

# A enquisa na investigación sociolingüística

**Fernando Ramallo**  
*Universidade de Vigo*

Departamento de Traducción e Lingüística  
Facultade de Filoloxía e Traducción  
Universidade de Vigo  
As Lagoas-Marcosende, s/n  
E-36200, Vigo, SPAIN  
framallo@uvigo.es

**Anxo M. Lorenzo**  
*Universidade de Vigo*

Departamento de Filoloxía Galega e Latina  
Facultade de Filoloxía e Traducción  
Universidade de Vigo  
As Lagoas-Marcosende, s/n  
E-36200, Vigo, SPAIN  
alorenzo@uvigo.es

Title: **The survey in the sociolinguistic research**

## **Abstract**

The aim of this paper is to present a description of the statistic survey since it constitutes an instrument traditionally related to the quantitative research and orientated towards social structure observation and analysis. The search for social object quantification has always been one of the most outstanding tendencies in the elaboration process of scientific knowledge, mainly from positivist theoretical perspectives. In some areas of sociolinguistic research the pertinence of surveys is obvious. After defining and contextualising surveys, we present a typology and a section about the questionnaire, which is the means of communication between the researcher and the analysed population. Then, we pay attention to the selection of observation unities and refer to the most important features of random and non-random samples. Finally, we offer a specific section about data treatment and we mention those most relevant theoretical aspects of descriptive and inferential statistics.

**Key words:** survey, methodology, sociolinguistics, quantitative methods.

## Resumo

O obxectivo do traballo é presentar unha descrición da enquisa estatística xa que constitúe un instrumento tradicionalmente vencellado á investigación cuantitativa orientada á observación e á análise da estrutura social. A procura da cuantificación dos obxectos sociais segue sendo unha das tendencias máis salientables no proceso de elaboración de coñecemento científico, en especial desde as perspectivas teóricas positivistas. E en determinadas áreas da investigación sociolingüística a pertinencia da enquisa é indubidable. Trala definición e contextualización da enquisa, presentamos unha tipoloxía de enquisas e dedicamos un epígrafe ó cuestionario, dispositivo de mediación entre o investigador e a poboación analizada. A continuación, centrámonos na selección das unidades de observación e comentamos as características máis salientables das mostraxes de tipo probabilístico e de tipo non aleatorio. Finalmente, dedicamos un apartado específico ó tratamento dos datos e nel comentamos os aspectos teóricos e metodolóxicos máis salientables dos procedementos da descrición e da inferencia estatística.

**Palabras clave:** enquisa, metodoloxía, sociolingüística, métodos cuantitativos.

## 1. Introducción

A enquisa estatística constitúe o instrumento tradicionalmente vencellado á investigación cuantitativa orientada á observación e á análise da estrutura social. A cuantificación de obxectos sociais segue sendo unha das tendencias máis salientables no proceso de elaboración de coñecemento científico, en especial desde a impronta hexemónica das perspectivas teóricas positivistas. Entre as técnicas de produción de datos relacionadas coa investigación científico-social, non é esaxerado dicir que a enquisa continúa a ser a máis utilizada, ademais de constituír un referente básico na análise de repertorios verbais estandarizados. O contexto emerxente das enquisas estatísticas hai que situalo nas propias orixes da investigación social empírica, no século XIX, precisamente cando os saberes sociais aspiran á consideración de ciencias obxetivas e universais. Sen embargo, non será ata a primeira metade do século XX cando o uso desta técnica se estenderá dominando boa parte do traballo empírico feito nas ciencias sociais en xeral (García Ferrando, 1994; Arias Astray & Fernández Ramírez, 1998)<sup>1</sup>.

---

<sup>1</sup> Os precedentes da enquisa remóntanse moitos séculos atrás, no seu uso como procedemento xudicial, primeiro eclesiástico e logo laico (Ibáñez, 1985: 235). Os seus antecedentes inmediatos cómpre situalos no século XVII. Arias Astray & Fernández Ramírez (1998) destacan os seguintes usos das enquisas antes do século XX: a *aritmética política*, é dicir, a recollida de datos demográficos por parte dos gobernos durante o século XVII, a *estatística moral*, orientada a acumular informes relacionados coa saúde pública, as cualidades morais da poboación, etc. (que serían a base da investigación de Durkheim para o seu estudio sobre o suicido) e o *Movemento de Enquisas e Monografía Sociais* durante o século XIX; este último tiña como principal característica “orientar, basándose en el estudio objetivo y cuantitativo de la sociedad, las reformas y mejoras necesarias para solventar los principales problemas sociales del momento” (p. 35). Ademais doutros autores menos relevantes do século XIX, Marx e Weber mostraron gran interese polas enquisas sociais (García Ferrando, 1994: 148). Marx iniciou unha investigación mediante enquisas, destinada a coñecer a situación da clase obreira francesa en 1880 e que esta-

Por enquisa enténdese unha técnica de produción de datos fundamentada na selección dunha mostra representativa de individuos dunha poboación concreta (denominada *universo*) ós que se lles subministra un cuestionario estandarizado, que pretende situar a tódolos entrevistados nunha mesma realidade contextual co obxectivo de extraer estatísticas extrapolables a toda a poboación obxecto de estudo. Polo tanto, seguindo a Fowler (1993: 1), a investigación mediante enquisas identifícase coas seguintes características:

1. O seu principal fin é producir estatísticas, é dicir, descrições cuantitativas ou numéricas dalgúns aspectos da poboación estudada
2. Os datos recóllense mediante un cuestionario. Por tanto, as respostas dadas constitúen a base dos datos que serán analizados
3. Por regra xeral, analízase só un grupo de individuos (mostra) coa finalidade de poder facer inferencias a todo o conxunto (poboación).

Na definición anterior, recollemos varios conceptos que é preciso distinguir da noción de enquisa e coa que manteñen unha estreita relación pero que non deben confundirse: entrevista e cuestionario. A *entrevista* (implicitamente recollida na nosa definición) é a situación na que se produce a interacción –asimétrica– entre entrevistador e entrevistado e como tal é unha parte fundamental de moitas enquisas<sup>2</sup>. Toda entrevista require unha planificación que se inicia coa propia selección aleatoria da persoa susceptible de ser entrevistada, a presentación, a propia realización e a supervisión e control (González Gómez & Padilla García 1998). Pola súa parte, o *cuestionario* é o instrumento que utiliza o entrevistador para guiar a súa entrevista; nel recóllense, ordenadamente, as variables obxecto de estudo da enquisa, a través de expresións (preguntas e respostas) que operativizan as variables da investigación. Trátase, sen dúbida, do instrumento máis utilizado para a obtención de datos en socioloxía empírica (García Ferrando 1994: 158)

A pertinencia da enquisa en determinadas áreas da investigación sociolingüística é indubidable. Como é ben sabido, o campo de estudo da sociolingüística é moi amplo, tal como se recolle nalgúns dos manuais relacionados con este campo do saber que foron publicados nos últimos anos (Wardhaugh, 1992; Chambers, 1995;

---

ba previsto que fose publicada na *Revue Socialiste*. Para tal fin, redactou un cuestionario de 100 preguntas que foi remitido a tódolos asalariados franceses. Unha tradución do seu cuestionario pode consultarse en VV.AA (1973).

<sup>2</sup> Ademais das entrevistas utilizadas na investigación socioestatística, hai outras modalidades de entrevista máis propias de deseños cualitativos, como pode ser a *entrevista aberta* (con distintos graos de estruturación), ou mesmo a *entrevista a grupos*.

Coulmas, 1996; Romaine, 1996; Boix & Vila, 1998; Moreno Fernández, 1998; García Marcos, 1999; Mesthrie et al., 2000). O abano de temas incluído nalgúns deles é tan diverso que, fóra dunha vaguidade, non é doado proporcionar unha definición que satisfaga igualmente a uns e a outros. Mesmo se partimos dunha distinción entre micro e macrosociolingüística atopamos tal cantidade de métodos, modelos e intereses en cada lado que tal variedade obriga a facer unha matización do alcance desta distinción, que, ó noso entender, só debe ser usada epistemoloxicamente, é dicir, como niveis de análise. No nivel microsociolingüístico analízanse as actividades e interaccións (acción social) mentres que no nivel macrosociolingüístico son os procesos colectivos (a estrutura social) os pertinentes; neste último caso o principal interese pasa a ser analizar a distribución dos membros dunha poboación ó longo dunha serie de parámetros lingüísticos e sociolóxicos que permiten agrupar ou diferenciar os individuos. O máis importante neste último tipo de análise é proporcionar modelos que dean conta do grao en que eses parámetros correlacionan entre si. Se concibimos as linguas como institucións organizadas estruturalmente dentro das sociedades, a análise de determinados aspectos da relación lingua-sociedade debe basearse na produción de datos mediante enquisa, dando conta de aspectos tan diversos como as actitudes lingüísticas, o contacto de linguas, o mantemento e a substitución lingüísticas, a demografía lingüística, a planificación lingüística, etc. Por exemplo, se partimos de que nunha situación de plurilingüismo social, hai unha determinada distribución dos diferentes grupos lingüísticos, mediante unha enquisa poderemos cuantificar e medir esa distribución. Ou se pretendemos coñecer cáles son as actitudes lingüísticas dos falantes ante unha determinada variedade lingüística e o seu uso nos sectores que favorecen o prestixio social, a enquisa serve tamén como dispositivo de produción de datos, sendo as actitudes, como son, inobservables de xeito directo<sup>3</sup>. En definitiva, a elaboración de mapas sociolingüísticos de diversa complexidade, que permitan obter descrições da distribución das linguas nas sociedades para afondar na planificación lingüística, xustifica o uso sistemático da enquisa como técnica de produción de datos.

Con todo, a enquisa non sempre constitúe un dispositivo adecuado na análise da relación lingua-sociedade, especialmente cando se analizan fenómenos sociolingüísticos difusos, semántica e pragmaticamente complexos e pouco homoxéneos (Navarro, 1995). Por exemplo, cando o obxectivo é estudar a rede de relacións dos individuos e o papel da lingua na construción desa rede de relacións, a enquisa é inoperativa. Ou se o obxecto de estudio é a construción das identidades lingüísticas

---

<sup>3</sup> “Hay que señalar que las encuestas son una de las escasas técnicas disponibles para el estudio de las actitudes, valores, creencias y motivos. Tan es así, que incluso en estudios experimentales los investigadores suplementan su diseño experimental con cuestionarios o entrevistas con el fin de conocer si una determinada variable tiene realmente los efectos que se pretenden” (García Ferrando, 1994: 152). Véxase tamén Blair (1990).

no discurso, teremos que recorrer a outros procedementos de creación de datos. Coincidimos con Ibáñez (1985: 229) cando di que “una situación de produción de datos es adecuada para investigar situaciones fuera del proceso de investigación que sean semejantes a ella”. Neste sentido, e seguindo co exemplo das actitudes lingüísticas, a actitude ante o uso do galego na publicidade, por exemplo, ten a forma dun cuestionario, no que cada individuo se situará dentro dun repertorio pechado de alternativas, que van desde o acordo total ó desacordo total. E nese caso, a enquisa é un artefacto adecuado. Outra cousa moi distinta é que o noso obxectivo sexa analizar as verbalizacións actitudinais producidas por unha técnica máis expresiva como é o grupo de discusión, por exemplo. Os obxectivos son distintos o cal implica que utilizemos técnicas de produción de datos diferentes.

## 2. Tipoloxía de enquisas

Aínda que a definición anterior recolle os elementos esenciais de toda enquisa, na práctica atopámonos con diversas formas de proceder. Se partimos da forma de administración do cuestionario, distinguiremos os seguintes tipos<sup>4</sup>:

Enquisa telefónica

Enquisa por correo

Enquisa persoal (“cara a cara”)

Enquisa electrónica (correo electrónico e internet)

A selección dun ou doutro tipo dependerá dos obxectivos e das posibilidades materiais da investigación. Para determinados obxectivos será máis adecuada a enquisa telefónica, mentres que noutros casos optaremos pola enquisa persoal. Por exemplo, se unicamente desexamos coñecer a lingua inicial dunha poboación determinada, será suficiente utilizar un cuestionario breve, de 4 ou 5 preguntas, facilmente administrado a partir dunha chamada telefónica. Sen embargo, se queremos realizar un estudio máis minucioso sobre a situación sociolingüística dunha determinada zona xeográfica, será preferible optar por enquisas persoais ou por correo. Ademais, á hora de tomar unha decisión cómpre ter en conta os recursos económicos e humanos que temos á nosa disposición, así coma o tempo planificado para levala a cabo. Xa que logo, cada unha das modalidades expostas ten a súas vantaxes e os seus inconvenientes. Nas próximas páxinas describiremos cada tipo sinalado e finalmente incluíremos un cadro resumo facendo especial énfase nas vantaxes e inconvenientes en cada caso.

---

<sup>4</sup> Esta tipoloxía inclúe tan só os catro modelos de enquisa máis frecuentes, os tres primeiros como exemplos clásicos e o último como modalidade emerxente. Con todo, e tal como recolle Díaz de Rada, citando a DeLeeuw & Nichols (1996), a tipoloxía de enquisas é algo máis complexa especialmente o que ten que ver coas enquisas automatizadas (Díaz de Rada, 2000b; Dillman, 1999).

### *Enquisa telefónica*

Esta modalidade de enquisa caracterízase por usar as redes telefónicas como a canle entre entrevistador e entrevistado, o que implica a ausencia física do primeiro. Na actualidade, gracias ós avances tecnolóxicos de apoio á enquisa telefónica como o sistema CATI (Bosch & Torrente, 1993; Wert, 1994; Cea D’Ancona, 1998: 247-250; Díaz de Rada, 2000a) ou a máis recente *entrevista telefónica automática* (Díaz de Rada, 2000b), o seu uso é moi habitual na investigación de mercados e nos estudos de opinión<sup>5</sup>. Sen embargo, na sociolingüística é aínda moi pouco usada, especialmente porque, como dixemos anteriormente, a principal característica da enquisa telefónica é a limitación temporal na duración da chamada. O tempo medio dunha enquisa deste tipo está entornado ós 10-15 minutos o cal dificulta enormemente a administración dun cuestionario extenso<sup>6</sup>. Con todo, en estudos sociolingüísticos que teñan como obxectivo recoller mostras de fala espontánea para analizar determinadas estratexias pragmáticas, o uso do teléfono pode ser de maior utilidade.

### *Enquisa por correo*

A principal característica da enquisa postal é a autoadministración derivada da ausencia de entrevistador (Bosh & Torrente, 1993). O que se fai nestes casos é seleccionar unha mostra a partir de bases de datos con enderezos persoais á que se remiten o cuestionario xunto coas instrucións para o seu adecuado cumprimento, que deben ser claras e suficientemente atractivas como para garantir a súa lectura e posterior resposta (Manzano & González Gómez, 1998). A diferenza doutras modalidades, a enquisa por correo ten unha taxa de resposta máis baixa, o que pode chegar a enfeblecer substancialmente a mostra. Na investigación sociolingüística tampouco é unha modalidade moi frecuente.

---

<sup>5</sup> A tecnoloxía CATI (acrónimo da expresión inglesa *Computer Assisted Telephone Interviewing*) ten as súas orixes a mediados da década dos anos 70. Trátase dun programa informático deseñado para incluír un cuestionario que o entrevistador ten na súa pantalla. A medida que se van obtendo as respostas o entrevistador vains incluíndo no cuestionario. Unha das principais vantaxes deste sistema é a eliminación da fase de gravación dos datos ademais de mellorar a calidade da entrevista, xa que o entrevistador ten en cada momento un control automatizado do ciclo da mesma. Outra vantaxe nada desdeñable é a proporción dunha selección ó azar dos números de teléfono que se utilizarán na enquisa. Pola súa parte, a *entrevista telefónica automatizada* caracterízase pola ausencia dun entrevistador; é un programa informático o que realiza a selección dos informantes, o que “fai” automaticamente a chamada, formula as preguntas e marca as respostas de entre as ofrecidas polo programa. Unha análise das diferencias entre CATI e *entrevista telefónica automática* pode consultarse en Díaz de Rada (2000a).

<sup>6</sup> En palabras de Wert (1994: 183-4) “se ha acreditado que mientras en las entrevistas de duración inferior a 15 minutos, las entrevistas válidas alcanzan casi el 50% de los contactos telefónicos efectivos, en las de duración superior a los 15 minutos, ese *ratio* desciende al 33%”.

*Enquisa persoal*

A enquisa persoal, tamén coñecida coa expresión “cara a cara”, constitúe, sen dubida, o procedemento de captación de datos máis utilizado na investigación sociolingüística interesada no coñecemento, uso e actitudes das linguas dunha determinada comunidade. A formato deste tipo de enquisa require a presenza dun entrevistador que realiza a súa entrevista na casa do entrevistado. Cando se trata de entrevistas de reducidas dimensións, poden realizarse tamén na rúa ou noutros espazos públicos como o lugar de traballo.

*Enquisa electrónica*

As posibilidades que ofrecen as novas tecnoloxías e o seu extraordinario desenvolvemento e aplicación na captura de datos primarios útiles para a investigación social, fai das modalidades de enquisa electrónica (correo electrónico, internet, telefonía móbil, etc.) un poderoso instrumento na investigación actual das ciencias sociais en xeral e da sociolingüística en particular (véxase Sobrino Blanco, neste mesmo volume; Dillman, 1999). A comodidade e rapidez para administrar a entrevista así como a escasa reactividade por parte do entrevistador son quizais as principais características deste tipo de enquisas. Pero, de momento, esta modalidade de enquisa non parece que vaia substituír os outros tres tipos expostos, debido fundamentalmente a tres factores (Dillman, 2002):

- a) limitada restricción de acceso,
- b) dificultade para garantir modelos mostrais, e
- c) moi baixas taxas de resposta, inferiores ás obtidas mediante unha enquisa telefónica.

Ofrecemos a continuación unha táboa resumo (Táboa 1) na que se comparan as vantaxes e inconvenientes dos catro tipos de enquisa comentados.

**Táboa 1.** *Vantaxes e inconvenientes dos distintos tipos de enquisas*

TIPO DE ENQUISA	VANTAXES	INCONVENIENTES
<i>Por teléfono</i>	<ul style="list-style-type: none"> <li>a) Custos baixos</li> <li>b) Accesibilidade a mostras dispersas</li> <li>c) Baixa reactividade</li> <li>d) Control de execución e supervisión (CACTI)</li> <li>e) Rapidez de execución</li> </ul>	<ul style="list-style-type: none"> <li>a) Limitacións mostrais ó excluír as persoas sen teléfono</li> <li>b) Taxa de resposta media</li> <li>c) Restriccións no cuestionario</li> <li>d) Brevidade</li> </ul>
<i>Por correo</i>	<ul style="list-style-type: none"> <li>a) Custos baixos</li> <li>b) Mínima infraestrutura</li> <li>c) Accesibilidade a mostras dispersas</li> <li>d) Disponibilidade temporal do entrevistado</li> <li>c) Mínima reactividade</li> </ul>	<ul style="list-style-type: none"> <li>a) Baixa cooperación</li> <li>b) Nulo control sobre a entrevista</li> <li>c) O entrevistado pode alterar a lóxica do cuestionario e contestar a determinadas preguntas despois de lelo completamente</li> </ul>
<i>Persoais</i>	<ul style="list-style-type: none"> <li>a) Alta cooperación</li> <li>b) Elevado control sobre a entrevista</li> <li>c) Calidade da entrevista</li> <li>d) Cuestionario máis extenso</li> </ul>	<ul style="list-style-type: none"> <li>a) Custos elevados</li> <li>b) Elevada reactividade</li> <li>c) Infraestrutura complexa</li> <li>d) Inaccesibilidade a certos grupos (elites)</li> <li>e) Tedioso proceso de gravación<sup>7</sup></li> </ul>
<i>Electrónicas</i>	<ul style="list-style-type: none"> <li>a) Baixos custos</li> <li>b) Rapidez</li> <li>c) Facilitade de tratamento dos datos</li> </ul>	<ul style="list-style-type: none"> <li>a) Imposibilidade de acceder a certos grupos poboacionais</li> <li>b) Baixa cooperación</li> </ul>

<sup>7</sup> Este aspecto pódese mellorar considerablemente se recorre a una entrevista persoal asistida por ordenador (CAPI, *Computer Assited Personal Interview*).



### 3. O dispositivo de mediación: o cuestionario

Como sinalamos anteriormente o cuestionario é a ferramenta que manexa o entrevistador para levar a cabo a entrevista<sup>8</sup>. Pero trátase dunha ferramenta especial. Lonxe de ser un mero instrumento de recollida de datos, o cuestionario é o dispositivo no que se sintetizan as variables obxectivos de estudio estandarizadas en forma de preguntas e respostas coas que entrevistador e entrevistado interactúan formalmente. É dicir, no cuestionario elabórase unha representación máis ou menos complexa do obxecto de estudio da enquisa. A formulación das preguntas e a acoutación da maioría das respostas crea unha *interacción orientada*, onde a mobilidade está restrinxida ós presupostos dos que parte o analista<sup>9</sup>. Por exemplo, preguntar nun cuestionario pola “lingua inicial”, proporcionándolle ó entrevistado unha reducida batería de respostas, implica unha concepción determinada desta noción: presupoñer que as linguas existen, que existen da mesma maneira para tódolos entrevistados, que tódolos entrevistados entenderán por esa noción as mesmas cousas, etc. Tales presupostos xorden precisamente dunha das características fundamentais dos cuestionarios que consiste en situar a cada un dos informantes nunha situación similar, independentemente do punto de partida de cada un deles, das súas realidades. Garantir isto é un dos obxectivos de todo cuestionario, isto é, homoxeneizar para afortalar a *fiabilidade* e a *validez* do instrumento de recollida de datos mediante a comparación de resultados.

Se tódalas etapas da investigación mediante enquisas son importantes, a elaboración dun cuestionario é, se cadra, a fundamental. Unha inadecuada elaboración pode botar por terra boa parte do traballo. Seguindo a Padilla et al. (1998: 121), na elaboración dun cuestionario debemos prestar atencións ós seguintes aspectos:

---

<sup>8</sup> Como da maior parte dos aspectos relacionados coa enquisa, a bibliografía específica sobre cuestionarios é moi extensa. Cada ano publicanse decenas de libros que teñen como obxecto fundamental a reflexión teórico-metodolóxica sobre o cuestionario e a súa aplicación. A seguinte selección é, logo, só unha mostra da enorme produción neste campo: Converse & Presser (1986); Fowler Jr. & Fowler (1995); varios dos volumes de Fink (1995); Fink & Kosecoff (1998); Foddy (1996); Padilla García, González Gómez & Pérez Meléndez (1998); Azofra (1999); Paterson (1999).

<sup>9</sup> Ibáñez (1985: 245) refírese á enquisa como a “una interacción verbal cerrada”, na que o cuestionario “contine todas las jugadas posibles”. Cfr. Navarro (1995) onde, fronte á perspectiva positivista (distributiva) na que tradicionalmente se desenvolven as enquisas, propón reciclar esta técnica para buscar a súa utilidade tamén na análise sociolóxica estrutural ó mesmo nivel cás técnicas relacionadas con esta perspectiva metodolóxica, o grupo de discusión e a análise do discurso. Desde esta perspectiva, as respostas a un cuestionario “sí bien encorsetadas en las opciones cerradas que se les ofrecen, son en realidade trasunto de sus respectivos intereses e ontoloxías [dos do entrevistado/a], a menudo discordantes en relación con las que el encuestador propone. Son respuestas que, por ello, constituyen un texto cualitativamente distinto al formulado en el cuestionario, a pesar de que éste fagocite y asimile ese texto por completo, en el curso del procesamiento estándar de la encuesta” (Navarro, 1995: 9).

- a) Determinación precisa dos obxectivos do cuestionario
- b) Planificación do cuestionario
- c) Elaboración das preguntas
- d) Análise da calidade das preguntas
- e) Disposición provisional das preguntas no caderno
- f) Análise da fiabilidade e validez do cuestionario
- g) Edición final do cuestionario

Comentaremos brevemente algúns destes aspectos tomando como exemplo unha recente investigación sociolingüística. Trátase do *Mapa Sociolingüístico de Galicia* (MSG), realizado polo Seminario de Sociolingüística da Real Academia Galega na década dos 90 (Seminario de Sociolingüística 1994, 1995, 1996). Nesta investigación utilizouse unha mostra de 38897 entrevistados maiores de 15 anos ós que se lles pasou un cuestionario de 148 ítems, o que dá conta da envergadura deste traballo.

#### **a) Determinación precisa dos obxectivos do cuestionario**

O obxectivo principal desta macro-enquisa era elaborar unha exhaustiva base de datos da relación entre lingua e sociedade en Galicia que permitira realizar unha descrición o máis completa posible da situación sociolingüística dos grupos lingüísticos desta comunidade, “de xeito que non só sirva para afondar no seu coñecemento, senón que ademais poida ser de proveito para futuras tarefas de planificación lingüística en Galicia” (Seminario de Sociolingüística, 1996: 11).

#### **b) Planificación do cuestionario**

A cuestionario dividiuse en varias partes. Os 148 ítems cos que contaba estaban estruturados como a seguir se indica (Seminario de Sociolingüística, 1994: 20):

- 1) Datos de control, onde se incluían os datos persoais do entrevistado (nome, teléfono, enderezo, etc.) o día e a hora na que tivo lugar a entrevista e os códigos de entrevistador, supervisor, gravador e depurador dos datos.
- 2) Variables sociais relacionadas coa persoa entrevistada (sexo, idade, clase social, lugar de nacemento e de residencia, nivel de ingresos, profesión, estudos, etc.).
- 3) Datos sociolingüísticos (lingua inicial, competencia lingüística, historia sociolingüística, usos lingüísticos, actitudes lingüísticas, conciencia sociolingüística, etc.).
- 4) Outras cuestións (observacións feitas polo entrevistador sobre as linguas de uso durante a entrevista, a situación social da entrevista, etc.).

### c) Elaboración das preguntas

A elaboración das preguntas é a parte central de calquera cuestionario. Por se tratar a entrevista dunha técnica eminentemente lingüística –un texto– que incide na función referencial do compoñente simbólico (Ibáñez, 1985), a formulación das preguntas requirirá unha especial atención, xa que unha deficiente formulación de preguntas e respostas devaluará seriamente todo o traballo de enquisa. Pero ¿como sabemos que estamos ante unha boa pregunta?. Non sempre é fácil rsponder a isto. En xeral, unha boa pregunta é aquela que produce respostas que son medidas fiables e válidas daquilo que queremos analizar<sup>10</sup>.

Na práctica, a construción do cuestionario require tempo e a redacción final das preguntas debe superar aínda un período de proba denominado *pre-test*. A fase de redacción das preguntas que foron incluídas finalmente na investigación que estamos a comentar ocupou varios meses. Dado que o cuestionario é, tamén, un instrumento de medida, intentamos ter en conta os seguintes aspectos que consideramos fundamentais para garantir unha boa mediación (Fowler & Fowler, 1995: 3-4):

- 1) As preguntas deben ser comprendidas coherentemente.
- 2) As preguntas deben ser administradas ou comunicadas ós entrevistados de xeito coherente.
- 3) As posibilidades de resposta (en preguntas pechadas) deben ser comunicadas coherentemente.
- 4) A menos que o obxectivo da pregunta sexa avaliar o coñecemento, tódolos entrevistados deben ter acceso á información necesaria para responder adecuadamente.
- 5) Os entrevistados deben estar dispostos a proporcionar as respostas esixidas polas preguntas.

Xa que logo, para garantir que cada entrevistado entenda as mesmas cousas en cada pregunta, é preciso seguir certas pautas, como por exemplo, evitar preguntas ambiguas, evitar preguntas que inclúan máis dunha cuestión, etc.

Nos puntos anteriores evidénciase, paradoxalmente, a principal limitación do cuestionario. Simplifícase a linguaxe para facela máis comprensible co cal se reducen as posibilidades de producir datos máis intensivos:

La pregunta y las respuestas han de formularse en un lenguaje comprensible, lo que obliga a simplificar las cosas. Lo que tiene (también) consecuencias semánticas –al exigir niveles mínimos de competencia lingüísti-

---

<sup>10</sup> Unha boa pregunta é fiable cando proporciona respostas constantes, coherentes ou estables en situacións comparables. De non ser así, o instrumento de medida deixa de medir adecuadamente e o resultado é insatisfactorio (*vid.* § 2.4)

ca sólo se alcanza lo general, pero no lo particular y lo singular– y pragmáticas –la encuesta forma parte de un dispositivo de homogenización (Ibáñez, 1985: 233).

As preguntas poden ser abertas ou pechadas. As primeiras permiten que o entrevistado elabore a súa resposta con menos restricións. Son moi adecuadas para acumular respostas menos previsibles. Por exemplo, para obter información relacionada coas profesións. Pódese optar por ofrecer unha batería pechada de profesións, pero dada a enorme variedade de posibilidades é máis adecuado preguntar abertamente por esta cuestión e posteriormente elaborar categorías. Por esta razón, teñen un uso máis limitado debido ós problemas que presenta o seu tratamento (categorización, codificación, etc.). Nas preguntas pechadas, o entrevistador presenta –implícita ou explicitamente– as posibles respostas, acoutando a pluralidade dos referentes. Neste caso, as respostas deben cumprir o principio de exhaustividade e ser, asemade, mutuamente excluíntes. A continuación proporcionamos un exemplo de cada tipo:

---

PREGUNTA ABERTA: *¿Cál é a súa profesión?*

PREGUNTA PECHADA:

*¿En que lingua(s) aprendeu a falar?*

En galego .....(1)

En castelán.....(2)

Nas dúas .....(3)

Noutra(s).....(4)

---

Non tódolos entrevistados teñen que contestar tódalas preguntas. Con frecuencia, os cuestionario inclúen preguntas *filtro*, isto é, preguntas que serven para reorientar a entrevista segundo se responda dun xeito o doutro.

#### **d) Análise da fiabilidade e validez do cuestionario**

O cuestionario, como artefacto de medida, debe aglutinar algunhas características. Entre elas, as máis relevantes son a *fiabilidade* e a *validez*. Tanto unha coma a outra son dous requisitos imprescindibles en toda investigación social mediante enquisa (Mayntz, Holm & Hübner, 1975; Latiesa, 1994; Cea D’Ancona, 1998). Avaliar un instrumento de medida é un proceso complexo. En primeiro lugar, cómpre determinar a súa validez, é dicir, comprobar que o que se está a medir é precisamente o que se quere medir. Se entendemos a medición como o proceso de vincular conceptos abstractos con indicadores empíricos (Carmides & Zeller, 1979), un instru-

mento de análise é válido se os seus indicadores –as preguntas do cuestionario– “indican lo que se pretende que indiquen; si ‘miden’ correctamente el *significado* dado al *concepto teórico* en consideración” (Cea D’Ancona, 1998: 150)<sup>11</sup>. Pola súa parte, a fiabilidade ten que ver coa consistencia do instrumento, coa súa capacidade de obter resultados semellantes en distintos momentos. Un cuestionario será tanto máis fiable canto máis estandarizado estea, é dicir, se é igualmente comprensible por todos os entrevistados<sup>12</sup>.

#### 4. Cooperación e cuestionario: o tratamento da ‘non resposta’

Unha dos principais desafíos cos que se debe enfrentar todo investigador que recorra á enquisa mediante cuestionario é a posibilidade de non cooperación do entrevistado, o que constitúe un tipo de erro non sempre recoñecido na investigación empírica. A falta de cooperación pode ser “total”<sup>13</sup>, cando se produce un rexeitamento ó conxunto do cuestionario ou “parcial”, cando o entrevistado decide non responder a determinados ítems<sup>14</sup>. Tanto nun caso coma no outro, elevadas taxas de *non resposta* poden producir nesgos relativos na interpretación e extrapolación dos datos (Fowler, 1993). Hai grupos determinados de poboación que nunca responden a cuestionarios, por diversas razóns nas que non é necesario entrar neste momento (Díaz de Rada, 2000a; Sánchez Carrión, 1995, 2000). O verdadeiramente relevante para a investigación é que en moitas ocasións o investigador social “asume” que os grupos que non acostuman contestar son similares ós que si teñen ese hábito, o cal está lonxe de ser certo. Xa que logo, nestes casos hai que optar por unha solución. Hai diversas estratexias que se poden seguir para minimizar o efecto da “non resposta”. Segundo Díaz de Rada (2000a), podemos basicamente distinguir dous tipos: as estratexias para *reducir* a non resposta e as estratexias para *mitigar* a non-resposta. Entre as primeiras están, por un lado, as que inciden de xeito principal no aumento da probabilidade de cooperación

---

<sup>11</sup> Na década dos 70 propuxéronse diversos tipos de validez e de fiabilidade. Entre os primeiros cómpre destacar a validez interna, a validez externa, a validez de constructo e a validez estatística. Entre os segundos, a fiabilidade como estabilidade, a fiabilidade como equivalencia e a fiabilidade como consistencia. Consúltase Latiesa (1994) para a caracterización de cada tipo citado.

<sup>12</sup> Para as estratexias de comprobación dos distintos tipos de validez e fiabilidade, véxase Latiesa (1994).

<sup>13</sup> A non-resposta total é máis frecuente nas enquisas telefónicas, por correo ou nas electrónicas. Sen embargo, determinadas cuestións relacionadas coa información de tipo persoal ou confidencial teñen unha alta taxa de non-resposta en enquisas persoais, debido precisamente á presenza do entrevistador.

<sup>14</sup> Cómpre matizar que moitos casos de non-resposta total non son derivados da falta de cooperación dun determinado informante, senón que a súa orixe está nunha insatisfactoria cobertura ou noutros problemas derivados da selección do entrevistado, como a ausencia do seu domicilio, falecemento, cambio de enderezo, etc. (Díaz de Rada, 2000a: 14-16).

dos entrevistados, acrescentando a motivación para responder e, por outro, atopámonos con estratexias que teñen como principal finalidade a re-entrevista mediante procesos de substitución dos non-cooperantes. Entre as segundas, as destinadas a mitigar os efectos, é preciso destacarmos as estratexias post-entrevista, é dicir aquelas que se actúan unha vez rematado o traballo de campo e que na maior parte dos casos tratan de solventar os efectos da non-resposta parcial. Estas últimas son basicamente extrapolacións e ponderacións dos datos recollidos, por exemplo: distribuír as non-respostas entre toda a mostra, respectando o principio de proporcionalidade, o cal conleva asumir que as taxas de non-resposta seguen un patrón regular en toda a mostra.

### **5. A selección das unidades de observación**

Un dos aspectos máis importantes dos protocolos de investigación que implican a utilización de enquisas e cuestionarios é a selección axeitada da poboación que vai ser o obxecto de estudo. Traballar con poboacións enteiras é un obxectivo realmente desbotado na práctica investigadora: en primeiro lugar, porque ás veces é imposible na práctica conseguir información de toda a poboación; e en segundo lugar porque os resultados non serían cribles. E se fosen cribles, xa non serían aplicables á poboación obxectivo, pois esta estaría transformada. Tendo en conta a necesidade de dispoñer dunha selección da poboación, a precisión e a rigorosidade deben extremarse xa que a análise desa parte do total producirá conclusións que aplicaremos ó conxunto da poboación.

O primeiro que debemos realizar é identificar o *universo* do noso estudo. Por exemplo, o *Mapa Sociolingüístico de Galicia* delimitaba o seu obxecto de estudo entre as persoas maiores de 15 anos residentes en Galicia. A continuación, hai que obter o *marco mostral*: tratáse dunha listaxe que conteña tódolos elementos relevantes da poboación definida. O contido do marco mostral –estrutura, información dispoñible e calidade da mesma– determinarán o deseño do mostra que imos utilizar. Tal como indica a este respecto Díaz de Rada (2001: 57-8), “en la medida en que el conocimiento de la poboación sea más perfecto la equivalencia entre el marco muestral y la población será mayor, siendo posible reducir determinados errores susceptibles de cometerse en la selección muestral.”

#### **5.1. O concepto de representatividade**

A cuestión da representatividade dunha mostra radica en establecer ata qué punto a investigación que concibimos e que desenvolvemos recolle unha imaxe fiel e realista da poboación estudada. Estamos, sen lugar a dúbidas, ante un tema de crucial importancia na investigación mediante enquisas. De pouco serve ter un cuestionario impecable e uns entrevistadores expertos se non está garantida a representatividade da mostra. Garantir que a nosa mostra *representa* unha determinada poboación que, polo tanto, podemos estar en condicións de facer inferencias dunha á outra é, en

definitiva, a principal finalidade da representatividade (Kish, 1995)<sup>15</sup>. A representatividade dunha mostra atópase en relación directa co tamaño, a *fiabilidade* e a precisión da base de mostraxe, a aleatoriedade, así como do rigor do protocolo de investigación, principalmente no que atinxe á recollida de datos.

## 5.2. Tipos de mostraxes probabilísticas

Xa que un dos obxectivos centrais da investigación desenvolvida a través de enquisas é a presentación de inferencias respecto da poboación obxecto de estudo, centrarémonos nesta sección só nos métodos de mostraxe de tipo probabilístico, e deixamos para o seguinte epígrafe os métodos non probabilísticos máis salientables.

Os deseños aleatorios posúen dúas características principais: as unidades de mostraxe da poboación teñen unha probabilidade coñecida e distinta de cero para seren seleccionadas, e a selección realízase mediante un procedemento que respecta as ditas probabilidades.

Existe unha diversidade de mostraxes probabilísticas das cales analizaremos a continuación catro: simple, sistemática, estratificada e de conglomerados.

### *Aleatoria simple*

É o tipo de mostraxe aleatoria máis sinxela e habitual na bibliografía sobre enquisas. Este método consiste en identificar, a partir dun marco, listaxe ou base de datos en que se atope toda a información precisa, tódalas unidades de poboación do marco mostral. A continuación, utilizando algún tipo de procedemento aleatorio –por exemplo, un bombo de lotería, unha táboa de números aleatorios, un programa de ordenador, etc.–, empregamos unha técnica aleatoria para seleccionarmos  $n$  unidades e tales unidades formarán parte da mostra –e serán o noso obxecto de estudo.

Este método presenta múltiples vantaxes derivadas da facilidade dos cálculos estatísticos e as probabilidades iguais de selección. Con todo, presenta inconvenientes que, ás veces, fan imposible utilizalo: a dificultade para dispoñer dunha listaxe –fiable e actualizada– da poboación, o caro que pode resultar dado a dispersión alcanzada pola mostra, ou as dificultades para establecermos comparacións entre grupos distintos dentro dunha mesma mostra. Tendo en conta as ditas vantaxes e desvantaxes, Díaz de Rada (2001: 61) afirma que a mostraxe aleatoria simple “se aplica, fundamental-

---

<sup>15</sup> “El *objetivo* del muestreo representativo reside en hacer que la muestra sea una miniatura que refleje y represente la población por medio de distribuciones *similares*” (Kish, 1995: 29). É interesante destacar o sintagma “distribucións similares” da cita anterior, na medida en que é preciso admitir a variación mostral, inherente a toda investigación social mediante enquisa. Con outras palabras, é pouco verosímil pensar nunha mostra *completamente* representativa dunha poboación. Na investigación de grandes poboacións, o principal obstáculo para lograr unha mostra excepcional está na dificultade para garantir a aleatoriedade das unidades de mostraxe. Na práctica, trátase principalmente dun problema de custos económicos.

mente, en investigacións sobre poboacións pequenas facilmente identificables, fundamentalmente si constitúen universos específicos y diferenciados.”

#### *Aleatoria sistemática*

Aínda que é similar ó anterior, xa que é preciso partir dunha listaxe da poboación, diferénciase del na forma de selección das unidades mostrais. Unha vez que se numeran de forma aleatoria as unidades do marco mostral, o proceso de selección comeza considerando en primeiro lugar o *coeficiente de elevación*: o número de persoas que están ‘representadas’ por cada unidade de mostra, coeficiente que se calcula dividindo o tamaño do universo entre o tamaño mostral. A continuación, selecciónase un número inferior ó coeficiente de elevación, que será o número de arrinque da selección. O segundo elemento para a mostra selecciónase sumándolle a ese número de arrinque o coeficiente de elevación, e así sucesivamente ata completarmos o tamaño total da mostra.

Este procedemento é máis sinxelo có aleatorio simple, pero ten o risco dos potenciais nesgos que se poderían suscitar no caso de que a ordenación dos elementos da poboación estivese realizada de acordo con algún criterio concreto. Se o universo de poboación non fose estrictamente aleatorio, é factible que a mostraxe efectuada non recolla toda a diversidade da poboación.

#### *Aleatoria estratificada*

Este tipo de mostraxe probabilística ten un amplo uso na investigación actual. Parte da división da poboación, que constitúe o obxecto de estudio, en subdivisións de maneira que os resultados obtidos para cada subdivisión poden ser analizados por separado e poden ser comparados entre si. Neste tipo de mostraxe establécense distintos tipos de restriccións derivadas da consideración por parte do investigador de que hai algún conxunto de variables que exercen influencia sobre a característica que se quere medir. É dicir, a poboación divídese en subconxuntos ou estratos e realízase unha mostraxe en cada un deles.

Por exemplo, se a variable que ten influencia é a lingua habitual, teranse en conta as categorías de *só galego*, *máis galego ca castelán*, *máis castelán ca galego* e *só galego* para establecer unha mostra con Z ‘só galego’, H ‘máis galego ca castelán’, L ‘máis castelán ca galego’ e M ‘só galego’. O resultado final será unha mostra de contía  $n = Z + H + L + M$ .

As vantaxes deste procedemento radican en que permite extraer mostras representativas de cada un dos distintos estratos nos que dividimos a poboación, reducindo as varianzas e aumentando, en consecuencia, a precisión dos cálculos e estimacións. Tocante ás desvantaxes, debemos salienta que se precisa moita información sobre o marco mostral, que ten un maior custo respecto a outros procedementos probabilísticos (por exemplo, en comparación co custo da mostraxe por conglomerados, que veremos de seguido), e que o tratamento estatístico dos datos é máis complexo.



*De conglomerados*

Ó igual ca no tipo anterior de mostraxe, os conglomerados son subconxuntos da poboación, é dicir, grupos. Pero a diferenza daquel, nos conglomerados a unidade mostral non son os individuos senón un conxunto de individuos que consideramos que forman unha unidade. O grupo que conforma un conglomerado pode ser unha familia, unha institución (por exemplo, un departamento universitario), unha urna electoral, unha área concreta de terreo (un bloque de edificios, por exemplo), etc. Mentres que na mostraxe estratificada os estratos debían ser internamente homoxéneos, no caso da mostraxe por conglomerados perséguese que estes sexan o máis heteroxéneos posible, coa finalidade de que cada conglomerado constitúa unha representación da variedade e diversidade dos compoñentes do universo.

Despois de seleccionar o número concreto de conglomerados, procederase a entrevistar a tódalas unidades do conglomerado. Este proceso adoita realizarse cando os conglomerados son pequenos, mentres que no caso de que estes teñan un tamaño importante, adóitase realizar unha mostraxe en distintas etapas, proceso denominado *polietápico*<sup>16</sup>.

Este procedemento presenta vantaxes indubidables: en primeiro lugar, para levalo a cabo non é preciso dispoñer de listaxes exhaustivas da poboación, senón que abonda co coñecemento e a descrición dos conglomerados; e en segundo lugar, o custo do traballo de campo que é mester desenvolver é máis reducido ca no caso das mostraxes vistas anteriormente. Con todo, non podemos obviar as desvantaxes que presenta: menor precisión nos datos obtidos e un tratamento estatístico moito máis complexo para calcular o erro mostral.

**5.3. Mostraxes non aleatorias**

En moitas ocasións resulta moi difícil cumprir tódalas condicións propias das mostraxes probabilísticas, ademais de seren, en xeral, caras e prolongadas no tempo. Unha alternativa é empregar algún modelo non probabilístico. En todo caso, trátase dun tipo de mostraxe que deita moitas dúbidas sobre a súa representatividade e obxectividade: ó partir dunha mostra non baseada en criterios aleatorios, xorden problemas pola subxectividade do investigador e polos nesgos propiciados por unha selección intencionada.

---

<sup>16</sup> A *mostraxe polietápica* constitúe un caso especial de mostraxe de conglomerados que consiste en que as unidades mostrais son subdivisións do conglomerado. Algúns autores aconsellan utilizar este procedemento cando sexa difícil obter unha listaxe detallada de tódolos compoñentes dunha poboación, cando se quere reducir o tempo e o custo dun traballo de campo, e cando se levan a cabo estudos de ámbito estatal ou internacional que supoñen unha importante dispersión mostral (véxase Cea D'Ancona, 1998).

Malia o dito, unha mostraxe non aleatoria pode ser moi interesante e factible en determinados casos<sup>17</sup>. Por exemplo, no caso de descoñecermos as unidades que compoñen unha poboación. Ademais, trátase dunha mostraxe fácil de realizar, rápida e barata, de maneira que é moi interesante para estudos mediante os cales só pretendemos obter unha imaxe xeral ou unha primeira impresión dunha determinada poboación: por exemplo, estudos piloto, investigacións de tipo comercial, etc.

Dos diferentes tipos de mostraxe non aleatorias (cf. Manzano 1998: 64-5), presentamos a continuación as características máis relevantes das mostraxes por cuotas e das mostraxes ‘bóla de neve’.

#### *Por cotas*

A mostraxe por cotas supón un procedemento sofisticado que ten un amplo uso. En principio recóllese información sobre algunhas variables relativas á poboación da cal se podería facer unha mostraxe estratificada. Pero o investigador pode optar por proceder deixándose guiar pola súa subxectividade. Na práctica, o método de cotas límitase a un número reducido de dimensións, principalmente aquelas que son de acceso relativamente doado, e asúmese que as restantes características están proporcionalmente representadas. A partir da información contida nunha táboa de continxencia, en que se inclúen os datos pertinentes dos enquisados cuns perfís determinados, formaranse ‘cotas’ conforme a determinadas distribucións de características xa coñecidas.

O realmente peculiar deste método, que o diferencia respecto da mostraxe estratificada, é que é o investigador quen escolle as persoas que vai entrevistar. É dicir, dentro de cada subpoboación, a selección realízase cada entrevistador por elección persoal, sen se basear en ningún tipo de criterio aleatorio.

Trátase dun tipo de mostraxe rápido e barato que adoita dar resultados moi satisfactorios en sondaxes comerciais. Pero, en cambio, conta coa desvantaxe de que o investigador non pode realizar ningún tipo de inferencia estatística sobre os seus resultados, de maneira que é totalmente desaconsellable utilizalo cando se deban tomar decisións estratéxicas.

#### *As mostraxes de ‘bóla de neve’*

Neste caso o investigador selecciona uns poucos individuos, segundo criterios diversos (a súa dispoñibilidade, o feito de seren especialistas, etc.), e a cantidade de

---

<sup>17</sup> Combessie (2000: 72) inclúe baixo a denominación de ‘sondaxes razoadas’ o método das cotas, as enquisas na rúa e as mostraxes de bóla de neve, e indica que se trata de “procedimientos de muestreo que pretenden garantir la representatividad de la muestra adaptando su estructura a las características de la población. La representación es directamente proporcional al número, a la fiabilidad (exactitud) y a la precisión y pertinencia (adaptación a los objetivos de la investigación) de los datos disponibles para el conjunto de la población de referencia, (...)”.

individuos enquisados vai aumentando cos propios contactos dos entrevistados, ata acadar o número previamente fixado. Convén sinalar que moitos metodólogos consideran que este tipo de mostraxe deben evitarse xa que ten un efecto deformador incontrolable.

## 6. Tamaño da mostra

A decisión de qué número de unidades cómpre seleccionar adoita ser un dos aspectos máis problemáticos ós que se enfrenta ó investigador á hora de definir unha mostra. E non adoita ser un aspecto doado porque a mostra debe ser representativa, e para iso o investigador corre o perigo de deseñar unha mostra, ben excesivamente pequena, ben demasiado grande, e ambos extremos non son en absoluto recomendables; no caso de ser pequena, os resultados acadados polo investigador serán moi dificilmente cribles; no caso de ser grande, suporá un estipendio abusivo de recursos materiais, temporais e económicos, como mínimo.

Para conseguir un tamaño óptimo, Manzano (1998: 61-2) ofrece catro tipos de consellos. En primeiro lugar, debe terse en conta o procedemento de mostraxe que eliximos. En segundo lugar, debemos definir con atención o conxunto de variables *importantes* do noso estudio. En terceiro lugar, a esixencia dunha contía determinada no tamaño da mostra está en relación directa co tipo de estudio que realizamos. En cuarto lugar, non se recomenda a utilización de táboas para calcular a mostra. E finalmente, hai que prever as estimacións independentes para cada diversos colectivos. Se queremos establecer conclusións significativas para subconxuntos da poboación, isto suporá aumentar o tamaño dos ditos subconxuntos, de maneira que a mostra aumentará necesariamente de tamaño.

Rodríguez Osuna (1991: 47 e ss.; 1994: 368 e ss.) fai referencia a tres elementos que interveñen na decisión de ampliar ou de reducir o tamaño dunha mostra. En primeiro lugar, a varianza poboacional, definida como o grao de homoxeneidade ou de heteroxeneidade percibida na poboación estudiada. En segundo lugar, o nivel de confianza elixido, definido como “acotaciones de la función de densidad en las cuales existe una determinada probabilidad de que se encuentre el valor poblacional que se pretende estimar” (p. 369). E en terceiro lugar, o erro mostral máximo permitido nas estimacións.

Estes tres factores –a varianza poboacional, o nivel de confianza e o erro mostral– permiten definir o tamaño dunha mostra. A fórmula máis utilizada –partindo de que temos unha poboación infinita (máis de 100.000 elementos) e no caso dunha mostra aleatoria simple–, é a seguinte:

$$N = \frac{K^2 P(1 - P)}{E^2}$$

onde,  $K$  é o nivel de confianza,  $E$  o erro máximo admitido<sup>18</sup> e  $P$  a varianza poboacional suposta.

En calquera caso, débese ter moi presente que facer unha mostra non equivale a traballar coa poboación, de maneira que non temos seguridade absoluta de que os resultados obtidos serían semellantes no caso de considerar toda a poboación no canto dunha porción desta.

## 7. Tratamento dos datos

Unha vez que os datos están xa recollidos e xa os codificamos e tabulamos, estamos en condicións de iniciar un tratamento estatístico. Consideraremos brevemente a continuación as dúas estratexias de tratamento dos datos que podemos afrontar a partir do deseño dunha enquisa: a descrición dos datos e as inferencias estatísticas a partir deles.

### 7.1. Descrición

A descrición estatística ten como finalidade ordenar, resumir e compendiar os datos obtidos. Unha vez que dispoñemos do ficheiro de datos, podemos iniciar a súa explotación. Debemos comezar coa descrición univariable recorrendo a recontos e porcentaxes no caso de variables nominais e a medidas de tendencia central (media, moda, mediana, etc.) e de dispersión (desviación típica, varianza, rango, etc.) no caso das variables codificadas ordinalmente.

Despois de realizarmos a descrición exploratoria da mostra seleccionada, debemos pasar ás análises que relacionan dúas –análise bivariante– ou máis de dúas variables –análise multivariante. Cando as variables dependentes son nominais, o axeitado é elaborar táboas de continxencia<sup>19</sup>. Unha táboa de continxencia é unha matriz de datos na que as fileiras son os valores dunha variable e as columnas os valores doutra. Cada cela da táboa é o cruce dun valor de fileira cun valor de columna. O máis aconsellable é indicar na táboa tanto as frecuencias coma as porcentaxes, para facilitar a súa interpretación.

---

<sup>18</sup> O principal erro que o investigador debe ter presente é o denominado *erro de mostraxe*, que consiste na diferenza entre o resultado que se obtivo na mostra e o resultado que se obtería no caso de considerar a poboación enteira. Ademais do erro de mostraxe hai outros tipos de erro que é preciso ter en conta na investigación mediante enquisa: erros de medición, erros de cobertura, erros derivados da non-resposta, etc.

<sup>19</sup> Na investigación social faise referencia a variables dependentes en relación a aquelas que hai que explicar e que conforman o obxecto de estudo. As variables independentes (predictoras) son as que se utilizan para analizar o seu efecto sobre as dependentes. O carácter de independente ou dependente non está dado pola propia variable senón polos intereses do investigador.

*Medidas de tendencia central*

Se traballamos con variables ordinais, poderemos ofrecer unha táboa descritiva máis completa, engadindo outros cálculos estatísticos básicos como a media, moda, mediana, rango, desviación típica, varianza, etc. A *media* é o estatístico máis utilizado con variables ordinais sempre e cando a variable que se estea a analizar non inclúa valores moi dispares que afecten, no sentido de desvirtuar, a esta medida. Nestes casos sería conveniente calcular a *mediana*, que é a puntuación que ocupa a posición central, ou recorrer á *moda*, que consiste na puntuación máis repetida dunha serie. A media calcúlase sumando tódolos casos e dividindo o resultado polo número total de casos.

Non sempre resulta doado decantarse por unha medida de tendencia central, xa que calquera delas presenta vantaxes e inconvenientes. Sempre que teñamos unha distribución sen valores moi extremos que afecten de maneira considerable á media, esta debería ser a medida principal utilizada, xa que, como acabamos de comentar, ten en conta tódolos datos. En caso contrario, a mediana será a mellor elección.

*Medidas de dispersión*

Xunto ás medidas de tendencia central, sempre debemos ofrecer algún estatístico que dea conta da maior ou menor variabilidade dos datos. Neste caso o máis habitual é utilizar a *desviación típica* o *estándar*, que resulta de calcular o promedio da desviación dos casos con respecto á media. Datos moi homoxéneos terán unha desviación típica pequena e datos moi heteroxéneos o contrario. Deste modo, facilitaremos a interpretación dos datos da nosa investigación.

**7.2. Inferencia**

No caso da estatística inferencial, utilizamos distintas estratexias para facer algún tipo de estimación de significación e de fiabilidade. A estratexia neste caso é formular un conxunto de hipóteses que os datos deben aceptar ou rexeitar.

Non todo o que aparentemente é significativo nunha mostra hai que interpretalo como igualmente relevante na poboación da que foi extraída. A solución a esta cuestión debemos buscala nas probas estatísticas, que nos permiten pasar dos estimadores mostrais ós parámetros poboacionais, cun nivel de confianza predeterminado polo analista e que na demolingüística se adoita situar no 95%, o que equivale a dicir que de cada 20 veces que fagamos a inferencia, 19 serán acertadas. Este nivel de confianza, aínda que de feito é moi elevado, nunca nos garante que as conclusións propostas sexan as correctas, pero esta é a arte da estatística. Ademais, temos que contar sempre co erro mostral, que como xa vimos antes virá determinado polo tamaño da mostra, o nivel de confianza adoptado, a varianza poboacional e o tipo de mostraxe utilizada.

As probas estatísticas divídense en *paramétricas* ou dependentes da distribución e en *non paramétricas* ou de distribución libre. O uso dunhas e outras non é arbitrario senón que vén determinado por unha serie de supostos. Aínda que ás veces estes ter-

mos son utilizados en diferentes sentidos, en xeral podemos afirmar que utilizar unha proba paramétrica implica asumir que os erros aleatorios que afectan ós resultados teñen unha distribución normal, é dicir, os contrastes paramétricos teñen que suponer unha normalidade aproximada da distribución poboacional. Pola súa parte, as probas non paramétricas non requiren este suposto (MacRae, 1995: 154). As probas paramétricas son máis adecuadas cando a medida se realizou cunha escala de intervalo ou de razón, nas que as diferencias entre os valores representan diferencias iguais nos obxectos medidos. As probas non paramétricas teñen normalmente menos potencia para rexeitar a *hipótese nula* cando deberíamos facelo (erro de tipo II, en termos estatísticos)<sup>20</sup>.

### *Análise da varianza*

A análise da varianza é unha das probas máis relevantes da estatística multivariante. Hai varias razóns para preferir unha análise multivariante a unha análise univariante. Quizais a máis importante sexa que os tests que teñen en conta só unha variable ignoran información importante como, por exemplo, a correlación entre variables. Ademais, sabemos que anque as variables poidan non ser significativas tomadas individualmente, si poden selo analizadas de maneira conxunta. Con todo, cando só é necesario estudar a influencia dunha soa variable independente sobre unha variable dependente cuantitativa, podemos recorrer á análise da varianza dun só factor; neste caso, asumiremos que o resto das ‘causas’ de variación debemos vincularlas ó compoñente aleatorio.

Aínda que a verdadeira análise multivariante é aquela que analiza o comportamento de varias variables dependentes a partir de dúas ou máis variables independentes, na práctica utilízase case exclusivamente na investigación experimental. O máis frecuente en sociolingüística é explorar a análise factorial da varianza, calculando as medias dunha variable dependente para cada un dos grupos das variables independentes e analizando as diferencias observadas entre elas.

Como xa comentamos anteriormente, unha das condicións para poder realizar unha análise da varianza é que a variable dependente sexa cuantitativa. En demolin-güística, e en xeral na investigación non experimental, é moi infrecuente utilizar este tipo de variables como dependentes polo que, en rigor, a análise da varianza ten un interese menor. As variables utilizadas son ordinais na súa maioría e non cumpren

---

<sup>20</sup> A *hipótese nula* é un forma precisa de afirmar que as diferencias que encontramos son debidas ó azar. O que se persegue nunha investigación é comprobar se podemos determinar unha *hipótese alternativa* que corrobore que “algo” distinto do chou é imprescindible para explicar os resultados obtidos. O erro de tipo I cométese ó descartar a *hipótese nula* cando non debería facerse, mentres que o erro de tipo II se produce cando concluímos que o resultado non é significativo e en realidade si que o é. É imposible eliminar os dous tipos de erros. Se pretendemos reducir o erro de tipo I aumentamos as posibilidades de cometer erros de tipo II.

cos requisitos que esixe a utilización da análise da varianza, que son os seguintes: a) suposto de normalidade: a distribución da poboación da que quitamos a mostra é normal; b) suposto de homoscedasticidade: as varianzas dos grupos cos que se traballa son iguais.

Sen embargo, cando as mostras son grandes ( $N > 30$ ), este tipo de probas non se ven moi afectadas malia a que non se cumpran os supostos necesarios (Sánchez Carrión, 1995: 334 e ss.). Ademais, como xa se dixo, a solución máis habitual pasa por dar un salto de nivel e considerar as escalas ordinais coma se fosen de intervalo, asumindo que as respostas dadas polos individuos poden ser interpretadas como puntuacións que permitan calcular medias. Isto non quere dicir, por poñermos un exemplo, que un individuo que diga que fala *só galego* (puntuación 4 nunha escala de 4 = *só galego* a 1 = *só castelán*) fale o dobre de galego que alguén que declare que a súa conducta lingüística é *máis castelán ca galego* (puntuación 2).

Unha vez que temos as medias da variable dependente para cada un dos valores das variables independentes, calquera programa estatístico dispón de varias probas que deciden se as diferencias son significativas e se, polo tanto, podemos extrapolarlas da mostra coa que traballamos ó conxunto da nosa poboación.

### *Análise factorial*

Denomínase análise factorial á técnica que se aplica para reducir un determinado número de ítems individuais nun número inferior de dimensións latentes (factores). Son varias as razóns polas que podemos decantarnos por unha análise factorial, entre as que destaca a de pretender explicar cómo correlacionan un grupo de variables (Comrey, 1985). Nesta proba, os factores representan as variables orixinais cunha perda mínima de información. Como condición previa para realizar unha análise factorial, as variables deben presentar unha determinada correlación, que permita que sexan consideradas coma subconxuntos. Se isto non fose así, carecería de sentido realizar tal proba. Unha vez que indicamos qué variables serán sometidas á análise, o programa calcula por defecto tantos factores coma autovalores iniciais iguais ou superiores a 1, aínda que esta opción se pode modificar. O normal é traballar con autovalores superiores a 1, xa que canto máis próximo está un autovalor a 1 máis seguros estamos de que cada factor está asociado a unha única variable (a suma de autovalores sempre será igual ó número de variables), polo que non tería sentido facer unha análise factorial.

### *Análise de correlacións*

As correlacións utilízanse para analizar datos obtidos de investigacións descritivas e con elas examinamos a relación existente entre variables. Consideramos, a modo de exemplo, unha investigación que examina a relación existente entre o nivel de estudos e a competencia escrita nunha lingua. Cando se obtén unha *correlación*

*alta* entre estas dúas variables, significa que hai unha relación estreita entre ambas. É dicir, isto significa que canto máis alto é o nivel de estudos dos individuos, maior nivel de competencia escrita teñen nunha lingua. E viceversa. Unha *correlación baixa* entre as variables significará que estas dúas variables non están relacionadas unha coa outra, é dicir, que a competencia lectora non está relacionado co nivel de estudos. Unha *correlación negativa* entre esas variables implicaría que canto menos nivel educativo acaden os individuos, máis alta será a competencia lectora, e neste caso ambas variables estarían inversamente relacionadas.

Para calcular as correlacións, o investigador precisa obter resultados de dous conxuntos de datos procedentes do mesmo grupo de individuos. É dicir, no caso hipotético que comentamos no parágrafo anterior, para examinar a relación entre o nivel de estudo e a competencia lectora nunha lingua debemos ter de tódolos suxeitos da nosa mostra dúas medidas: unha do nivel de estudos e outra de competencia lectora. Dependendo do tipo de datos (numéricos, ordinais, etc.), poderemos aplicar diferentes tipos de fórmulas para calcular a correlación. Os resultados acadados exprésanse mediante un coeficiente de correlación que varía desde  $-1.00$ , que indica unha correlación negativa perfecta, ata  $1.00$ , que indica unha correlación positiva perfecta.

## 8. Informática e investigación cuantitativa

O papel da informática na investigación cuantitativa das ciencias sociais é moi notable<sup>21</sup>. Practicamente en todo o proceso de investigación estamos xa afeitos a recorrer ó uso de ferramentas informáticas, en calquera que sexa a etapa: na preparación do material, na ordenación dos datos acadados, na análise de tales datos e na presentación dos resultados. A utilización de ferramentas informáticas faise especialmente imprescindible na parte analítica de explotación dos datos. Nesta fase, o recurso máis habitual é a estatística, disciplina mediante a cal pasamos dos números illados á súa distribución probabilística na realidade social que estamos a analizar –inferencia. Neste sentido, a estatística cumpriría entendela como unha iniciativa cualitativa.

O programa informático de maior relevancia e extensión na actualidade para a análise estatística neste tipo de investigaciónes é o SPSS (acrónimo de *Statistical Package for the Social Sciences*)<sup>22</sup>. Ademais do sistema SPSS, hai outros programas

---

<sup>21</sup> Véxase Ramallo (1999) para a informática na análise de datos cuantitativos. Para o caso da investigación cualitativa, véxase Lorenzo (1999).

<sup>22</sup> Sobre o SPSS existen diversos manuais que presentan, polo xeral dunha forma gráfica e clara, os procedementos deste programa nalgunha das súas últimas versións. Entre eles, podemos salientar, no ámbito hispánico, Ferrán Aranaz (1997) e Visauta Vinacua (1997), ademais do propio manual do programa. A través de internet pódese consultar o sitio web: [www.spss.com/SPAIN](http://www.spss.com/SPAIN) que ofrece información sobre 30 produtos da familia SPSS, ademais de bibliografía, cursos en liña e a posibilidade de incorporar diversos programas de demostración dos distintos módulos.



estadísticos que poden ser utilizados neste tipo de investigacións relacionados con enquisas. Se o que se necesita é unha análise estatística elemental, algunhas das follas de cálculo máis coñecidas inclúen ferramentas que satisfan esta demanda. Ademais, outros programas específicos son MINITAB, SAS, BMDP (adquirido por SPSS Inc. en 1995), ESP, OSIRIS, etc. Aínda que para a súa utilización se require un certo nivel de coñecemento de fundamentos estatísticos, non é necesario que os ditos coñecementos sexan moi exhaustivos<sup>23</sup>.

### Referencias bibliográficas

- Arias Astray, A. & B. Rodríguez Ramírez (1998). “La encuesta como técnica de investigación social”. En A.J. Rojas Tejada, J.S. Fernández Prados & C. Pérez Meléndez (eds.), 31-49.
- Azofra, M<sup>a</sup> J. (1999). *Cuestionarios*. Madrid: Centro de Investigaciones Sociológicas.
- Blair, F. (1990). *Survey on a Shoestring*. Dallas: Summer Institute of Linguistics/The University of Texas at Arlington.
- Boix i Fuster, E. & F.X. Vila i Moreno (1998). *Sociolingüística de la llengua catalana*. Barcelona: Ariel.
- Bosch, J.L.C & D. Torrente (1993). *Encuesta telefónicas y por correo*. Madrid: CIS.
- Butler, C.S. (1985). *Statistics in Linguistics*. Oxford: Blackwell.
- Carmides, E.G. & R.A. Zeller (1979). *Reliability and Validity Assesment Series: Quantitative Applications in the Social Sciences*. Londres: Sage.
- Cea D’Ancona, M<sup>a</sup> A. (1998). *Metodología cuantitativa. Estrategias y técnicas de investigación social*. Madrid: Síntesis.
- Chambers, J.K. (1995). *Sociolinguistic Theory*. Oxford: Blackwell.
- Combessie, J-C. (2000). *El método en sociología*. Madrid: Alianza.
- Comrey, A.L. (1985). *Manual de análisis factorial*. Madrid: Cátedra.
- Converse, J.A. & S. Presser (1986). *Survey Questions: Handcrafting the Standardized Questionnaire*. Londres: Sage.
- Coulmas, F. (ed.) (1996). *The Handbook of Sociolinguistics*. Oxford: Blackwell.
- Díaz de Rada, V. (2000a). *Problemas originados por la no respuesta en investigación social: definición, control y tratamiento*. Navarra: Universidad Pública de Navarra.

---

<sup>23</sup> A maioría destes programas é modular, polo que o investigador poderá adquirir os compoñentes que realmente serán imprescindibles para o seu traballo. Todos eles teñen grande versatilidade e o seu manexo carece de complicacións, entre outras cousas porque dispoñen dunha interface moi intuitiva. Unha breve comparación entre MINITAB e SPSS, baseada nas versións de hai algúns anos de ámbolos programas, pode atoparse en Butler (1985: 154-167). Pola súa parte, Stevens (1996: 23-34) ofrece un resumo das potencialidades comparativas de SAS e de SPSS.

- Díaz de Rada, V (2000b). "Utilización de nuevas tecnologías para el proceso de 'recogida de datos' en la investigación social mediante encuesta". *Revista Española de Investigaciones Sociológicas* 91, 137-166.
- Díaz de Rada, V. (2001). *Organización y gestión de los trabajos de campo con encuestas personales y telefónicas*. Barcelona: Ariel.
- Dillman, D.A. (1999). *Mail and Internet Surveys: the Tailored Design Method* (2ª edición). Nova York: John Wiley & Sons.
- Dillman, D.A. (2002). "Navigating the rapids of change: some observations on survey methodology in the early 21<sup>st</sup> century". Disponible en <http://survey.sesrc.wsu.edu/dillman/>.
- Ferrán Aranaz, M. (1997). *SPSS para Windows. Programación y análisis estadístico*. Madrid: McGraw-Hill.
- Fink, A. (ed.) (1995). *The Survey Kit*. Thousand Oaks: Sage (9 volumes).
- Fink, A. & J. Kosecoff (1998). *How to Conduct Surveys. A Step by Step Guide*. Thousand Oaks: Sage.
- Foddy, W. (1999). *Como perguntar: teoria e prática da construção de perguntas em entrevista e questionários*. Oeiras: Celta.
- Fowler, F.J. Jr. (1993). *Survey Research Methods*. Newbury: Sage.
- Fowler, F.J. Jr. & F.J. Fowler (1995). *Improving Survey Questions: Design and Evaluation*. Londres: Sage.
- García Ferrando, M. (1994). "La encuesta". En J. García Ferrando, J. Ibáñez & F. Alvira (comp.), 147-176.
- García Ferrando, J., J. Ibáñez & F. Alvira (comp.) (1994). *El análisis de la realidad social. Métodos y técnicas de investigación* (2ª edición revisada e ampliada). Madrid, Alianza Editorial.
- García Marcos, F. (1999). *Fundamentos críticos de sociolingüística*. Almería: Universidad de Almería.
- González Gómez, A. & J.L. Padilla García (1998). "La entrevista". En A.J. Rojas Tejada, J.S. Fernández Prados & C. Pérez Meléndez (eds.), 141-153.
- Hayman, H. (1971). *Diseño y análisis de las encuestas sociales*. Bos Aires: Amorrortu.
- Ibáñez, J. (1985). *Del algoritmo al sujeto. Perspectivas de la investigación social*. Madrid: Siglo XXI.
- Kish, L. (1995). *Diseño estadístico para la investigación*. Madrid: CIS.
- Latiesa, M. (1994). "Validez y fiabilidad de las observaciones sociológicas". En J. García Ferrando, J. Ibáñez & F. Alvira (comp.), 335-364.
- Lorenzo Suárez, A.M. (1999). "Sociolingüística cualitativa y lingüística informática". En J. Gómez Guinovart, A.M. Lorenzo Suárez, J. Pérez Guerra & A. Álvarez Lugiés (eds.). *Panorama de la investigación en lingüística informática* (número monográfico da *Revista Española de Lingüística Aplicada*). Logroño: Asociación Española de Lingüística Aplicada, 247-262.

- Manzano, V.G. (1998). “Selección de muestras”. En A.J. Rojas Tejada, J.S. Fernández Prados & C. Pérez Meléndez (eds.), 51-93.
- Manzano, V.G., A.J. Rojas & J.S. Fernández (1996). *Manual para encuestadores*. Barcelona: Ariel.
- Manzano, V.G., & A. González Gómez (1998). “Selección del encuestado”. En A.J. Rojas Tejada, J.S. Fernández Prados & C. Pérez Meléndez (eds.), 99-113.
- MacRae, S. (1995). *Modelos y métodos para las ciencias del comportamiento*. Barcelona: Ariel.
- Mayntz, R., K. Holm & P. Hübner (1975). *Introducción a los métodos de la sociología empírica*. Madrid: Alianza.
- Mesthrie, R., J. Swann, A. Deumert & W. L. Leap (2000). *Introducing Sociolinguistics*. Edimburgo: Edinburgh University Press.
- Moreno Fernández, F. (1998). *Principios de sociolingüística y sociología del lenguaje*. Barcelona: Ariel.
- Navarro, P. (1995). “La encuesta como texto: un enfoque cualitativo”. Comunicación presentada no *V Congreso Español de Sociología*, Granada, setembro de 1995. Disponible en <<http://www.netcom.es/pnavarro/Publicaciones/EncuestacomTexto.html>>.
- Paterson, R.A. (1999). *Constructing Effective Questionnaires*. Thousand Oaks: Sage.
- Padilla García, J.L., A. González Gómez, & C. Pérez Meléndez (1998). “Elaboración del cuestionario”. En Rojas Tejada, A.J., J.S. Fernández Prados & C. Pérez Meléndez (eds.), 115-140.
- Pires de Lima, M. (2000). *Inquérito sociológico. Problemas de metodologia* (5ª edición). Lisboa: Presença.
- Ramallo, F. (1999). “Informática y sociolingüística cuantitativa”. En J. Gómez Guinovart, A.M. Lorenzo Suárez, J. Pérez Guerra & A. Álvarez Lugrís (eds.). *Panorama de la investigación en lingüística informática* (número monográfico da *Revista Española de Lingüística Aplicada*). Logroño: Asociación Española de Lingüística Aplicada, 263-290.
- Rodríguez Osuna, J. (1991). *Métodos de muestreo*. Madrid: Centro de Investigaciones Sociológicas.
- Rodríguez Osuna, J. (1994). “La muestra: teoría y aplicación”. En J. García Ferrando, J. Ibáñez & F. Alvira (comp.), 365-402.
- Rojas Tejada, A.J., J.S. Fernández Prados & C. Pérez Meléndez (eds.) (1998). *Investigar mediante encuestas. Fundamentos teóricos y aspectos prácticos*. Madrid: Síntesis.
- Romaine, S. (1996). *El lenguaje de la sociedad*. Barcelona: Ariel.
- Sánchez Carrión, J.J. (1995). *Manual de análisis de datos*. Madrid: Alianza.
- Sánchez Carrión, J.J. (2000). *La bondad de la encuesta: el caso de la no respuesta*. Madrid: Alianza.

- Seminario de Sociolingüística (1994). *Lingua inicial e competencia lingüística en Galicia*. A Coruña: Real Academia Galega.
- Seminario de Sociolingüística (1995). *Usos lingüísticos en Galicia*. A Coruña: Real Academia Galega.
- Seminario de Sociolingüística (1996). *Actitudes lingüísticas en Galicia*. A Coruña: Real Academia Galega.
- Stevens, J. (1996). *Applied Multivariate Statistics for the Social Sciences*. Nova Jersey: Lawrence Erlbaum Ass.
- Visauta Vinacua, B. (1997). *Análisis estadístico con SPSS para Windows*. Madrid: McGraw-Hill.
- VV.AA (1973). *Encuesta a los trabajadores*. Madrid: Castellote.
- Wardhaugh, R. (1992). *Introducción á sociolingüística*. Santiago de Compostela: Universidade.
- Wert, J. I. (1994). “La encuesta telefónica”. En J. García Ferrando, J. Ibáñez & F. Alvira (comp.), 177-188.