

**EL PROBLEMA DE LA REPRESENTATIVIDAD DE LAS
MUESTRAS EN LA INVESTIGACIÓN
FONÉTICA EXPERIMENTAL**

**THE PROBLEM OF SAMPLES REPRESENTATIVITY IN
EXPERIMENTAL PHONETICS RESEARCH**

DANIEL IGNACIO PEREIRA
Universidad de Concepción, Chile
danielpereira@udec.cl

Artículo recibido el día: 9/12/2010
Artículo aceptado definitivamente el día: 14/09/2011
Estudios de Fonética Experimental, ISSN 1575-5533, XX, 2011, pp. 57-69

RESUMEN

El presente artículo es una reflexión acerca de un problema metodológico de la fonética experimental: la selección y representatividad de las muestras usadas en la investigación acústica y perceptiva. Se hace la distinción entre *datos del corpus*, *informantes* y *sujetos* de la investigación. En fonética acústica se trabaja con corpora de datos obtenidos de informantes, mientras que en fonética perceptiva, se analizan los juicios de los sujetos sobre ciertos estímulos. De acuerdo con la bibliografía específica y en consonancia con los trabajos revisados, se llega a la conclusión de que no existe acuerdo entre los investigadores sobre este punto y observa una preferencia de estos por las muestras pequeñas, por lo que se propone relativizar este concepto y ajustarlo al de homogeneidad de rasgos.

Palabras clave: *muestra de datos, informantes, sujetos, representatividad, fonética.*

ABSTRACT

The following article is an afterthought about a methodological problem of the experimental phonetics: the selection and representativeness of the samples used in the acoustic and perceptive investigation. The distinction here among *data of the corpus*, *informants* and *subjects* of the investigation is done. Acoustic phonetics deals with corpora of data obtained from informers, while in perceptive phonetics the judgments of the subjects about certain reactive patterns are analyzed. According to the specific bibliography and regarding to the works reviewed, it is therefore concluded that there is not agreement among authors on this matter and a preference towards small samples is observed, thus a relativization of this concept is proposed plus an adjustment to the homogenization of patterns.

Keywords: *data sample, informants, subjects, representativeness, phonetics.*

1. INTRODUCCIÓN

Escribir acerca de problema metodológicos no es tarea fácil, pues implica poseer una idea fundamentada y sólida acerca de nociones mucho más profundas que el *hacer científico*, nociones relacionadas con la concepción misma de la ciencia y, por lo tanto, nociones sobre epistemología. Finalmente, es preciso tener una idea acerca de la *verdad* o de los tipos de verdades que podemos vislumbrar con los criterios científicos. Este trabajo está lejos de eso. El objetivo que se persigue es hacer una reflexión acerca de un problema bastante acotado -y polémico- del proceso experimental en las investigaciones fonéticas de tipo acústico: el tamaño y representatividad de la muestra.

A pesar de lo acotado y circunscrito que podría parecer, no ha sido fácil aunar criterios en torno a algo tan controversial. Algunos piensan que la proporción de la muestra con respecto a su población debe estar dada por procedimientos estadísticos, basada en los conceptos de *representatividad* y de *azar*; otros, en cambio, defienden la postura de que proceder de ese modo es absurdo en fonética, disciplina en la que los informantes no son los datos en sí mismos, sino sólo la fuente, y que, por lo tanto, las variables se pueden controlar si se tiene una muestra homogénea de informantes, aunque no necesariamente representativa. Por último, hay quienes trabajan con un solo informante y de él obtienen todos los datos requeridos para el análisis.

Nuestra tesis es que en fonética, por lo menos en la variante acústica, la representatividad del tamaño de la muestra de informantes es relativa a las necesidades de la investigación y debe estar fundamentada en la homogeneidad de los rasgos pertinentes para el estudio y no necesariamente en el tamaño poblacional. Es complicado establecer una ecuación o un método que se ajuste a todos los requerimientos de las disciplinas que dialogan con la fonética acústica y, al parecer, tampoco se puede hallar alguno más recomendado y que se ajuste a todos los experimentos del área.

En la primera parte, se explicará la naturaleza de los datos fonéticos, determinando los tipos de datos de acuerdo a determinadas clases de investigaciones, dado el complejo funcionamiento de los sonidos del lenguaje; en el segundo apartado, se presentarán los distintos tipos de muestra recomendados y utilizados -he aquí una sutil, pero fundamental diferencia- en fonética acústica; por último, en las conclusiones se expondrán las reflexiones suscitadas por las lecturas realizadas en relación al problema central de este trabajo.

2. LA NATURALEZA DE LOS DATOS FONÉTICOS

El objeto de estudio de la fonética son los sonidos del habla, entendiéndose *sonido* como un fenómeno consistente en la propagación de ciertas vibraciones por el aire, perceptibles por el oído. Por lo tanto, estamos frente a un hecho físico. Si a eso le agregamos *del habla*, debemos decir que dichos hechos físicos son provocados intencionadamente por los seres humanos, mediante la puesta en funcionamiento de una serie de órganos fonatorios, para establecer comunicación con sus semejantes (Cf. Martínez Celdrán 1996).

Ahora bien, la fonética divide su estudio en tres grandes ramas: la fonética articuladora, que estudia la manera en que son producidos los sonidos por los órganos fonatorios; la fonética acústica, que se dedica a analizar las propiedades físicas de dichos sonidos en su paso de la boca al oído; y la fonética perceptiva, que intenta describir las percepciones de los sonidos y su interpretación por parte del cerebro (trayecto oído-cerebro) para proponer un modelo de percepción del habla.

Si analizamos con cuidado lo dicho hasta aquí, nos daremos cuenta de que la fonética es, dentro de las disciplinas lingüísticas, la más cercana a las ciencias naturales, pues opera con datos físicos procedentes de un organismo vivo que tiene una determinada constitución fisiológica. Es en esta peculiar combinación de factores físico-fisiológicos donde se origina la variabilidad en la producción del habla. Y un dato más peculiar incluso: cuando un individuo habla su lengua en el seno de su comunidad lingüística, se comunica, a pesar de las diferencias físicas de su producción acústica, ocasionadas por la morfología de sus órganos fonoarticulatorios.

Como se ve, cada una de las subdivisiones de la fonética tiene bastante delimitada su parcela de estudio. Nosotros nos centraremos en el sonido como hecho físico, por lo tanto, nuestro interés central estará puesto en la fonética acústica aplicada a la lingüística.

Desde esta perspectiva, el sonido puede ser analizado en todos o algunos de los rasgos que lo caracterizan como una onda sonora: amplitud o intensidad de la onda (medida de su potencia en decibeles), frecuencia de onda (medida de la cantidad de ciclos por segundos), relación entre los armónicos y la frecuencia fundamental; y duración absoluta o relativa de un determinado sonido (medida en segundos o milésimas de segundo).

Ahora bien, el estudio del sonido desde el punto de vista físico, implica que el analista debe conocer, por un lado, las características físicas del fenómeno y, por otro, los instrumentos que debe utilizar para tomar muestras y realizar los análisis correspondientes para comprobar su hipótesis, de acuerdo a la metodología experimental de la lingüística, la que, según Solé (1981) consiste en

observar un cierto fenómeno lingüístico, hipotetizar sobre las posibles causas de este fenómeno, encontrar algunos puntos en que hipótesis o explicaciones alternativas hagan predicciones diferentes, y diseñar experimentos para contrastar estas predicciones en los mismos hablantes de la lengua. (1981: 4-5)

En las ciencias, los datos de los fenómenos que se desea explicar deben ser recogidos cuidadosamente de la realidad empírica, de acuerdo a ciertos procedimientos o protocolos delimitados por el mismo científico o por la ciencia en cuestión (Cf. López Morales 1994: 59). Y la fonética acústica no escapa a tal condición, tal como lo afirma Llisterra (1991: 60): *en la mayoría de los trabajos de fonética es el mismo investigador quien debe recogerlos [los datos], acudiendo a hablantes que cumplan los requisitos establecidos en su diseño experimental.*

3. ALGUNOS MÉTODOS DE MUESTREO UTILIZADOS EN FONÉTICA

Atendiendo a lo dicho anteriormente, el fonetista acústico debe encontrar los métodos más adecuados para conseguir una muestra del fenómeno estudiado, de acuerdo a su diseño experimental y a los objetivos que persigue.

Pero, ¿qué es una muestra? Tal como la definen Hernández, Fernández y Baptista (2006: 236), la muestra *es un subgrupo de la población de interés (sobre el cual se recolectarán datos, y que tiene que definirse o delimitarse de antemano con precisión), éste deberá ser representativo de la población.* En las Ciencias Sociales y, por ende, en la lingüística, para determinar y asegurar el grado de representatividad de la muestra se emplean con frecuencia métodos estadísticos, en especial en los estudios sociolingüísticos (Cf. Moreno Fernández 1990; López Morales 1994). Estas problemáticas serán discutidas en el apartado siguiente.

Aun cuando existen diversos métodos para la extracción de muestras en los manuales metodológicos generales y de lingüística, presentamos a continuación algunos de ellos, presentes en las investigaciones de fonética acústica revisadas para el caso.

3.1. Muestras probabilísticas

Es este tipo de muestras *se parte del principio de que todos y cada uno de los componentes de la población tienen alguna probabilidad de ser seleccionados para formar parte de la muestra*. (Moreno Fernández 1990: 82; cf. Hernández, Fernández y Baptista 2006: 241). Dentro de este tipo de muestras se explora el azar. Llisterri (1991: 102-105) propone tres métodos de azar: el *sampling frame* (en el que la muestra se toma a partir de un listado completo de la población), el muestreo estratificado (en el que del listado se elige un individuo cada cierto cantidad hasta completar el número deseado) y el muestreo por bloque (donde escogen los individuos a partir del primero y completando la cantidad deseada para la muestra, sin saltarse ningún sujeto).

También existen las muestras probabilísticas estratificadas, en las que los investigadores *comparan sus resultados entre segmentos, grupos o nichos de la población* (Hernández, Fernández y Baptista 2006: 247). Esto requiere del *conocimiento de las variables que van a intervenir en el estudio* (Llisterri 1991: 103), con el objetivo de delimitar con exactitud las características que deben tener los individuos seleccionados para la conformación de los estratos.

Por último, cabe mencionar las muestras por racimos, en la que se estudian grupos ya conformados por otras razones, tales como segmentos sociales, clases de profesionales, etc. (Cf. Hernández, Fernández y Baptista 2006: 249; Moreno Fernández 1990: 87).

Aunque no son las más utilizadas en la fonética acústica aplicada a la lingüística, existen investigaciones que han empleado métodos de azar para conseguir sus muestras. Por ejemplo, García Riverón (1996) en un estudio sobre la entonación del español de Cuba, obtuvo grabaciones en terreno de 124 sujetos de un total poblacional de 1.941.854, correspondiente a la cantidad total de habitantes de la Ciudad de La Habana (Cuba), que equivale al 0,01% del universo. Los sujetos no fueron seleccionados directamente, pero sí los centros laborales de donde se obtendría la muestra. Como se ve, la investigadora utilizó un método por racimo, en la que eligió los grupos (racimo), pero no los individuos de esos grupos.

3.2. Muestras no probabilísticas

En este tipo de muestras, *la elección de los elementos no depende de la probabilidad, sino de causas relacionadas con las características de la investigación o*

de quien hace la muestra (Hernández, Fernández y Baptista 2006: 241). Moreno Fernández (1990: 87-88) enumera tres tipos de muestras no probabilísticas: la accidental, en la que se utilizan individuos que se cruzan en el camino, sin tener cuidado en sus características; la muestra por cuotas, que parte de una estratificación de la población de acuerdo a ciertas características y luego precisa buscar sujetos que se adscriban a los estratos identificados; por último, el muestreo intencionado, *se basa en el juicio del investigador para seleccionar los individuos que deben aparecer en la muestra* (Moreno Fernández 1990: 88).

En fonética acústica, debido a su fuerte rasgo instrumental, muchos de estos procedimientos muestrales se hacen impracticables, pues a menudo es preciso que los sujetos se trasladen a lugares especialmente acondicionados para tomar los datos, específicamente, a laboratorios. Además, tomar una muestra probabilística dependiente de una población extensa (el habla de una zona geográfica amplia, un país, por ejemplo), implica un enorme esfuerzo económico y temporal del que muchas veces no se dispone o no es razonable disponer¹. Por esta razón, en las investigaciones acústicas aplicadas a la lingüística es común encontrar otro tipo de muestras, las de voluntarios, que son frecuentes en las ciencias sociales y en las ciencias que estudian la conducta (medicina, psicología, sociología, por ejemplo). Se trata de

muestras fortuitas (...) donde el investigador elabora conclusiones sobre casos que llegan a sus manos de forma casual. (...). En estos casos, la elección de los participantes depende de circunstancias muy variadas. Este tipo de muestras se usa en estudios experimentales de laboratorio (...) (Hernández, Fernández y Baptista 2006: 565-566).

Este último tipo de muestras corresponde a investigaciones cualitativas, de acuerdo a la catalogación que establecen Hernández, Fernández y Baptista (2006), lo que es discutible, debido al marcado rasgo experimental y cuantificable que ostentan los datos.

López Morales (1994) describe un tipo de muestra que podríamos propugnar como la utilizada en gran parte de los estudios fonético-acústicos. Se trata de las llamadas muestras textuales, en las que *la unidad de muestreo no es un sujeto, sino un texto*

¹ Considérense las cláusulas de viabilidad e impacto de muchos fondos concursables, que no financian investigaciones cuya inversión inicial no implique réditos transferibles a la industria o a otras áreas de conocimiento.

(p. 73), entendiendo *texto* como *una porción de lengua grabada para ser analizada y descrita* (Crystal 2000: 547). Es decir, el N muestral no debe determinarse necesariamente sobre los sujetos, sino sobre las producciones orales de dichos sujetos. Las muestras textuales pueden ser aleatorias – y entonces se someten a cualquiera de las técnicas de azar antes descritas- o empíricas. Estas últimas suelen estar determinadas por los criterios impuestos tanto por la investigación misma como por los del investigador.

4. EL TAMAÑO DE LA MUESTRA

A la luz de lo expuesto en el anterior acápite, el concepto de *representatividad* de la muestra debe ser relativizado o ponderado en su justa medida, incluso a pesar de lo que afirma Moreno Fernández (1990: 44):

la fiabilidad se pone en duda porque los datos cosechados y las personas que los han proporcionado no son representativos. La estadística moderna y las ciencias que la aplican no conceden fiabilidad a aquellos estudios que, trabajando con fenómenos sociales, no se realizan sobre muestras de poblaciones suficientemente representativas.

Si consideramos que las muestras fonético-acústicas son los datos reales y numéricamente medibles de las producciones de los sujetos, la cantidad de individuos que participen en la investigación no es necesariamente determinante para validarla o invalidarla. Sin embargo, lo que no se debe abandonar al azar es la selección de los informantes, la que debe atender estrictamente a los requerimientos de la investigación (Llisterri 1991: 95ss; Moreno Fernández 1990: 78). Es decir, es preciso cuidar ciertas variables procedentes de las experiencias o características de los sujetos de la muestra que podrían modificar los resultados o invalidarlos definitivamente; variable tales como el sexo, la edad, el estrato social, la procedencia, la permanencia en el lugar, las lenguas que habla, etc. Lo mismo dígame para el número de informantes necesarios para obtener los datos. A propósito de esto último, a menudo se escucha a algunos investigadores, sobre todo a aquellos procedentes de otras áreas de la lingüística o de disciplinas afines, criticando la escasa representatividad de los informantes de algunos trabajos, debido a que han incluido pocos informantes. Al respecto, es necesario hacer una distinción conceptual entre informante y sujeto. Llisterri (1991: 48) la establece con eminente claridad:

En el caso de un análisis de tipo articulatorio o acústico, los datos constituyen un corpus que es realizado -es decir leído, repetido o pronunciado espontáneamente- por un grupo de hablantes a los que llamaremos informantes. En cambio, si se trata de un estudio auditivo-perceptivo, nuestros datos vienen dados por las respuestas de un conjunto de personas a las que podemos denominar sujetos del experimento.

En otras palabras, la muestra de datos es el corpus recogido y no de donde se obtuvieron esos datos.

Muchas críticas realizadas al respecto no consideran esta distinción y se equivocan al generalizar a todos los ámbitos de la fonética lo que le corresponde sólo a algunos. En efecto, las investigaciones acústicas, por ejemplo, en las que se analizan datos referentes a frecuencias (fundamental y formantes), intensidad y duración (vocálica, consonántica, de pausas, de enunciado), se prefiere trabajar con pocos informantes, debido a la cantidad enorme de datos que es posible obtener de sólo unos minutos de grabación, por cada uno de los informantes. De ahí, la rigurosidad que aplican los fonetistas al momento de seleccionar a quienes colaborarán en calidad de informantes, aplicando pautas de información relativa a los individuos, su comportamiento lingüístico y el de su entorno, su movilidad geográfica, etc., como ya hemos mencionado (Cf. Mettas 1979: 32-65; Llisterri 1991: 96).

Desde esta perspectiva, cualquier individuo que sea proficiente en su lengua será representativo en cuanto a los datos que de él podamos obtener. La rigurosidad en la selección del informante se justifica sólo en el mayor o menor control de las variables que sean pertinentes en la investigación.

El investigador debe controlar en qué momento empieza a aparecer un comportamiento regular en los hablantes, y debe saber también que, mucho más importante que la cantidad de la muestra es su calidad. Cuanto más dispersa sea ésta, un número mayor de hablantes necesitará para que se manifiesten tendencias claras en los fenómenos; en cambio, cuanto mayor sea el grado de homogeneidad, más rápidamente observará el surgimiento de algún tipo de regularidad. (Llisterri 1991: 106).

Este concepto de *homogeneidad* del cual se habla en la cita de Llisterri es el que fundamenta la *representatividad* de una muestra en estudios de este tipo: si se han seleccionado los rasgos pertinentes adecuadamente y con rigurosidad y si existe

uniformidad entre ellos, los resultados de una investigación pueden ser extrapolables a grupos más grandes (Cf. Castro y Castro 2001: 179).

Diferente es el caso en fonética auditiva o perceptual, específicamente de aquella que analiza juicios perceptuales. En este tipo de estudios no se requieren informantes, sino sujetos que den un juicio de expertos sobre los fenómenos sobre los cuales trate la investigación. Los datos ya no son aquí las manifestaciones acústicas (físicas) o articulatorias (fisiológicas) de determinada producción de habla, sino las representaciones psíquicas que dichas manifestaciones suscitan en los sujetos. En estos estudios es deseable un N muestral definido estadísticamente o suficientemente representativo de una población, pues lo que se mide son tendencias en la interpretación, siempre más subjetivas que los datos puramente físicos.

5. CONCLUSIONES

A pesar de que los métodos estadísticos para obtener muestras representativas de una población nos ofrecen una alternativa confiable y válida, es preciso señalar que su incorporación a las investigaciones dependerá de los distintos diseños requeridos para estudiar determinados fenómenos y de los objetivos que busque el estudio.

Sin embargo, no hay que confundir los datos con su fuente. Como hemos visto, en fonética acústica aplicada a la lingüística, los datos se obtienen de las producciones de los individuos y no equivalen a los individuos que los proporcionan: de ahí que hayamos postulado que las muestras más usadas en estas investigaciones son las llamadas textuales empíricas, en las que el investigador opera sobre las grabaciones de las producciones de unos sujetos seleccionados de acuerdo a criterios concordantes con la investigación que realiza o de los casos fortuitos que tiene a su alcance.

Tampoco hay que confundir los conceptos de informante y sujeto de una investigación. Los informantes son justamente fuentes de los datos que se constituyen como objeto de la investigación, mientras que los sujetos son los datos mismos, en cuanto que sobre ellos opera el proceso investigativo.

Por último, cabe destacar que el concepto de representatividad asociado al tamaño de la población debe ser relativizado. En otras disciplinas, como la medicina, muchas veces se trabaja con las muestras que esporádicamente llegan a manos de

los especialistas. Enfermedades rarísimas o nuevas construyen su descripción sobre datos exiguos proporcionados solo por un individuo o por casos aislados, sin que por ello se dé lugar a discusiones sobre la representatividad de dicha descripción, bajo el supuesto de que lo que le ocurre a un miembro de la especie, potencialmente podría ocurrirle a cualquiera.

En el plano de la lingüística, a veces el investigador se ve en la aporía de tener sólo un informante, por ejemplo, en el caso de las lenguas a punto de desaparecer, sin que podamos argüir que sus datos no son representativos. La representatividad en fonética, entonces, se relaciona más con la consistencia interna de los métodos aplicados en la investigación que con la cantidad de informantes participantes en el estudio, aunque la conformación muestral debe ser homogénea, para que los resultados obtenidos sean generalizables. De todos modos, esta hipótesis sobre la relativización del concepto de representatividad muestral basada en la cantidad de individuos de la población y la defensa de una representatividad basada en la homogeneidad de la muestra debe ser demostrada (o refutada) experimentalmente para verificar su validez.

AGRADECIMIENTOS: Este trabajo se ha beneficiado de una beca de la Comisión Nacional de Investigación Científica y Tecnológica (CONICYT), del Gobierno de Chile.

6. REFERENCIAS BIBLIOGRÁFICAS

- CASTRO NOGUEIRA, M. Á. y L. CASTRO NOGUEIRA. (2001): «Cuestiones de metodología cualitativa», *EMPIRIA. Revista de Metodología de Ciencias Sociales*, 4, pp. 165-190.
- CRYSTAL, D. (2000): *Diccionario de lingüística y fonética*, Barcelona, Octaedro.
- GARCÍA RIVERÓN, R. (1996): *Aspectos de la entonación hispánica I. Metodología*, Cáceres, Universidad de Extremadura.
- HERNÁNDEZ, R.; C. FERNÁNDEZ y P. BAPTISTA. (2006): *Metodología de la investigación*, México, McGraw-Hill.
- LÓPEZ MORALES, H. (1994): *Métodos de investigación lingüística*, Salamanca, Ediciones Colegio de España.

-
- LLISTERRI, J. (1991): *Introducción a la fonética: el método experimental*, Barcelona, Anthropos.
- MARTÍNEZ CELDRÁN, E. (1996): *El sonido en la comunicación humana. Introducción a la fonética*, Barcelona, Octaedro.
- METTAS, O. (1979): *La prononciation parisienne. Aspects phonétiques d'un sociolecte parisien (Du Faubourg Saint-Germain a la Muette)*, Paris, Selaf.
- MORENO FERNÁNDEZ, F. (1990): *Metodología sociolingüística*, Madrid, Gredos.
- SOLÉ, M. J. (1981): «La experimentación en fonética y fonología», *Estudios de Fonética Experimental*, I, pp. 1-70.