugenia (2001). *La arquitectura lingüística del compromiso*. México, El Colegio de México.

Apéndice



Foto 1. Grabadora Uher 'Royal de Luxe', utilizada para el trabajo de campo en los años 1967-1972



Foto 2. Micrófonos Uher utilizados durante el levantamiento de encuestas (1967-1972)

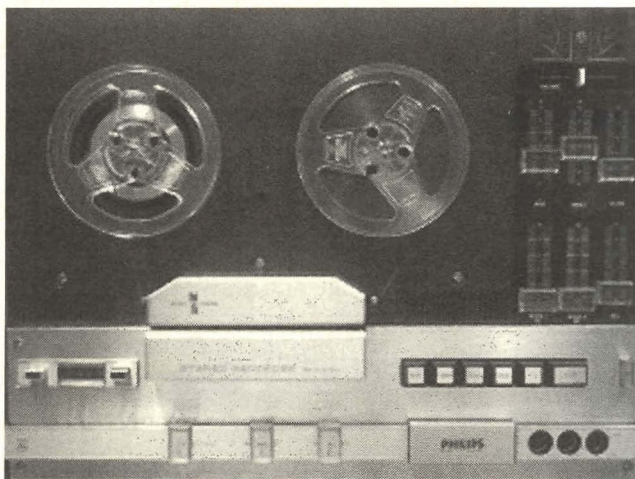


Foto 3. Grabadora Philips 'N4416', que se utiliza actualmente para realizar la digitalización