

*Supplements of the Anuario de Filología Vasca "Julio de Urquijo", LIII*

## TOOLS FOR LINGUISTIC VARIATION

Gotzon Aurrekoetxea, Jose Luis Ormaetxea  
(eds.)

orriari la zabel zazu



Universidad Euskal Herriko  
del País Vasco Unibertsitatea  
Bilbao Bilbo  
2010



INTRODUCCIÓN A LOS PROBLEMAS Y MÉTODOS  
SEGÚN LOS PRINCIPIOS DE LA ESCUELA  
DIALECTOMÉTRICA DE SALZBURGO  
(CON EJEMPLOS SACADOS  
DEL “ATLANTE ITALO-SVIZZERO”, AIS)

Hans Goebel

Universidad de Salzburgo

**Abstract**

*This paper documents the many taxometric and cartographic achievements of the Salzburg school of dialectometry. The paper discusses the following topics: 1) problems of measurement (“taxatation”) of linguistic atlas data (with particular consideration of Romance linguistic atlases), 2) establishment of the data matrix, 3) choice of the similarity index (here: Relative and Weighted Identity Value), 4) generation of the respective similarity and distance matrices, 5) their subsequent cartographic exploitation, which encompasses the following cartographic tools: similarity maps, parameter maps, interpoint maps, dendrograms (and their spatial projection), and correlation maps. The ultimate purpose of these highly sophisticated cartographic tools (choropleth and isopleth maps) is to increase our knowledge of the complex mechanisms of the “dialectal management of space by man”. From a methodological point of view our paper deals with problems related to Romance dialectology and linguistic geography, historical linguistics, numerical classification, statistics and statistical cartography.*

*The examples are drawn from the Italian linguistic atlas AIS (Atlante-Italo-Svizzero, recte: Sprach- und Sachatlas Italiens und der Südschweiz published by Karl Jaberg and Jakob Jud, Zofingen: Ringier, 1928-1940, 8 volumes) whose data have been dialectometrically analyzed between 2007 and 2009.*

*The taxometric calculations and their respective visualizations are realized by a powerful computer program called “Visual DialectoMetry” (VDM), created by Edgar Haimmerl between 1997 and 2000 in Salzburg, which is freely available for research purposes.*

**Key words:** *dialectometry, geolinguistics, linguistic geography, dialectology, cartography, numerical classification.*

**1. Prólogo**

Es bien sabido que el AIS (= “Atlante italo-svizzero”) es el segundo atlas lingüístico, por lo que hace a su importancia, de la filología románica. Fue elaborado por dos romanistas suizos —Karl Jaberg y Jakob Jud— de 1919 al 1927 y publicado en

ocho volúmenes entre los años 1928 y 1940. La realización de las encuestas dialectales estuvo a cargo de tres lingüistas de renombre y con una gran experiencia: de la parte norte y central de la red del AIS se encargó Paul Scheuermeier (1888-1962), quien además era un excelente etnógrafo y un gran fotógrafo; del sur de Italia y Sicilia se encargó Gerhard Rohlfs (1892-1986) —más tarde fue catedrático de filología románica en Tübinga y Munich y llegó a ser uno de los más grandes romanistas— y de Cerdeña se encargó Max Leopold Wagner (1880-1962), en cuyo honor se han nombrado actualmente en ella calles y plazas.

La palabra y el método *dialectometría* fueron, como es bien sabido, acuñados en el año 1973 por el dialectólogo tolosano Jean Séguy (1914-1973). Séguy fue también el autor del “Atlas linguistique et ethnographique de la Gascogne” (ALG): ante la enorme variabilidad interna de los datos del ALG, se sintió retado a buscar una posibilidad que con métodos cuantitativos le permitiera comprender de forma global esa variabilidad, es decir, desistiendo de muchos detalles cualitativos.<sup>1</sup>

Séguy se introdujo más de forma intuitiva que premeditada en el camino de la “taxonomía numérica”, con cuya utilización ya se habían obtenido a principio de los años setenta del siglo pasado grandes éxitos en el análisis cuantitativo de masas de datos biológicos, económicos y psicológicos. Debido a su temprana muerte (1973), Séguy lamentablemente no pudo llegar a madurar su teoría.

Mis primeros intentos dialectométricos empezaron paralelamente a los de Jean Séguy a principios de los años setenta<sup>2</sup> y desde un principio estuvieron explícitamente marcados por la clasificación numérica<sup>3</sup> y la visualización sistemática de los resultados obtenidos a través de la utilización consecuente de la entonces todavía muy joven informática.

Junto a la dialectometría existen actualmente una multiplicidad de otras *metrías*, las cuales comparten todas las siguientes específicas características metodológicas:

- a) El tratamiento de datos masivos.
- b) El esfuerzo en recoger y descubrir de forma cuantitativa, es decir, con la ayuda de las matemáticas y la estadística, las estructuras ocultas y las regularidades subyacentes en el conjunto de los datos analizados.
- c) El afán en medir primero de forma adecuada la masa de datos analizada, lo cual está vinculado a una transferencia de la información que, a veces, puede resultar muy difícil.
- d) La dependencia de los métodos cuantitativos aplicados a los principios y planteamientos de cada una de las respectivas disciplinas. Así como la psicometría depende y viene determinada por la psico-logía, y la socio-metría por la socio-logía y la econo-metría por la econo-mía, del mismo modo depende la dialecto-metría de la dialecto-logía y sigue sus pautas.

---

<sup>1</sup> Xavier Ravier da cuenta de la enunciación siguiente de Jean Séguy hecha poco antes de su muerte (en 1973): “Désormais je peux crever tranquille. L'idée fixe qui me hantait depuis trente ans est réalisée: à partir de mille millions de chiures de mouche scrupuleusement intégrées, arriver, par une série d'abstractions à la fois mathématiques et réalistes, à faire tenir le gascon dans une formule ou un schéma.” (Ravier 1976: 390).

<sup>2</sup> Véanse mis contribuciones de 1971, 1975 y 1976.

<sup>3</sup> Cfr. las obras todavía fundamentales de Sneath-Sokal (1973), Bock (1974) y Chandon-Pinson (1981).

La dialectometría no es en ningún caso —así lo veo yo por lo menos desde el punto de vista de Salzburgo— una subdisciplina de la estadística o de la geografía cuantitativa, sino de la dialectología o (dicho de forma más exacta) de la dialectología o la geografía lingüística románica. Subrayo de forma especial la palabra *románica*, porque entre la dialectología o la geografía lingüística de los romanistas, los germanistas y los anglicistas, etc., existen ciertas diferencias que no deben ser pasadas por alto ni ignoradas.

La dialectometría que se lleva a cabo en Salzburgo se basa en un fundamento teórico específico y una no menos específica superestructura metodológica.<sup>4</sup> Desde el punto de vista teórico parto de la idea de que la multiplicidad dialectal de nuestros países es el resultado de una actitud lingüística especial de sus habitantes respecto al espacio por ellos habitado. Parece ser que las personas no sólo gestionan de forma especial el espacio por ellas habitado con el trabajo de sus manos, sino que además también lo hacen a través de su “*faculté langagière*”. Por eso hablo yo desde hace algunos años de una “gestión basilectal del espacio por parte del *Homo loquens*”.

Aunque los principios teóricos y metodológicos de la dialectometría practicada en Salzburgo fueron publicados ya en el año 1984 en mi trabajo de habilitación con el título de “*Dialektometrische Studien*”, fue a partir del año 1999 cuando los métodos definidos llegaron, desde el punto de vista informático, a “alzar el vuelo”. Fue entonces cuando fueron implementados por mi amigo Edgard Haimerl en un genial programa informático llamado “*Visual Dialectometry*” (VDM). Todos los gráficos dialectométricos de este artículo han sido generados mediante el uso del VDM.

Ya he mencionado que la dialectometría de Salzburgo se basa en la geografía lingüística románica<sup>5</sup> e intenta enriquecerla y desarrollarla con los conocimientos de la dialectometría. Una de las características principales de la geografía lingüística románica es el uso intensivo de mapas. Este uso tiene sus inicios en la utilización de “mapas mudos” con los que desde un principio se evaluó el atlas lingüístico más importante de los estudios románicos, es decir, el “*Atlas Linguistique de la France*” (ALF),<sup>6</sup> y termina con la utilización de los mapas de similitud, los mapas de parámetros, los mapas interpuntuales, los mapas de parámetros y los mapas correlativos que van a ser presentados y comentados a continuación. En realidad, en la romanística, desde hace más de cien años, cada geolingüista debería ser un buen cartógrafo.

Al final de esta presentación expondré la quintaesencia de mi exposición resumiéndola en siete tesis.

---

<sup>4</sup> Para una buena comprensión del desarrollo y la situación actual de la dialectometría de Salzburgo pueden verse —con preferencia a las contribuciones escritas en lenguas románicas— mis siguientes libros o artículos: (1981), (1983), (1984) —libro capital desde el punto de vista metodológico—, 1987, 1992, 2003, 2005 y 2008. Véanse también los trabajos dialectométricos de Gotzon Aurrekoetxea (1992), Roland Bauer (2009), Lluís Polanco Roig (1984), Xulio Fernández Sousa (2006) y Paul Videsott (2009).

<sup>5</sup> Para una visión global de esta véase el primero volumen de Pop (1950).

<sup>6</sup> Cfr. la presentación de mapas mudos para la explotación geolingüística de los mapas originales del ALF, hecha por Karl Jaberg ya en el año 1906. Véase también el libro introductorio al ALF del mismo autor publicado dos años más tarde (Jaberg 1908) del cual existe igualmente una versión española de 1959.

## 2. De los mapas-AIS originales a los mapas de trabajo A: algunos principios cartográficos

Los ocho volúmenes del AIS contienen 1705 mapas de gran formato, cada uno de los cuales nos posibilita un fascinante vistazo a la dinámica y biología de los dialectos de la Italo-, Sardo- y Retorromania. Pero yo no puedo, ni quiero conformarme, como dialectómetra, con el estudio de la variación de mapas aislados del atlas, sino que mi intención es la de ir a la búsqueda —tal como hizo Séguy con el ALG— de una variación global más amplia, que incluya la variación interna total del conjunto de los 1.705 mapas del AIS.

Cuando se tiene entre manos un objetivo tan amplio, se tiene que llevar a cabo con métodos escogidos de manera precisa y exacta, lo cual en algunos casos sobrepasa de forma amplia los límites de la lingüística. Estos métodos son, sobre todo, de naturaleza estadística y cartográfica. Voy a empezar con la exposición de los métodos cartográficos.

Contemplando la red del AIS como un espacio, es decir, como una parte de la superficie terrestre habitada por los humanos, se puede imaginar que este espacio fue gestionado (y lo sigue siendo) no sólo desde el punto de vista agrícola, urbanístico o económico, sino además desde el punto de vista dialectal. En el marco del AIS se llevó a cabo la investigación de la gestión basilectal de Italia dentro de una red de 407 puntos de encuesta. En estas 407 localidades se utilizó —lamentablemente no en todas— el cuestionario *normal* del AIS,<sup>7</sup> de manera que, debido a las lagunas de los datos correspondientes, siempre muy molestas desde el punto de vista taxométrico, se tuvo que renunciar, durante la dialectometrización integral del AIS, a los datos de 29 localidades.<sup>8</sup> Quedan así 382 localidades, que pueden representarse como los nudos de una red que pueden por una parte *triangularse* y por la otra *poligonizarse*.

En el lenguaje especializado de la cartografía y de la geometría se habla de “triangulación de Delaunay”. Existe además el proceso de *poligonización*, que para nuestro objetivo es mucho más importante. El nombre técnico es “poligonización de Voronoi”. Entre la *triangulación* y la *poligonización* existen simples relaciones geométricas.<sup>9</sup>

En un primer momento se triangula la red objetivo de nuestro análisis; después se trazan sobre los lados del triángulo las mediatrices. Estas mediatrices se alargan hasta que se encuentren con las restantes mediatrices. De esta manera surgen los puntos de los ángulos de los polígonos. El resultado final es la teselación de Voronoi.

<sup>7</sup> Véanse a este propósito la versión original de la introducción al AIS de Karl Jaberg y Jakob Jud de 1928 y también la traducción italiana publicada en 1987.

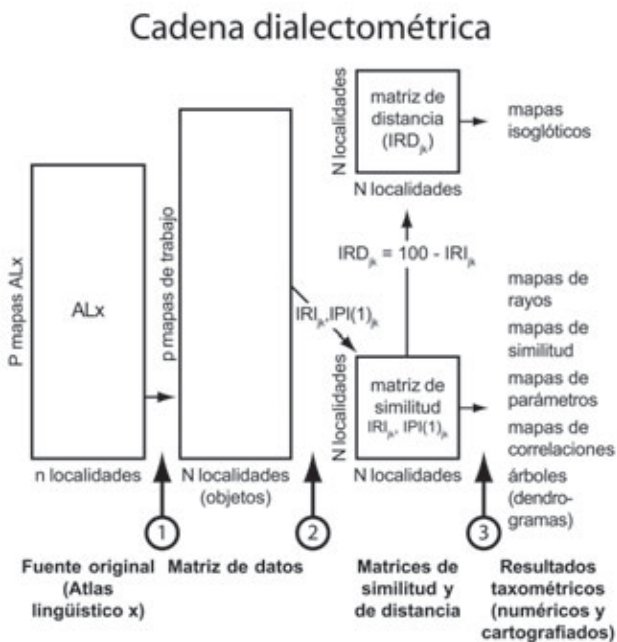
<sup>8</sup> La obtención de la cifra 382 resulta todavía más compleja. Es preciso añadir, a los 407 puntos de encuesta del AIS, cinco puntos suplementarios, correspondientes a las ciudades de Bolonia, Florencia, Milán, Turín y Venecia, donde fueron llevadas a cabo dos encuestas normales, sociolingüísticamente diferenciadas, y dos puntos artificiales equivalentes al italiano y al francés. De la suma así calculada ( $407 + 5 + 2 = 414$ ) hay que deducir tres encuestas no románicas (griego: PP-AIS 748 y 792; albanés: P-AIS 751) y 29 encuestas realizadas por medio de un cuestionario reducido:  $414 - 32 = 382$ .

<sup>9</sup> Para todos los problemas geométricos y computacionales relativos a la triangulación de Delaunay y a la poligonización de Voronoi véase el libro fundamental de Okabe, Boots y Sugihara (1992).

El procedimiento aquí descrito fue empleado por primera vez en el marco de la geografía lingüística general ya en el año 1898 por el germanista Karl Haag, pero fue lamentablemente olvidado posteriormente.

### 3. De los mapas-AIS originales a los mapas de trabajo B: algunos principios de metrología y de taxación

Paso ahora a presentar los métodos dialectométricos utilizados en Salzburgo, que en la práctica constituyen una forma de cadena: véase la figura 1. A la izquierda el rectángulo representa cada uno de los atlas lingüísticos que va a ser dialectometrizado; aquí se trata del AIS. A partir de una medición basada en principios filológicos tradicionales<sup>10</sup> es posible obtener una matriz de datos de los datos del AIS que está formada por N puntos de encuesta y p mapas de trabajo.<sup>11</sup> Aquí se trata de 382 loca-



Flecha 1: aplicación de la taxación

Flecha 2: selección de la medida de similitud (o de distancia)

Flecha 3: determinación de la explotación taxométrica y visual deseada

**Figura 1**

Esquema de los métodos dialectométricos utilizados por la Escuela Dialectométrica de Salzburgo

<sup>10</sup> Véase la flecha 1 en la figura 1.

<sup>11</sup> Para una discusión y explicación del concepto de “mapa de trabajo” cfr. Goebel (1984: I, 31-40) y (2008: 33-35).

lidades y de 3.911 mapas de trabajo. Desde el punto de vista de la estadística, a cada mapa de trabajo corresponde un atributo nominal. Además, la matriz de datos con sus 3.911 atributos nominales es bastante grande. Los resultados obtenidos a partir de ella son, por tanto, muy fiables.

Después de elaborar la matriz de datos, ésta tiene que ser evaluada por medio de una medida de similitud adecuada.<sup>12</sup> Para esto la estadística nos ofrece muchas posibilidades, entre las cuales debemos escoger, como dialectómetra, la adecuada a los propios objetivos. Para nuestro objetivo se han mostrado como especialmente adecuadas las medidas IRI (“índice relativo de identidad”)<sup>13</sup> e IPI (“índice ponderado de identidad”).<sup>14</sup>

Con el IRI o IPI se calcula una matriz de similitud que tiene una dimensión de N por N y por lo tanto es cuadrada. Por medio de una sencilla transformación ( $s + d = 1$ ) se puede derivar de la matriz de similitud (s) una matriz de distancia (d).

La matriz de similitud pasa a contener la totalidad de las variaciones cualitativas recogidas primero en la matriz de datos, pero en forma cuantitativa compacta. El principio de la compactación cuantitativa de la información cualitativa es muy importante; ahora bien, su comprensión no es fácil. Muchos humanistas no pueden o no quieren comprender ni aceptar este principio. Que esto es así ya lo constaté en los años 70. Hoy en día es todavía así, lamentablemente, en muchos casos. Para comprender la dialectometría se debe, por tanto, comprender el principio de transferencia de la información del nivel de la cualidad al de la cantidad.

A la derecha de la figura 1 se pueden ver diferentes valoraciones cuantitativo-cartográficas que nosotros empleamos en Salzburgo para visualizar nuestros cálculos dialectométricos.<sup>15</sup> La estadística y también la cartografía o la visualística moderna tienen disponible todavía muchas más posibilidades de explotación de las matrices de similitud. Nosotros en Salzburgo, de manera consciente, siempre nos hemos limitado a unas pocas de estas posibilidades especialmente escogidas. Ello se debe a que estas se adaptan bien a los planteamientos de la geografía lingüística románica y permiten así tender un puente claramente reconocible entre el saber tradicional de la geografía lingüística románica y el conocimiento nuevo de la dialectometría. Además, contemplamos el espacio a analizar como invariable y las diferentes formas en que se manifiesta la “gestión basilectal del espacio” como las variables a analizar.

¡Empecemos nuestro paseo en el marco de la cadena dialectométrica a la izquierda, es decir, en el momento de la realización de los llamados “mapas de trabajo”!

Se trata de una actividad que en el marco de la geografía lingüística románica se ha realizado desde hace más de cien años,<sup>16</sup> con gran éxito, miles de veces. No sólo existen en nuestra disciplina muchas publicaciones al respecto sino que además hay un amplio consenso sobre cómo y por qué se realizan estos análisis. Se pueden llevar a cabo taxaciones fonéticas y léxicas: véanse los mapas 1 y 2. En Salzburgo denominamos “taxación” al proceso de realización de los mapas de trabajo nominales. Cada mapa de tra-

<sup>12</sup> Véase la flecha 2 en la figura 1.

<sup>13</sup> Para la explicación y definición del IRI cfr. Goebel (1981: 357-361) y (1984: I, 75-79).

<sup>14</sup> Para la explicación y definición del IPI cfr. Goebel (1987: 67-79).

<sup>15</sup> Véase la flecha 3 en la figura 1.

<sup>16</sup> Véanse a este propósito los mapas tipificados (y coloreados) en Brun-Trigaud, Le Berre y Le Dù (2005) y en Veny (2007-2009).



bajo contiene un número exactamente definido de “taxatos” a los cuales corresponden áreas geográficas más o menos grandes y dotadas de configuraciones muy variables.

Una taxación fonética se puede efectuar sólo con aquellos mapas del AIS que se basan en un mismo étimo. El mapa 1 se refiere al mapa 27 del AIS (*il suo cognato*) y al étimo latino Cogná tu para “cuñado”. En este caso se analizó solamente la U latina final. El mismo mapa del AIS permitiría analizar también el desarrollo de la C inicial, de la o pretónica, del grupo -gn-, de la a acentuada o de la -t- intervocálica. Se puede apreciar, pues, que mapas del AIS con una etimología homogénea se pueden taxar de múltiples maneras desde el punto de vista fonético. En nuestro análisis dialectométrico del AIS hemos extraído a partir de 257 mapas originales del AIS, cada uno de ellos con un mismo étimo, un total de 1766 mapas de trabajo fonéticos.

El mapa 2 constituye uno de los 1225 mapas de trabajo léxicos, que hemos realizado en Salzburgo. Se trata de las denominaciones utilizadas para *sobrino*, que se encuentran en el mapa 22 (*la vostra nipote*) del AIS.

Cuando se interpreta la estratificación geográfica de estos taxatos a la luz de la lingüística histórica, se puede reconstruir el surgimiento de esta estratificación. Cada una de las taxaciones de este tipo nos cuenta en cierta manera una pequeña historia. Se puede suponer que de la síntesis de muchas de estas historias parciales se puede llegar a conseguir algo semejante a una historia global del espacio analizado.

Téngase en consideración que lo que se ve en los mapas de trabajo está basado sobre todo en una multitud de diferentes áreas. En el mapa 1 figuran diez áreas de diferentes tamaños y de muy diferentes configuraciones geográficas. Las diferencias de tamaño son fáciles de indicar: el área más grande —aquí con el número uno, de color rosa— comprende 198 puntos de encuesta, es decir 198 polígonos. Las áreas más pequeñas —aquí con los números 10 y 11 y señaladas con tonos azules— están constituidas por un único punto de encuesta o por un solo polígono.

Cada uno de los mapas de trabajo tiene, por lo que se refiere al número de sus taxatos (o áreas), así como a la extensión y la forma de las respectivas áreas en el espacio, un perfil diferente. En el marco de la dialectometría de Salzburgo es muy importante el análisis exacto de las áreas y de sus dimensiones.

El análisis cuantitativo de las áreas taxatorias de las cuales dispone cada uno de los puntos de una red de atlas, es muy clarificador para una mejor comprensión de los mecanismos y regularidades que determinan el funcionamiento comunicativo de la respectiva red. Se puede decir metafóricamente hablando que cada área taxatoria funciona como una “oportunidad comunicativa”. Cuanto más grandes son las superficies medias de las áreas taxatorias de una localidad, tanto mejores son sus oportunidades comunicativas. La figura 2 visualiza los patrimonios areales de dos puntos de nuestra red. El punto 1 (Breil/Brigels), ubicado en los Grisones, dispone de oportunidades comunicativas muy inferiores a las del punto 376 (Venecia). Esto se explica por el hecho de que tiene en su patrimonio areal muchas áreas de tamaño muy reducido: véase la primera columna del histograma con 1288 áreas cuya extensión varía entre 1 y 38 (= 10% de 382) puntos, mientras que en el patrimonio areal de Venecia se encuentran sólo 571 áreas de tamaño muy reducido y, por el contrario, muchas áreas de tamaño más grande: véanse al respecto las tres últimas columnas del histograma de Venecia (a la derecha) que señalan la presencia de 411, 610 y 799 áreas cuyas extensiones varían entre 271 (= 71% de 382) y 308, 309 (= 81% de 382) y 347 así

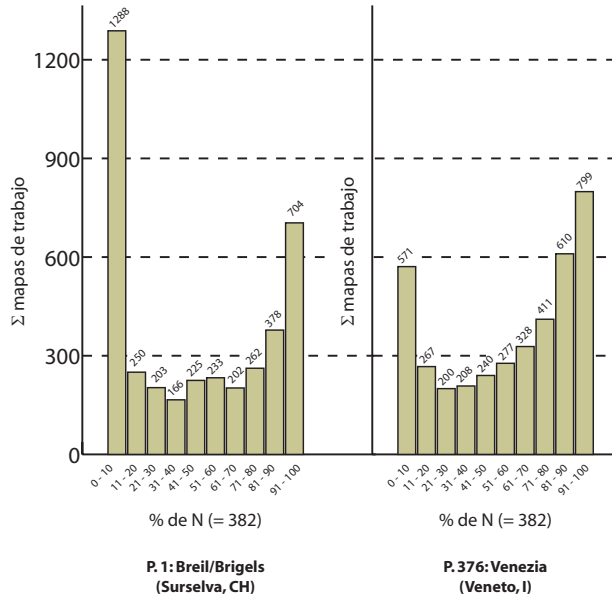


Figura 2

Histogramas relativos a la composición interna de los patrimonios areales de los puntos-AIS 1 (Breil/Brigels) y 376 (Venecia)

como 348 (= 91% de 382) y 382 puntos respectivamente. El rendimiento análogo del punto 1 es, de lejos, inferior.

La determinación de los taxatos y al mismo tiempo de sus áreas sólo puede llevarse a cabo aplicando los conocimientos que provienen del fondo clásico de la lingüística románica. Se puede decir entonces que la dialectometría de la Escuela de Salzburgo, en lo que se refiere a sus fundamentos, se sustenta sobre dos pilares metodológicos:

- 1) áreas geolingüísticas (de carácter nominal);
- 2) la participación, en el proceso de constitución de la matriz de datos, de los conocimientos especializados (*expert knowledge*) de la lingüística románica.

Esto es lo que la diferencia, fundamentalmente, de otras escuelas dialectométricas como, por ejemplo, de la de Groninga. Pero pasemos ahora a la medición de similitud.

#### 4. De la matriz de datos a la matriz de similitud

Como índice estándar de la medición de la similitud se puede considerar el IRI (“índice relativo de identidad”),<sup>17</sup> que constituye la mejor medida de similitud de

<sup>17</sup> El nombre originario alemán de este índice es “Relativer Identitätswert”, el cual se tradujo inicialmente en francés en la forma “índice relatif d’identité” (IRI). Posteriormente las restantes denominaciones románicas fueron acuñadas según la designación francesa.

nuestra dialectometría. Desde el punto de vista matemático el IRI es muy sencillo. Mide la relación entre el número de parejas de mapas de trabajo idénticos y la cantidad de mapas de trabajo disponibles que existen en los vectores de dos puntos de encuesta a comparar. Cuando en un mapa de trabajo aparece el mismo taxato entre los dos puntos de encuesta a comparar, hablamos de una *co-identidad*. En el caso contrario hablamos de una *co-diferencia*. En el *numerador* de la fórmula del IRI se halla siempre el número de las *co-identidades* y en el *denominador* la suma de las *co-identidades* y *co-diferencias*. Los resultados de todos los cálculos IRI se recogen en la matriz de similitud. Ésta es siempre cuadrada, tiene a lo largo de la diagonal siempre el valor uno (o cien) y está compuesta de dos mitades totalmente idénticas. Los valores de un vector de la matriz de similitud son la base de una mapa de similitud.

Véanse los mapas 3 y 4.

El mapa de similitud es el instrumento heurístico más importante, y a la vez el más simple, de la dialectometría de Salzburgo. Al mismo tiempo es la base de todos los complejos instrumentos heurísticos de los que disponemos. Cada uno de los mapas de similitud tiene un punto de referencia que siempre aparece como un polígono blanco: aquí son la ciudad piemontesa de Turín (mapa 3) y la pequeña localidad grisona de Domat/Ems (mapa 4). El resto del mapa muestra la muy variable estratificación del espacio en cuanto a las similitudes cuantitativas de los locoslectos de los restantes 381 puntos de encuesta con respecto al locoslecto de la localidad de referencia. De este modo el mensaje principal de cada uno de los mapas de similitud es cuantitativo, y ya no cualitativo, como es el caso de los mapas de trabajo (véanse los mapas 1 y 2). Los colores ordenados según el arco iris sirven para representar la variación numérica de los valores de medición de la distribución de similitud.

En el mapa 3 los valores de la distribución de la similitud varían entre 45,88% y 83,10%; en el mapa 4 la misma variación se sitúa entre 38,34% y 84,11%. La formación de las seis clases (o intervalos) se lleva a cabo utilizando un algoritmo especial que aquí se llama MINMWMAX. Para la confección de los mapas 17-20 y 27-32 utilizamos el algoritmo MEDMW que se diferencia ligeramente de MINMWMAX creando perfiles coropléticos más accidentados.<sup>18</sup>

Las técnicas de visualización que se verán a continuación no son, en principio, nada especial: se corresponden con los estándares internacionales de la cartografía cuantitativa. A pesar de ello he realizado aquí una selección especial, que se corresponde exactamente con las necesidades de nuestra dialectometría y que en su totalidad fue adoptada por el programa dialectométrico VDM.

El valor heurístico de los mapas de similitud reside en la generación de patrones espaciales cuantitativos, que son típicos para el correspondiente punto de referencia. En el mapa 3 se advierte que las mayores similitudes con el locoslecto de Turín aparecen en el Piemonte y las mayores diferencias con este dialecto se encuentran en Cerdeña, en amplias zonas del sur de Italia, en el Valle de Aosta y en los Grisonos. Son significativos en este sentido los polígonos de color azul claro y azul oscuro. Además todos los polígonos con colores calientes se encuentran por encima de la media arit-

<sup>18</sup> Para una presentación detallada de estos algoritmos cfr. Goebel (1981: 361-368), (1983: 370-376), (1984: I, 93-98) y (1987: 79-83).

métrica de la correspondiente distribución de similitud;<sup>19</sup> los polígonos con colores fríos se encuentran por debajo.

En los mapas 3-14 se constata, además, que la disminución de la similitud lingüística en el espacio se corresponde claramente de alguna manera con el aumento de la distancia del punto de referencia. El patrón de disminución de un mapa de similitud puede calificarse como típico del punto de referencia seleccionado y de sus alrededores. Perfiles coropléticos similares al del mapa 3 se pueden obtener también de los otros puntos de encuesta del Piamonte. Lo mismo vale por lo que respecta al resto de los mapas de similitud.

La cantidad de datos en la que se basan los mapas de este artículo es enorme. Se trata de 3.911 mapas de trabajo y de 43.564 áreas taxatorias pertenecientes a todas las categorías lingüísticas. Además, es importante constatar que los mapas de similitud también se pueden interpretar de forma extralingüística. De esta manera es posible hacer comparaciones analógicas con el comportamiento de abonados de una red telefónica o de misioneros que intentan difundir sus ideas en el interior de una red comunicativa espacialmente definida. Es importante tener en cuenta esto para comprender la relevancia interdisciplinar de la dialectometría.

Consideremos ahora los mapas 4 y 6: cada uno de estos dos mapas de similitud tiene muy diferentes puntos de referencia. El mapa 4 tiene un punto de referencia en la parte central de los Grisones (P. 5: Domat/Ems); el mapa 6, por el contrario, en la ciudad de Venecia (P. 376). Las diferencias entre ambos perfiles coropléticos se corresponden con la totalidad de sus seis intervalos. Mientras en el mapa 4 las similitudes superiores a la media aritmética se extienden desde el lococeto de Domat/Ems hasta aproximadamente los Apeninos septentrionales (es decir la conocida "línea La Spezia-Rimini"), en el mapa 6 las similitudes por encima de la media aritmética abarcan casi la totalidad del norte de Italia e, incluso, la totalidad de la Italia central hasta aproximadamente la "línea Roma-Ancona". Obsérvese, además, con atención en ambos casos la extensión en el espacio de los polígonos de color rojo y naranja: por ellos se reconoce el núcleo del correspondiente tipo dialectal.

Los mapas 7 y 8 muestran de una parte la posición relacional de un dialecto toscano septentrional y de la otra la del italiano estándar, el cual ha sido agregado a nuestra red de investigación como punto artificial. ¡Nótese la práctica identidad de los dos perfiles coropléticos!

Se puede variar la forma de los perfiles coropléticos de los mapas de similitud mediante la selección de otros algoritmos de intervalación u otras medidas de similitud: véanse los mapas 9 y 10.

A la izquierda (mapa 9) se ve un perfil de similitud típico de la Campania, que ha sido realizado mediante el algoritmo de intervalación estándar MINMWMAX. A la derecha (mapa 10) se visualizaron los mismos valores de medición mediante otro algoritmo de intervalación (MEDMW), que hace destacar de forma más clara la división interna del perfil coroplético. El programa VDM nos permite realizar este tipo de variaciones en la visualización de manera muy rápida.

---

<sup>19</sup> En el mapa 3 la media aritmética es de 61,88%. Este valor constituye siempre el umbral superior de la clase 3. En el mapa 4 el mismo parámetro es de 50,13%.

#### 4.1. La posición dialectométrica de una isla lingüística occitana situada en el sur de Italia: Guardia Piemontese (P. 760)

Préstese atención, en los perfiles coropléticos de los mapas 9 y 10, a los polígonos diseminados de color verde y amarillo en medio de la zona roja del sur de Italia y de Sicilia. Se trata en su mayoría de polígonos que pertenecen a las islas lingüísticas de tipo galorrománico (en Sicilia: PP. 865, 836, 817), occitano (en Calabria: P. 760) y francoprovenzal (en Apulia: P. 715).

Ahora pasemos a considerar los mapas 11 y 12, los cuales constituyen un caso muy especial de utilización de la medida de similitud. Se trata de mostrar el grado de integración lingüística de la isla lingüística occitana de Guardia Piemontese en la totalidad de la red del AIS. La localidad de Guardia Piemontese fue fundada en el siglo xiii por colonos del sudoeste del Piamonte, quienes llevaron su dialecto occitanoalpino hacia Calabria. Este dialecto está, pues, desde hace ya más de siete siglos en contacto con su entorno calabrés.

En el mapa 11 se ve un típico perfil de similitud calabrés: las mayores similitudes abarcan hasta la “línea La Spezia-Rimini”, y el respectivo punto de referencia (P. 761-Mangone) se encuentra en los alrededores inmediatos de Guardia Piemontese (P. 760).

A la derecha (mapa 12) se halla el perfil de similitud de Guardia Piemontese, el cual es muy diferente del de Mangone. Las mayores similitudes —representadas por medio de los polígonos de color rojo y naranja— se encuentran en la parte occidental del norte de Italia. El valor de similitud más grande (65,28%) se halla en una localidad (P. 181)<sup>20</sup> que se encuentra apenas a unos 60 kilómetros, desde la que hace más de 700 años fue colonizada Guardia Piemontese.<sup>21</sup> Este ejemplo nos muestra que, en ciertas condiciones, es posible determinar la patria lingüística de una isla lingüística mediante la medición de la similitud. Pero esto presupone que esta patria ha de encontrarse en el interior de la red analizada.

Véanse los mapas 13 y 14.

Sabiendo que la intensidad de un contacto lingüístico es normalmente mayor en el léxico que en la fonética, puede plantearse también en este caso la misma pregunta. Y de hecho es así también aquí: el perfil de similitud de la izquierda (mapa 13) muestra que la vinculación fonética del dialecto de Guardia Piemontese con su antigua patria está todavía totalmente intacta, mientras que el mapa coroplético de la derecha (mapa 14) muestra de manera totalmente clara que —por lo que se refiere al léxico— el contacto lingüístico con el ambiente calabrés ha llevado a una mayor asimilación.

<sup>20</sup> En el mapa 12 el polígono del punto 181 está marcado por un rayado blanco cruzado y una flecha.

<sup>21</sup> Se trata de la parte superior de la Val Pellice (en el Piamonte occidental).

## 5. Al otro lado de la matriz de similitud

Otro capítulo muy interesante de la dialectometría de la Escuela de Salzburgo es el de los “mapas de parámetros”, de los cuales aquí sólo voy a presentar por falta de espacio el mapa de sinopsis de los “coeficientes de asimetría de Fisher”.

En los mapas de parámetros se trata de realizar una sinopsis de determinados parámetros de la distribución de similitud para luego visualizarlos. Así, aparecen perfiles coropléticos muy útiles para aclarar determinados fenómenos de la geografía lingüística.

Véanse otra vez los mapas 3 y 4.

Consideremos de nuevo los mapas 3 y 4 que constituyen visualizaciones de dos distribuciones de similitud diferentes. Lo que aquí nos interesa de manera especial son las diferencias estadísticas que subyacen entre las dos distribuciones de similitud: préstese atención a las diferentes formas de ambos histogramas en la mitad inferior derecha de los dos mapas.

El histograma del mapa 3 muestra una distribución de similitud muy simétrica, mientras que el histograma del mapa 4 muestra una distribución de similitud asimétrica y desplazada hacia la izquierda. Las diferencias en la simetría de una distribución de similitud tienen un significado geolingüístico especial. Nos muestran hasta qué punto un dialecto está integrado en la totalidad de la red, y también si toma parte en muchos fenómenos de contacto lingüístico o si, por el contrario, está aislado. Con la ayuda del “coeficiente de asimetría” propuesto por el estadístico inglés Ronald A. Fisher (1890-1962) [CAF] se puede medir muy bien la variación de la simetría en una distribución de similitud.<sup>22</sup>

Véanse los mapas 15 y 16.

El mapa 15 muestra la visualización del resultado de la sinopsis de los 382 valores del CAF. Para poder comprender claramente su perfil coroplético tan bien estructurado, es necesario poder interpretar correctamente desde el punto de vista lingüístico los valores de medición así como los colores correspondientes: los polígonos azules —los cuales visualizan siempre los valores del CAF más bajos— muestran una gran interacción con el resto de la red del AIS, mientras los polígonos de color rojo —correspondientes a los valores del CAF más altos— muestran por su parte una muy escasa interacción. Denomino esta interacción “compromiso lingüístico”, siguiendo el ejemplo de la germanística, en la que este fenómeno está muy bien descrito bajo el nombre alemán de “Sprachausgleich”.<sup>23</sup>

La ubicación geográfica de los polígonos de color azul oscuro es típica: los encontramos a lo largo de los Apeninos, en Liguria, en los Abruzos y en el Lacio, así como en el interior del norte de Italia a lo largo del río Adigio y justo al sur de los dialectos retorrománicos de los Grisonos y Ladinia. Todas estas zonas siempre han llamado la atención de los dialectólogos por su intenso contacto e intercambio lingüístico.

En el norte de Cerdeña y en el sur de Italia hay también polígonos de color azul oscuro. Los de norte de Cerdeña indican una intensiva toscanización de la zona desde hace alrededor de 300 años, y los de sur de Italia así como los de Sicilia apuntan a las islas lingüísticas galorrománicas (de tipo galoitálico: PP: 865, 836, 817, de tipo occi-

<sup>22</sup> Para una presentación pormenorizada del CAF cfr. Goebel (1981: 394-401) y (1984: I, 150-153).

<sup>23</sup> Véase el libro inspirador de Besch (1967).

tano: P. 760 y de tipo francoprovenzal: P. 715) que allí se encuentran. Puesto que cada isla lingüística tiene que ver *per definitionem* con contacto e intercambio lingüístico, es evidente que allí tienen que aparecer valores de CAF muy bajos: el valor mínimo del CAF se encuentra de hecho en la isla lingüística de Guardia Piemontese (P. 760).

Los valores de CAF muy elevados por encima de cero se corresponden con los polígonos rojos; estos nos muestran las zonas que tienen poco que ver con contacto lingüístico y que, por lo tanto, disponen de una homogeneidad lingüística interior relativamente alta. Añado también que en el mapa 15 la visualización no se refiere a seis, sino a ocho clases de valores: esto permite hacer más evidentes las múltiples gradaciones del efecto del “compromiso lingüístico”.

Debido a la fina división del perfil del mapa puede ser preferible realizar una visualización no en la forma acostumbrada de un mapa coroplético sino en forma de un estereograma: véase el mapa 16 de la misma página.

Los mapas 17 y 18 muestran que los perfiles coropléticos logrados mediante la sinopsis de los coeficientes de asimetría de Fisher pueden variar según las diferentes categorías lingüísticas a las que pertenecen los mapas de trabajo analizados. Desde el punto de vista lingüístico las diferencias entre los mapas 17 y 18 son muy interesantes: mientras en el mapa 17 (relativo a datos fonéticos) la ubicación de los polígonos de color azul oscuro crea algunas agrupaciones lineares que fraccionan ante todo el espacio central de Italia de manera muy clara, en el mapa 18 (relativo a datos léxicos) los polígonos de color azul oscuro —típicos de un alto nivel de interacción lingüística— se concentran sobre todo en Emilia, Romaña y Liguria. Surge así la certeza de que los intercambios léxicos (mapa 18) tienen una implantación espacial diferente de sus equivalentes fonéticos (mapa 17).

## 6. La dialectometría interpuntual: mapa isoglótico y mapa de rayos

Véanse los mapas 19 y 20.

El mapa 19 constituye una síntesis dialectométrica de isoglosas, cuyo mensaje se corresponde de manera amplia con los obtenidos por la síntesis de isoglosas tradicionales como las que, por ejemplo, fueron realizadas para la red del AIS por los lingüistas Gerhard Rohlfs (1947) o Robert A. Hall Jr. (1943). El mapa isoglótico de tipo dialectométrico es naturalmente mucho más exacto y rico en detalles. Además se basa en muchos más datos de los que se pueden abarcar por medio de un análisis hecho a mano.

La totalidad de la sintaxis de la imagen se basa en 970 lados de polígonos, en los que se visualiza un número igual de medidas no de similitud sino de distancia. Estas medidas de distancia se refieren siempre a las diferencias lingüísticas entre dos puntos de encuesta contiguos: de ahí el nombre de *análisis inter-puntual*.<sup>24</sup>

El principio cartográfico es sencillo: cuanto más grandes sean las medidas de distancia, más gruesas y azules se visualizan. Y a la inversa: las medidas de distancia más pequeñas se visualizan más finas y más rojas. Se reconocen efectos muy claros de división del espacio sobre todo por medio del grosor de los lados de los polígonos de color azul oscuro y su distribución espacial; préstese atención a la múltiple división interna del

<sup>24</sup> El término “interpoint” fue acuñado por el lingüista francés Théodore Lalanne (1953: 266).

norte de Italia, la clara separación de las variantes galorrománicas del Valle de Aosta y del Piamonte Occidental y el aislamiento de las variedades retorrománicas en el norte.

Señalo además que la “línea La Spezia-Rimini” está mucho mejor marcada que la no menos citada “línea Roma-Ancona”. Finalmente se puede ver claramente, como era de esperar, que las islas lingüísticas del sur de Italia y del norte de Cerdeña están rodeadas por gruesos haces de isoglosas.

A la derecha (mapa 20) se ve un mapa de rayos, que es la contrapartida cartográfica del mapa isoglótico. La sintaxis de la imagen se basa en 970 lados de triángulo, a lo largo de los cuales se visualiza el mismo número de valores de similitud. La concentración de gruesos lados de triángulo de color rojo nos señala paisajes lingüísticos interpuntuales especialmente bien enlazados. En este sentido, sobresalen claramente la Toscana, parte de Lombardía y el Véneto, así como Sicilia y el sur de Cerdeña. Se pueden apreciar, sin embargo, zonas que están marcadas por un mínimo contacto entre las localidades vecinas: estas son las zonas donde aparecen muchos triángulos finos de color azul. De este modo, allí donde en el mapa de la izquierda (mapa 19) se ven lados de polígono gruesos de color azul, en el mapa de la derecha (mapa 20) se ven lados de triángulo finos de color azul.<sup>25</sup>

Hemos utilizado para la visualización de los cálculos interpuntuales el algoritmo MEDMW, con ocho clases para conferir una estructura más accidentada a los perfiles interpuntuales de los dos mapas. El algoritmo MEDMW utiliza, como el algoritmo MINMWMAX, la media aritmética para la separación de los valores “altos” y “bajos”. Pero en el interior de los recorridos estadísticos situados entre la media aritmética y el máximo de una parte y la media aritmética y el mínimo de otra, el algoritmo MEDMW<sup>26</sup> crea clases con un número de polígonos tan igual como sea posible.

## 7. La dialectometría dendrográfica

Paso ahora a la presentación de la dialectometría dendrográfica, cuya práctica es muy sencilla con el VDM. Los análisis dendrográficos son habituales en muchas ciencias humanas y naturales, pero para poder emplearlos de manera conveniente, hay que conocer bien el funcionamiento de los algoritmos en ellos empleados. Esto es imposible, desde luego, si no se tiene un mínimo de conocimientos matemáticos.

El programa VDM puede, basándose en seis algoritmos dendrográficos diferentes,<sup>27</sup> no sólo calcular y visualizar los respectivos árboles muy rápidamente, sino que, además, puede mostrar el resultado de la clasificación de estos árboles en el espacio (“espacialización”), es decir, proyectarlo sobre el mapa correspondiente. Esto sucede de maneras muy variadas y gracias a la utilización de colores.

<sup>25</sup> Préstese atención al hecho de que las siluetas de los histogramas de los mapas 19 y 20 son enteramente simétricas.

<sup>26</sup> Para el algoritmo MEDMW cfr. Goebel (1984: I, 95) y (1987: 81-82).

<sup>27</sup> Se trata de los algoritmos jerárquico-aglomerativos siguientes: *Single Linkage*, *Complete Linkage*, *Ward Method*, *Simple Average Linkage*, *Average Linkage (UPGMA)* y *Centroid Method*. En la mayoría de los casos los algoritmos *Centroid Method* y *Single Linkage* producen resultados que no son útiles para la dialectometría.



A la izquierda (mapa 21) figura un árbol genealógico calculado con el método de “Complete Linkage”.<sup>28</sup> El cálculo ha sido realizado a partir de los datos de las matrices de similitud que fueron utilizadas anteriormente. En este caso se trata de una matriz de similitud calculada con el índice de similitud IRI y mediante una matriz de datos que comprende nuestro corpus integral de 3.911 mapas de trabajo (los cuales abarcan todas las categorías lingüísticas).

Todos los árboles calculados con algoritmos jerárquico-aglomerativos tienen una estructura binaria y se construyen de izquierda a derecha, es decir, de las hojas a la raíz. Esto se lleva a cabo mediante parejas de fusiones de los elementos más similares desde el punto de vista cuantitativo, siendo el algoritmo utilizado el que define la correspondiente similitud cuantitativa.

El árbol obtenido puede ser interpretado por los lingüistas en dos sentidos: primero de la raíz a las hojas; una interpretación de este tipo es relevante desde el punto de vista diacrónico y se utiliza a menudo en la estadística léxica y en la glotocronología. El otro sentido interpretativo va de las hojas a la raíz. Éste se revela muy importante, pues sirve para mostrar las correspondientes dependencias o similitudes.

En el presente caso he definido en el árbol de clasificación calculado 16 unidades importantes desde el punto de vista geolingüístico, a las que denomino “dendremas”: véase el mapa 21. En la espacialización correspondiente (véase el mapa 22) a cada uno de los dendremas le corresponde un sector espacial muy coherente, al que llamo “corema”. Los dendremas y los coremas de los mapas 21 y 22 tienen siempre el mismo número y el mismo color.

Para interpretar el resultado de un análisis dendrográfico de este tipo es necesario utilizar siempre el árbol y el mapa de forma paralela. Al observar el árbol (mapa 21) se aprecia que en la primera bifurcación después de la raíz (A), de las dos ramas superiores (C y D) cuelgan los dendremas 1-8 y de las dos ramas inferiores (E y F) los dendremas 9-16. Sólo al observar la espacialización (mapa 22) se puede apreciar que en el mapa la línea separadora de estos dos grupos (A versus B) *grosso modo* discurre paralela a los Apeninos, es decir, a la conocida “línea La Spezia-Rimini”. Es, por tanto, el fraccionamiento más importante de nuestra red que se puede encontrar mediante la utilización de este método.

Cuando pasamos en la parte superior del árbol a la bifurcación C, observamos que la totalidad de este maxi-dendrema relativo a la parte superior de Italia se divide en pasos sucesivos en el romanche de los Grisonos y el ladino dolomítico (dendrema 1) y en un gran sub-dendrema, el cual incluye los dendremas (y los respectivos coremas) 2-7 que abarcan el resto del norte de Italia a excepción de las zonas de hablas galorrománicas a lo largo de la cadena alpina occidental (dendrema/corema 8 = rama D).

Interesantes son también las conexiones de zonas complejas desde el punto de vista clasificatorio: así los Grisonos romanches y la Ladinia dolomítica (dendrema/corema 1) están ligados a un bloque nordoccidental de la parte superior de Italia, el de los dendremas 2 a 7 y que contiene Lombardía, el Véneto, Emilia-Romaña y el Friul.

<sup>28</sup> Para una presentación pormenorizada de los algoritmos jerárquico-aglomerativos en general y del algoritmo “Complete Linkage” en particular, véanse Sneath-Sokal (1973: 356 s.), Bock (1974: 356 s.) y Chandon-Pinson (1981: 94 s.).

Las zonas galorrománicas del Piamonte constituyen un propio dendrema-corema, el número 8, que, a su vez, está segregado del resto del Piamonte, el número 3.

No menos interesantes son las conexiones de las islas lingüísticas del sur de Italia: véanse a este respecto los dendremas/coremas 13 (que abarcan los dialectos galoitalicos de los PP. 817, 836, 865) y 14 (que comprende el occitano de Guardia Piemontese [p. 760] y el francoprovenzal de Faeto [P. 715]).

Pero el gran valor heurístico del análisis dendrográfico sólo se puede apreciar cuando se utiliza de forma comparativa y se comparan árboles que, basados en los mismos datos, han sido calculados utilizando diferentes algoritmos.

## 8. La dialectometría correlativa

Llegamos así a un capítulo especialmente sugestivo de la dialectometría de Salzburgo, que existe desde 2004, cuando el creador de VDM, Edgar Haimlerl, lo incluyó en su programa: se trata de la “dialectometría correlativa”.<sup>29</sup> Esta permite comparar respectivamente los  $N$  vectores de dos matrices de similitud (o de proximidad) establecidas, por parejas, mediante un cálculo de las correlaciones y visualiza de forma inmediata los  $N$  valores de correlación obtenidos. Los mapas de correlaciones producidos de esta manera nos informan de si las dos dimensiones correlacionadas gestionan el espacio analizado de manera convergente o de manera divergente. Recuerdo aquí a este respecto el concepto mencionado al principio de “la gestión dialectal del espacio por el hablante”, que ha sido desarrollado en Salzburgo.

Además, la experiencia adquirida hasta el momento en el análisis dialectométrico de diferentes corpóra lingüísticos corrobora la hipótesis de que la gestión dialectal del espacio por el hablante —que en todas las ocasiones nos ha dado claros y muy ordenados perfiles espaciales— viene regida por “leyes espaciales específicas”, semejantes a las regularidades o “leyes” válidas para el cambio fonético o la variación lingüística a lo largo del eje diacrónico.

Como es sabido, las leyes fonéticas fueron descubiertas en el último cuarto del siglo xix por los Neogramáticos de Leipzig, siendo más tarde corroboradas en todas las ocasiones. Las leyes espaciales mencionadas serían la correspondencia diatópica de las leyes descubiertas por los Neogramáticos de Leipzig, las cuales funcionan en la dimensión diacrónica.

Véanse los mapas 23-26.

Demuestro primero las posibilidades de correlación entre el lenguaje y el espacio. En el mapa 23 se observa un nuevo perfil de similitud: se basa en la medición de similitud correspondiente al punto de referencia de Nápoles (en Campania). En el gráfico opuesto (mapa 24) se halla un tipo de mapa que aún no hemos presentado: se trata de un mapa de proximidad. Visualiza la estratificación de las proximidades geográficas a partir de un punto de referencia. En él se calculan las proximidades espaciales —no las distancias— entre los 382 puntos de encuesta del AIS mediante la utilización del teorema de Pitágoras. Para ello se utilizan las coordena-

<sup>29</sup> Véanse a este propósito nuestra contribución introductora de 2005 y los respectivos esquemas de cálculo en Goebel (2005: 328) y (2008: 53).

das  $x$  e  $y$  de todos los puntos de encuesta, las cuales están archivadas desde el principio en la base de datos del VDM. Desde los puntos de vista taxométrico y cartográfico las matrices calculadas de proximidad euclídea se pueden tratar como cualquier otra matriz de similitud lingüística.

¿Qué significa, pues, la diferencia entre ambas visualizaciones? A la izquierda (mapa 23) se encuentra un mapa de similitud basado en los principios de la gestión basilectal de Italia a través de sus habitantes; a la derecha (mapa 24) se halla un mapa que se refiere a la gestión euclídea del mismo espacio. En ambos casos se puede apreciar una disminución de los valores de medición en tanto que aumenta la distancia con respecto a Nápoles (P. 721) como punto de referencia. Cada uno de estos mapas se basa en 381 valores de medición. Cuando correlacionamos las dos series de valores de medición mediante la utilización del coeficiente de correlación  $r(\text{BP})$  de (Auguste) Bravais (1811-1863) y (Karl) Pearson (1857-1936), obtenemos un valor de +0,891, que se puede observar en la parte central de la respectiva página (véanse los mapas 23 y 24). Este valor, en principio, como lingüistas, no nos dice nada.

En los mapas 25 y 26 se puede observar el mismo proceder a partir de un punto de referencia en Friul (P. 318). Las tendencias que aparecen en ambos perfiles coropléticos son las mismas que antes: disminución de los valores de medición con el aumento de la distancia con respecto al punto de referencia: a la izquierda (mapa 25) podríamos decir de forma “humana” y a la derecha (mapa 26), si se quiere, de forma “natural”. El valor de correlación  $r(\text{BP})$  correspondiente es +0,803. Viendo esto nos puede venir la idea de realizar el mismo cálculo de correlación para la totalidad de los 382 puntos de encuesta y visualizar sobre el mapa los 382 valores obtenidos de forma sinóptica.

¿Qué aspecto tiene el perfil coroplético que se obtiene de la correlación de los  $N$  vectores acoplados de una matriz lingüística de similitud y una matriz euclídea de proximidad?

Los mapas 27 y 28 proporcionan la respuesta: constituyen dos visualizaciones que, aun basadas en los mismos datos estadísticos, son diferentes en la fineza de su presentación. La visualización de la derecha (mapa 28) —hecha con ocho clases icónicas— muestra más detalles y una más fina escala de niveles que la de la izquierda (mapa 27), hecha con sólo seis clases icónicas. De nuevo se necesita, para poder interpretar estos dos mapas, una exacta comprensión del significado lingüístico de los valores de medición y de los colores.

En general, el color azul se refiere a valores pequeños —algunas veces incluso negativos— de la medición de  $r(\text{BP})$ . El color rojo se refiere a valores grandes de la medición  $r(\text{BP})$ .

Valores de medición pequeños, es decir, polígonos azules, significan que la gestión lingüística del espacio y la euclídea suceden según principios diferentes y por lo tanto no van a la “misma cadencia” ni en armonía.

Elevados valores de medición, es decir, polígonos de color rojo, significan lo contrario: el hecho de que la difusión de las similitudes lingüísticas en el espacio se sucede de una forma amplia según principios euclidianos o “naturales”.

Allí donde encontramos polígonos azules y, por tanto, tiene lugar una fisura entre las gestiones del espacio y de la lengua, tienen que entrar en juego fuerzas que no son de origen natural, sino que tienen un origen antrópico. Realmente la distribución de los colores rojo y azul es, así como la de todos los semitonos, muy indicativa.

Zonas con elevado disturbio antrópico —aquí de color azul oscuro (correspondiente a las clases icónicas 1 y 2)— se encuentran en el Véneto, el norte de la Toscana y en Liguria, cuando se dejan de lado las cinco islas lingüísticas galorrománicas del sur de Italia y Sicilia, así como la parte norte de Cerdeña. Además, aparece el valor de medición más pequeño de todo el mapa en la isla lingüística de Guardia Piemontese (P. 760), ya conocida. En realidad las islas lingüísticas son productos antrópicos, que *per definitionem* se basan en una desvinculación artificial de las relaciones surgidas de forma natural entre la lengua y el espacio.

El Véneto y muy especialmente el norte de la Toscana son, según nuestro diagnóstico, zonas que en la historia de la lengua italiana tienen que haber jugado el papel de mediadores lingüísticos debido a un fenómeno de importación y exportación lingüística masivo, el cual ha alterado de forma radical su incorporación inicial al espacio de Italia.

De hecho existen puntos de referencia históricos que nos muestran que, especialmente la ciudad de Venecia y sus alrededores, han desarrollado una relación muy estrecha con la Toscana y la Italia media en general, que es totalmente atípica para el resto del norte de Italia. Hay que precisar, sin embargo, que todavía no se sabe con exactitud cuándo sucedió esto. Se puede tratar de un fenómeno acaecido en tiempos de la romanización del Véneto en el siglo II antes de nuestra era. Se puede tratar también, por otra parte, de un fenómeno que se remonta al siglo xiv, cuando la ciudad de Venecia empezó a conquistar su *terraferma*, es decir, el Véneto, y asimilarla lingüísticamente.<sup>30</sup>

Se podría decir que el sur de Italia, con Sicilia y la mayor parte de Cerdeña, se encuentra en una situación de armonía “natural” entre lengua y espacio, mientras que la dinámica evolutiva del desarrollo lingüístico diacrónico se concentra en ambas laderas de los Apeninos del Norte.

Véanse los mapas 29 y 30.

Se puede diferenciar este análisis de nuevo según la fonética y el léxico. En este caso se dan resultados parcialmente diferentes. Mientras que el perfil de la izquierda (mapa 29) se corresponde ampliamente con aquel que ya conocemos, el perfil de la derecha (mapa 30), sobre todo en la zona noroccidental de Italia, tiene claramente otro trazado. Pero el mensaje global permanece invariable: a un sur armónico y quieto desde el punto de vista evolutivo, se contraponen un norte agitado y turbulento desde el mismo punto de vista.

La dialectometría correlativa puede ser empleada también para comparar entre sí categorías intra-lingüísticas como el léxico, la fonética o el vocalismo y el consonantismo. Se puede comprobar así de qué manera estos componentes lingüísticos convergen o divergen en la “gestión dialectal del espacio por parte de los hablantes”.

Véanse los mapas 31 y 32.

Nuevamente se dan aquí perfiles claramente estructurados, que deben de ser interpretados de forma conveniente y comparados con la información ya disponible sobre la historia de la lengua italiana y de sus dialectos. Mientras que en el mapa 31 (a la izquierda) aparece un gran dinamismo en el norte y el este de la Toscana, éste aparece en el mapa 32 (a la derecha) de nuevo en el Véneto, Liguria, la franja de transición entre Roma y Ancona, y en algunas islas lingüísticas del sur de Italia.

<sup>30</sup> Véanse, en este sentido, las respectivas discusiones en Goebel (2008: 58-61).

De hecho la ya mencionada “centralización” del Véneto parece haber sucedido sobre todo en el ámbito del vocalismo, que así se ha alejado del consonantismo por lo que se refiere a su típica gestión del espacio. Estas conclusiones se deducen de la ubicación de las zonas de color azul oscuro en el mapa 32.

## 9. Observaciones finales

Llegado al final de mi exposición termino con la presentación de las siete tesis siguientes.

1. El uso de la “*faculté langagière*” del *Homo loquens* está estrechamente vinculado a las dimensiones naturales del espacio geográfico.
2. El *Homo loquens* se encuentra, pues, en la necesidad de apropiarse —es decir: de gestionar— lingüísticamente el espacio geográfico.
3. La gestión lingüística del espacio lingüístico por parte del *Homo loquens* se efectúa gracias a todos los aspectos de su “*faculté langagière*”.
4. Los atlas lingüísticos representan instrumentos muy útiles para el estudio empírico de la gestión basilectal del espacio.
5. Como los datos de los atlas lingüísticos son (por definición) datos masivos, es preciso utilizar, para su análisis global, métodos cuantitativos.
6. La dialectometría representa la aplicación, a los datos de cualquier atlas lingüístico, de métodos cuantitativos (estadísticos) y cartográficos (visualizadores) debidamente escogidos. Permite el descubrimiento de estructuras geolingüísticas «escondidas» (o «profundas») e incluso de «leyes diatópicas». Estas leyes representan el equivalente epistemológico de las bien conocidas «leyes fonéticas» descubiertas por los Neogramáticos de Leipzig a finales del siglo xix.
7. La orientación teórica, empírica y metódica de la dialectometría es claramente interdisciplinaria. Disciplinas afines son: la clasificación numérica, la estadística, la cartografía y la geografía cuantitativas, la genética de poblaciones y todas las ciencias humanas ocupadas del estudio del espacio como, por ejemplo, la antropología, la etnografía, la sociología, la ecología, etc.

## 10. Agradecimientos

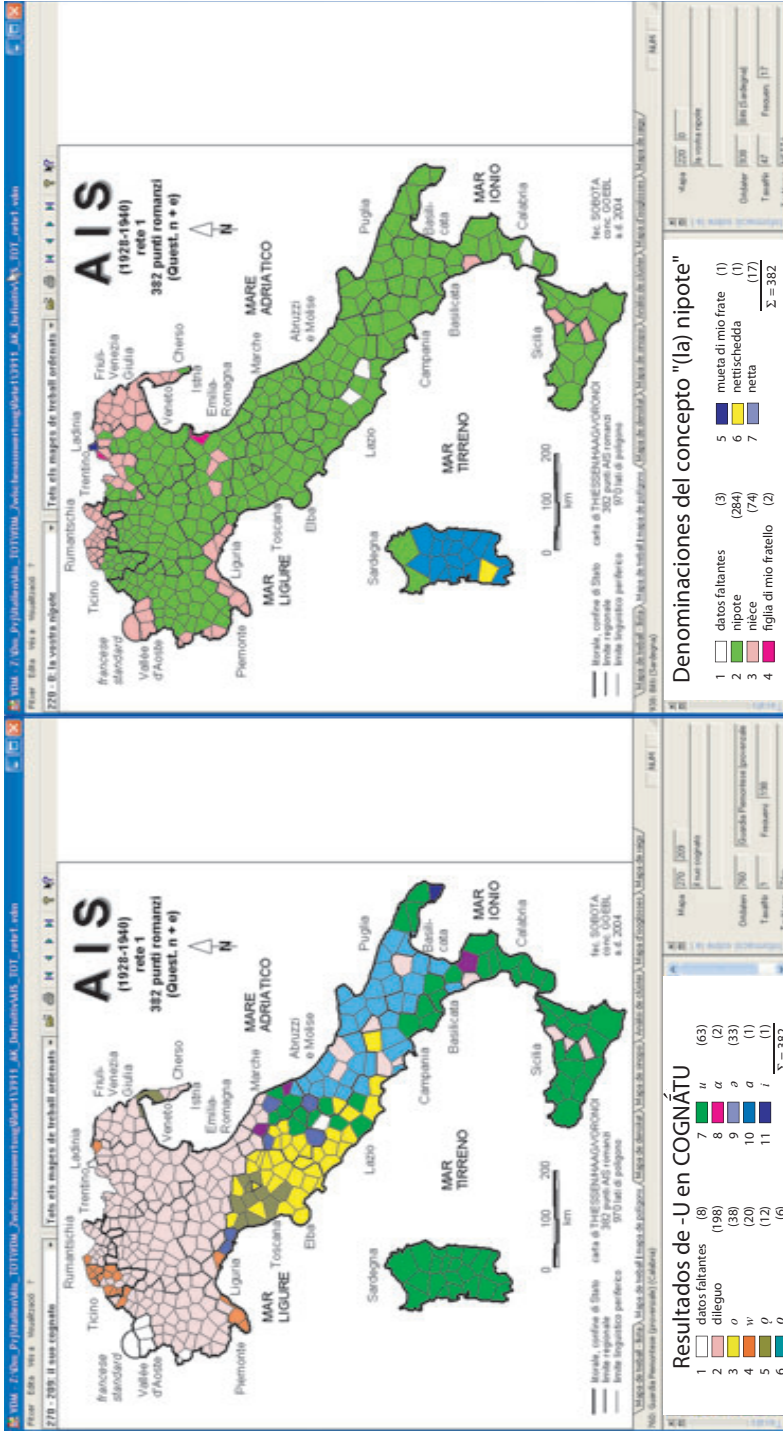
Quiero manifestar mi agradecimiento sincero a cuantas personas me han ayudado en la tarea de elaboración tanto de la ponencia oral de Vitoria-Gasteiz como de este artículo:

- Traducción castellana del original alemán: Xavier Casassas (Salzburgo).
- Supervisión de la corrección estilística de la versión final: Ramón de Andrés Díaz (Oviedo/Uviéu).
- Producción de los gráficos: Werner Goebel (Viena) y Slawomir Sobota (Salzburgo).
- Mantenimiento y perfeccionamiento continuo del programa VDM: Edgar Haimerl (Seattle) y Slawomir Sobota (Salzburgo).

## 11. Referencias bibliográficas

- AIS: Jaberg, K. & J. Jakob (eds.), 1928-1940, *Sprach- und Sachatlas Italiens und der Südschweiz*, Ringier, Zofingen, 8 vols. (reimpresión: Kraus, Nendeln, 1971).
- ALG: Séguy, J., 1954-1974, *Atlas linguistique et ethnographique de la Gascogne*, CNRS, París, 6 vols.
- Aurrekoetxea, G., 1992, «Nafarroako euskara: azterketa dialektometrikoa», *Uztaro* 5, 59-109.
- Bauer, R., 2009, *Dialektometrische Einsichten. Sprachklassifikatorische Oberflächenmuster und Tiefenstrukturen im lombardo-venedischen Dialektraum und in der Rätoromania*, Istitut ladin «Micurà de Rü», S. Martin de Tor (Ladinia monografica 01).
- Besch, W., 1967, *Sprachlandschaften und Sprachausgleich im 15. Jahrhundert. Studien zur Erforschung der spätmittelalterlichen Schreibdialekte und zur Entstehung der neuhochdeutschen Schriftsprache*, Francke, Munich.
- Bock, H. H., 1974, *Automatische Klassifikation. Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten (Cluster-Analyse)*, Vandenhoeck & Ruprecht, Göttingen.
- Brun-Trigaud, G., Le Berre, Y. & J. Le Dù, 2005, *Lectures de l'Atlas linguistique de la France de Gillieron et Edmont. Du temps dans l'espace. Essai d'interprétation des cartes de l'Atlas linguistique de la France de Jules Gillieron et Edmond Edmont augmenté de quelques cartes de l'Atlas linguistique de la Basse-Bretagne de Pierre Le Roux*, Éditions du CTHS, París.
- Chandon, J.-L. & S. Pinson, 1981, *Analyse typologique. Théories et applications*, Masson: París, Nueva York, Barcelona, Milán.
- Goebel, H., 1971, «Projekt einer sprachstatistischen Auswertung von in Sprachatlanten gespeicherter linguistischer Information mit Hilfe elektronischer Rechenanlagen», *Linguistische Berichte* 14, 60-61.
- , 1975, «Dialektometrie», *Grazer linguistische Studien* 1, 32-38.
- , 1976, «La dialectométrie appliquée à l'ALF (Normandie)», in A. Várvaro (ed.), *Atti del XIV Congresso Internazionale di Linguistica e Filologia Romanza*, Macchiaroli, Nápoles / Benjamins, Amsterdam, vol. 2, 165-195.
- , 1981, «Éléments d'analyse dialectométrique (avec application à l'ALS)», *Revue de linguistique romane* 45, 349-420.
- , 1983, «Parquet polygonal et treillis triangulaire: les deux versants de la dialectométrie interponctuelle», *Revue de linguistique romane* 47, 353-412.
- , 1984, *Dialektometrische Studien. Anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*, Niemeyer, Tübingen, 3 vols.
- , 1987, «Points chauds de l'analyse dialectométrique: pondération et visualisation», *Revue de linguistique romane* 51, 63-118.
- , 1992, «Problèmes et méthodes de la dialectométrie actuelle (avec application à l'ALS)», in Euskaltzaindia/Académie de la Langue Basque (ed.), *Nazioarteko Dialektologia Biltzarra. Agiriak/Actes du Congrès International de Dialectologie*, Bilbo/Bilbao, 429-475.
- , 2003, «Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALF): relations quantitatives et structures de profondeur», *Estudis Romànics* 25, 59-120.
- , 2005, «La dialectométrie corrélative: un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme», *Revue de linguistique romane* 69, 321-367.
- , 2008, «La dialettomettrizzazione integrale dell'ALS. Presentazione dei primi risultati», *Revue de linguistique romane* 72, 25-113.
- Haag, C., 1898, *Die Mundarten des oberen Neckar- und Donaufales (Schwäbisch-alemannisches Grenzgebiet: Baarmundarten)*, Hutzler, Reutlingen.
- Hall, R. A. Jr., 1943, «The Papal States in Italian Linguistic History», *Language* 19, 125-140.

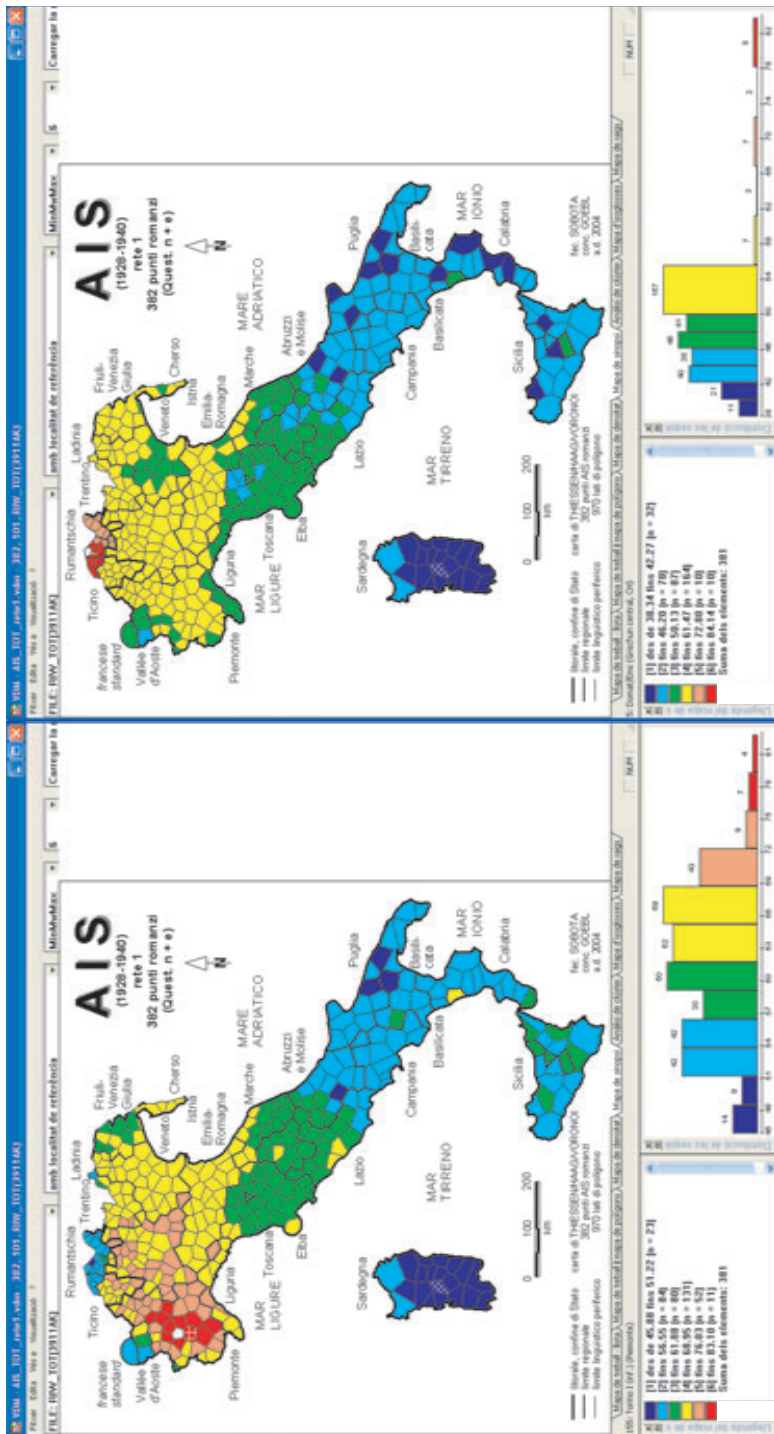
- Jaberg, K., 1906, «Zum Atlas linguistique de la France», *Zeitschrift für romanische Philologie* 30, 512.
- , 1908, *Sprachgeographie. Beitrag zum Verständnis des Atlas linguistique de la France*, Aarau: Sauerländer (versión española: *Geografía lingüística. Ensayo de interpretación del «Atlas lingüístico de Francia»*, traducción de A. Llorente y M. Alvar, Universidad de Granada; Secretariado de Publicaciones, Granada, 1959).
- & J. Jud, 1928, *Der Sprachatlas als Forschungsinstrument. Kritische Grundlegung und Einführung in den Sprach- und Sachatlas Italiens und der Südschweiz*, Niemeyer, Halle (reimpresión: Kraus, Nendeln, 1973).
- & —, 1987, *Atlante linguistico ed etnografico dell'Italia e della Svizzera meridionale, vol. I: L'atlante linguistico come strumento di ricerca. Fondamenti critici e introduzione*, edición italiana por G. Sanga y S. Baggio, Unicopli, Milán.
- Lalanne, T., 1953, «Indice de polyonymie. Indice de polyphonie», *Français moderne* 21, 263-274.
- Okabe, A., Boots, B. & K. Sugihara, 1992, *Spatial Tesselations. Concepts and Applications of Voronoi Diagrams*, Wiley, Chichester/Nueva York/Brisbane/Toronto/Singapore.
- Polanco Roig, L. B., 1984, «Llengua o dialecte: solucions teòriques i aplicació al cas català», in *Actes du XVII<sup>ème</sup> Congrès International de Linguistique et Philologie Romanes (Aix-en-Provence, 29 août-3 septembre 1983)*, Université de Provence, Aix-en-Provence / Jeanne Laffitte, Marsella, vol. 5: 13-30.
- Pop, S., 1950, *La dialectologie. Aperçu historique et méthodes d'enquêtes*, Duculot, Gembloux / chez l'auteur, Lovaina, 2 vols.
- Ravier, X., 1976, «Jean Séguy et la traversée du langage gascon. Réflexions sur une topogénèse géolinguistique», *Revue de linguistique romane* 40, 389-402.
- Rohlf, G., 1947, «Sprachgeographische Streifzüge durch Italien». [con 4 figuras y 29 mapas lingüísticos], Munich, in *Sitzungsberichte [Memorias] der Bayerischen Akademie der Wissenschaften, philosophisch-historische Klasse, Jahrgang 1944/46, Heft 3*, 1-67.
- Séguy, J., 1973, «La dialectométrie dans l'Atlas linguistique de la Gascogne», *Revue de linguistique romane* 37, 1-24.
- Sousa Fernández, X., 2006, «Aproximación á análise dialectométrica da variedades xeolingüísticas galegas: un estudo comparativo», in M. C. Rolão Bernardo & H. Mateus Montenegro (ed.), *Actas do I Encontro de Estudos Dialectológicos*, Instituto Cultural de Ponta Delgada, Ponta Delgada, 345-362.
- Veny, J., 2007-2009, *Petit atlas lingüístic del domini català*, Institut d'Estudis Catalans, Barcelona.
- Videsott, P., 2009, *Padania scrittologica: analisi scrittologiche e scrittometriche di testi in italiano settentrionale antico dalle origini al 1525*, Niemeyer, Tubinga.



**Mapa 1.** Mapa de trabajo fonético: distribución geográfica de los resultados italo-, sardo- y retorrománicos del nexo -u en COGNÁTU

**Mapa 2.** Mapa de trabajo léxico: distribución geográfica de las denominaciones italo-, sardo- y retorrománicas del concepto «sobrina» (según AIS 22 la vostra nipote)



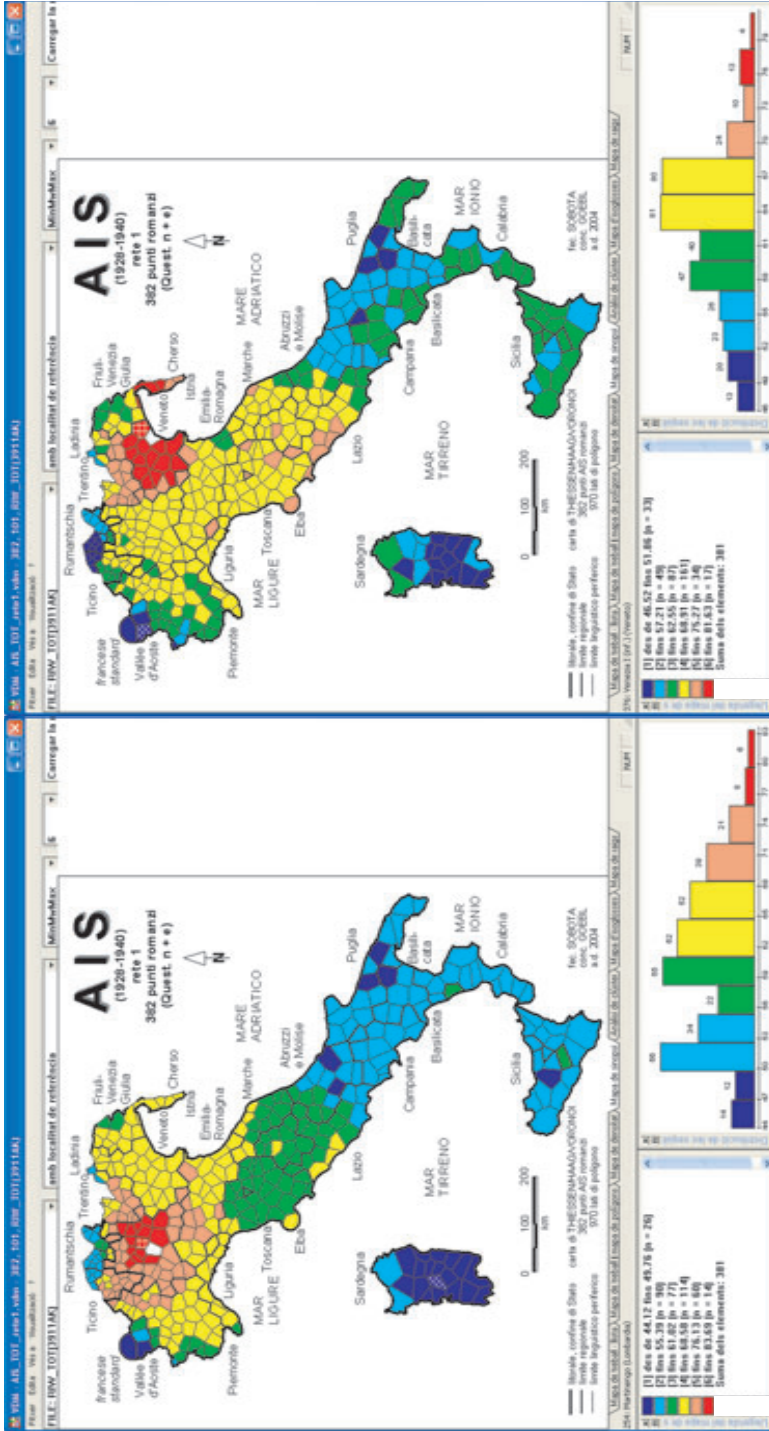


Mapa 3. Mapa de similitud relativo al P-AIS 155 Turin (Piamonte)

Algoritmo de visualización: MINMWMAX 6-tuplo  
 Medición de similitud:  $IRI_{jk}$   
 Corpus: corpus integral (3.911 mapas de trabajo)

Mapa 4. Mapa de similitud relativo al P-AIS 5 Domat/Ems (Grisones, CH)

Algoritmo de visualización: MINMWMAX 6-tuplo  
 Medición de similitud:  $IRI_{jk}$   
 Corpus: corpus integral (3.911 mapas de trabajo)

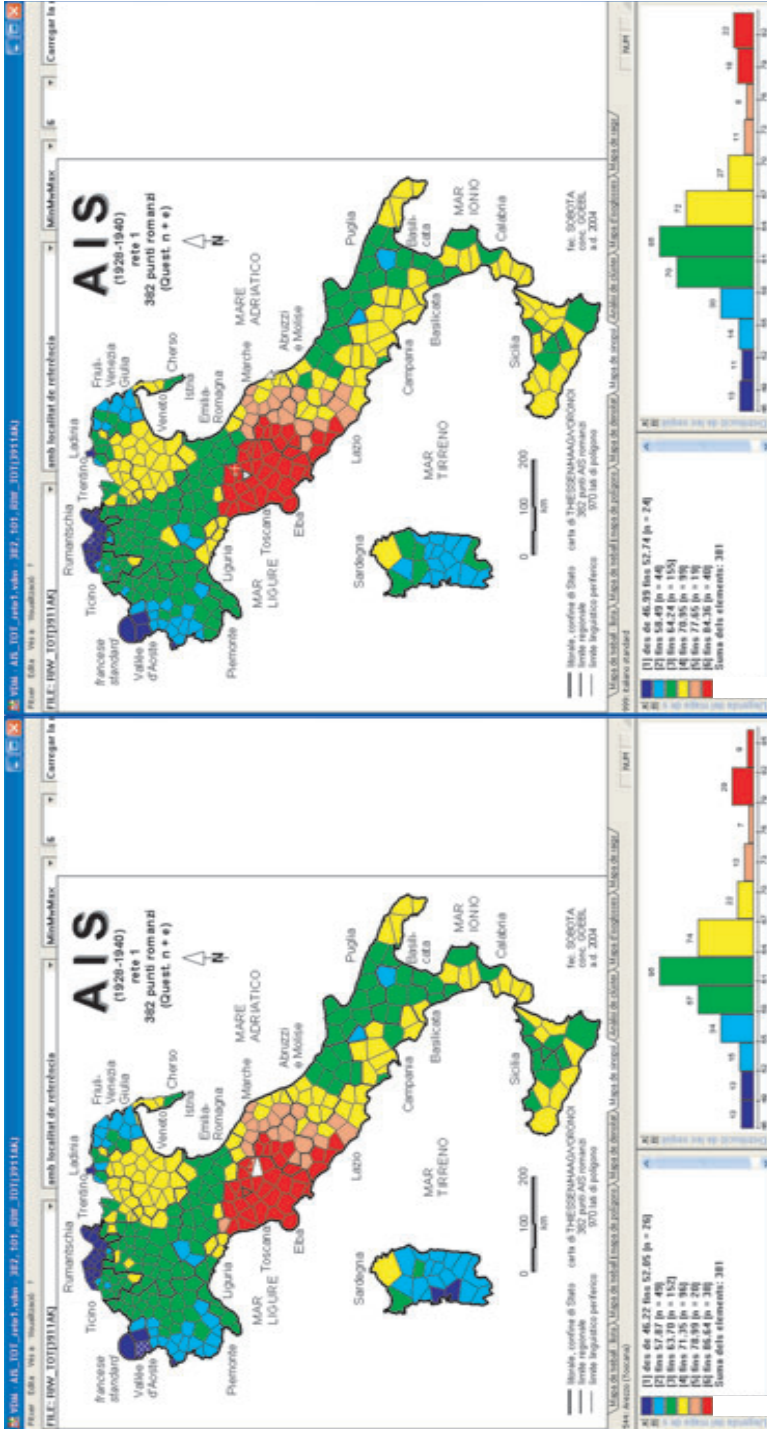


**Mapa 5** Mapa de similitud relativo al P-AIS 254 Martinengo (Lombardia)

Algoritmo de visualización: MINWMAX 6-tuplo  
 Medición de similitud:  $IRI_{jk}$   
 Corpus: corpus integral (3.911 mapas de trabajo)

**Mapa 6.** Mapa de similitud relativo al P-AIS 376 Venezia

Algoritmo de visualización: MINWMAX 6-tuplo  
 Medición de similitud:  $IRI_{jk}$   
 Corpus: corpus integral (3.911 mapas de trabajo)

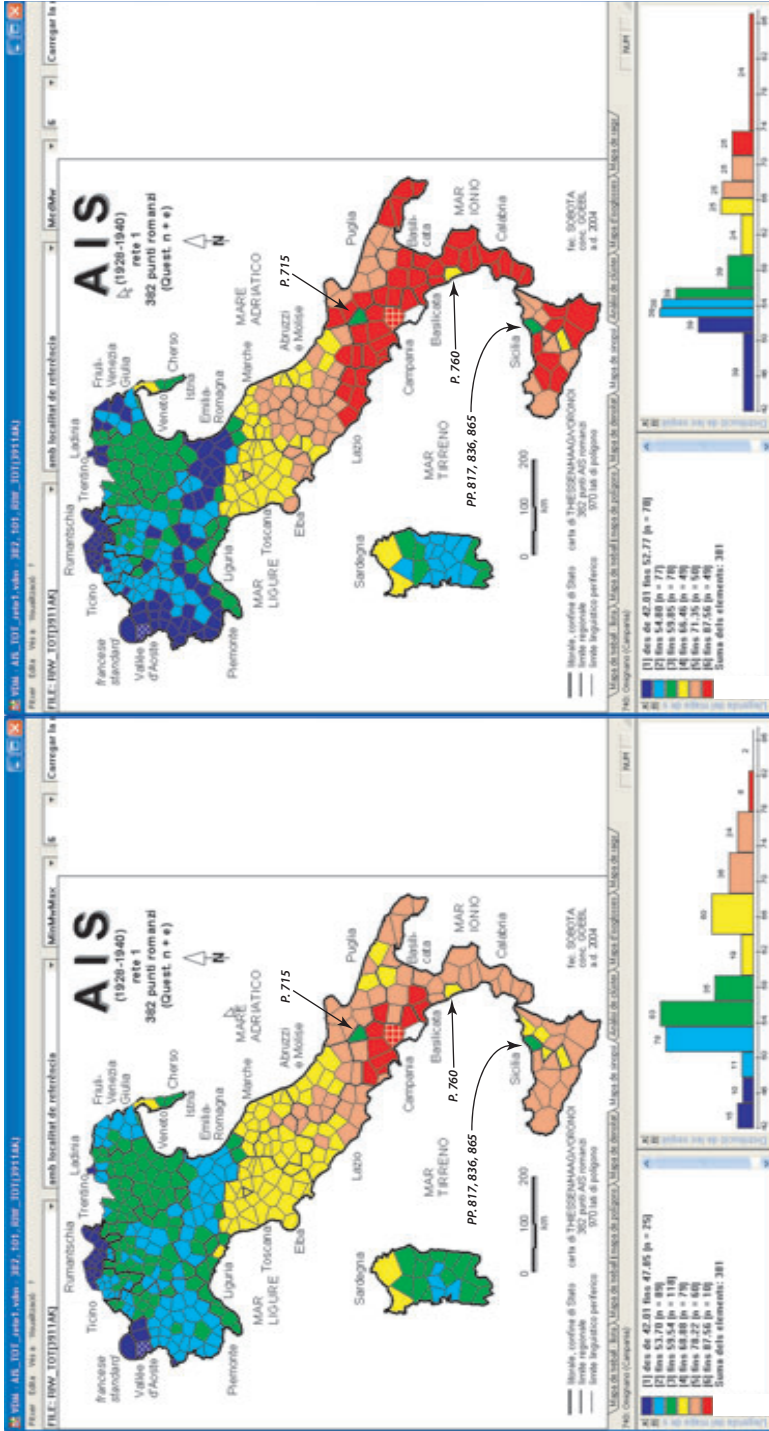


Mapa 7. Mapa de similitud relativo al P-AIS 544 Arezzo

Algoritmo de visualización: MINMWMAX 6-tuplo  
Medición de similitud:  $IRI_{j,k}$   
Corpus: corpus integral (3,911 mapas de trabajo)

Mapa 8. Mapa de similitud relativo al P-AIS 999 (italiano estándar)

Algoritmo de visualización: MINMWMAX 6-tuplo  
Medición de similitud:  $IRI_{j,k}$   
Corpus: corpus integral (3,911 mapas de trabajo)

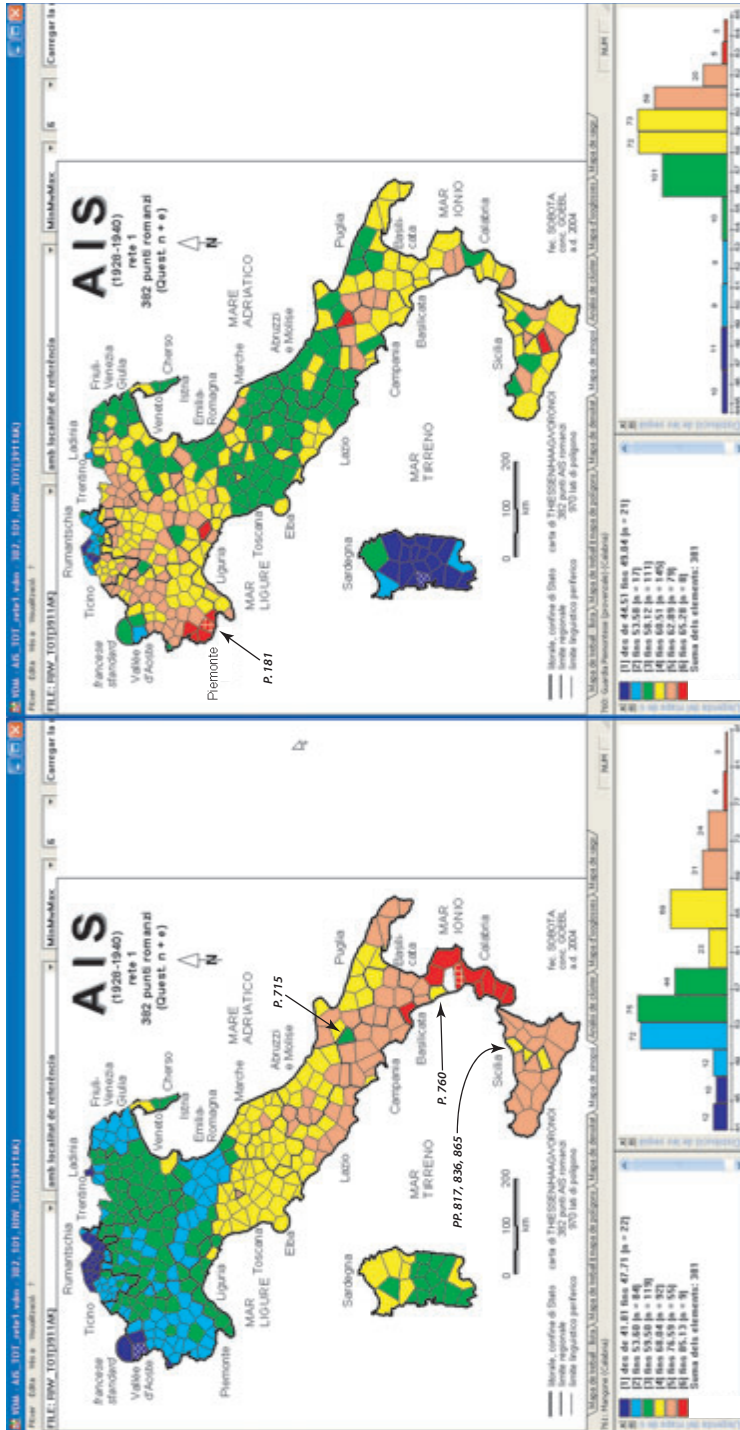


Mapa 9. Mapa de similitud relativo al P-AIS 740 Omignano (Campania)

Algoritmo de visualización: MINMWMAX 6-tuple  
 Medición de similitud:  $IRI_{ik}$   
 Corpus: corpus integral (3.911 mapas de trabajo)

Mapa 10. Mapa de similitud relativo al P-AIS 740 Omignano (Campania)

Algoritmo de visualización: MEDMW 6-tuple  
 Medición de similitud:  $IRI_{ik}$   
 Corpus: corpus integral (3.911 mapas de trabajo)

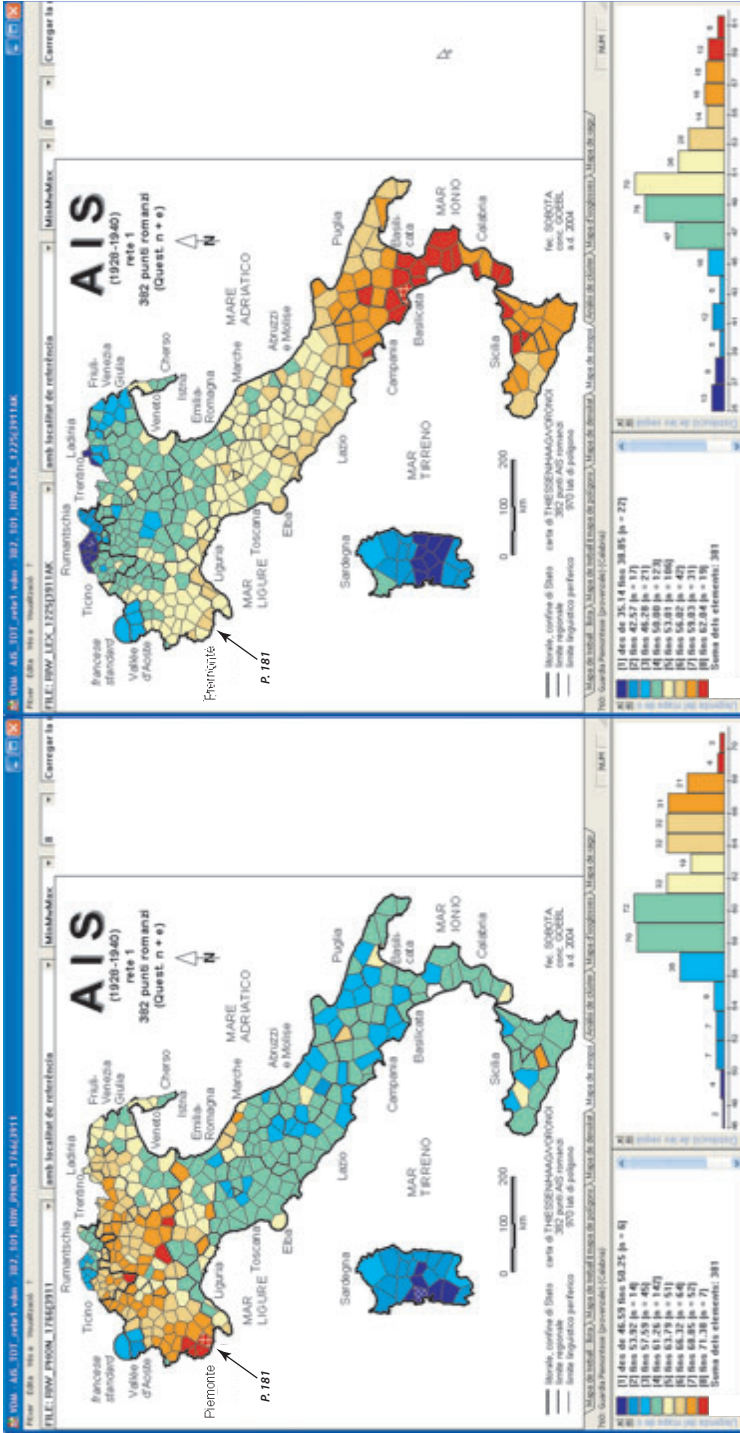


Mapa 11. Mapa de similitud relativo al P-AIS 761 Mangone (Calabria)

Algoritmo de visualización: MINMWMAX 6-tuplo  
 Medición de similitud:  $IRI_{j,k}$   
 Corpus: corpus integral (3,911 mapas de trabajo)

Mapa 12. Mapa de similitud relativo al P-AIS 760 Guardia Piemontese (Calabria)

Algoritmo de visualización: MINMWMAX 6-tuplo  
 Medición de similitud:  $IRI_{j,k}$   
 Corpus: corpus integral (3,911 mapas de trabajo)

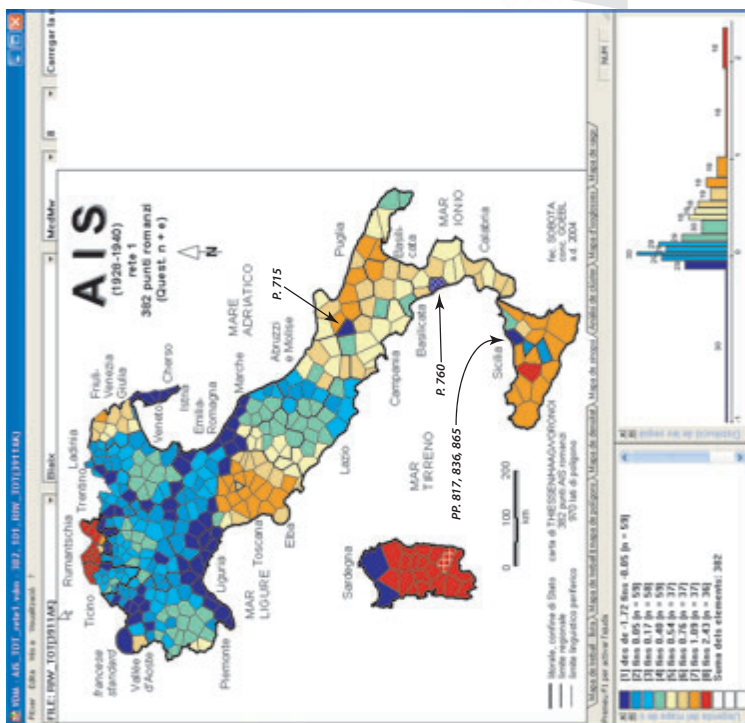


Mapa 14. Mapa de similitud relativo al P.-AIS 760 Guardia Piemontese (Calabria)

Algoritmo de visualización: MINMWMAX 6-tuplo  
Medición de similitud:  $IRI_{jk}$   
Corpus: subcorpus léxico (1.225 mapas de trabajo)

Mapa 13. Mapa de similitud relativo al P.-AIS 760 Guardia Piemontese (Calabria)

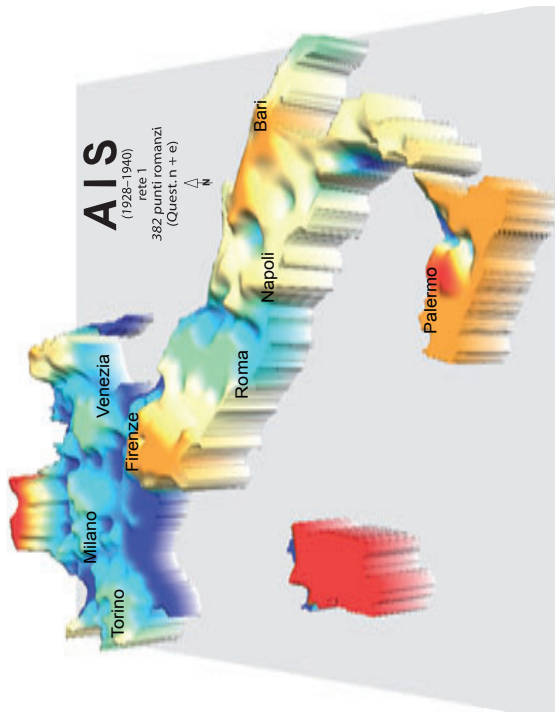
Algoritmo de visualización: MINMWMAX 6-tuplo  
Medición de similitud:  $IRI_{jk}$   
Corpus: subcorpus fonético (1.766 mapas de trabajo)



**Mapa 15.** Mapa de parámetros: sinopsis de 382 coeficientes de asimetría de Fisher (CAF)

Algoritmo de visualización: MEDMW 6-tuplo  
Medición de similitud:  $IRI_{jk}$   
Corpus: corpus integral (3,911 mapas de trabajo)

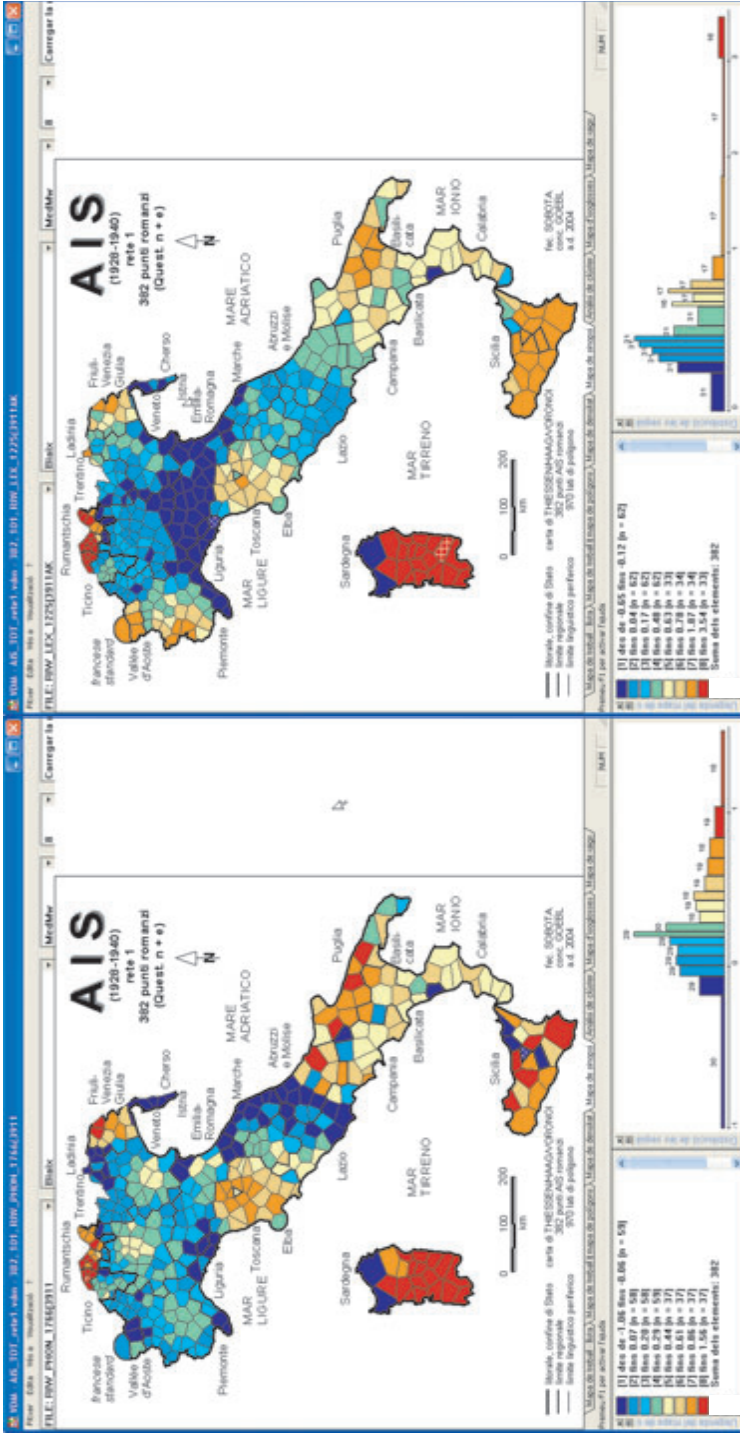
**Estereograma**



Punto de observación: desde sur  
Inclinación: 45°

**Mapa 16.** Estereograma de la sinopsis de 382 coeficientes de asimetría de Fisher (CAF)

Punto de observación: desde el sur  
Inclinación: 45°



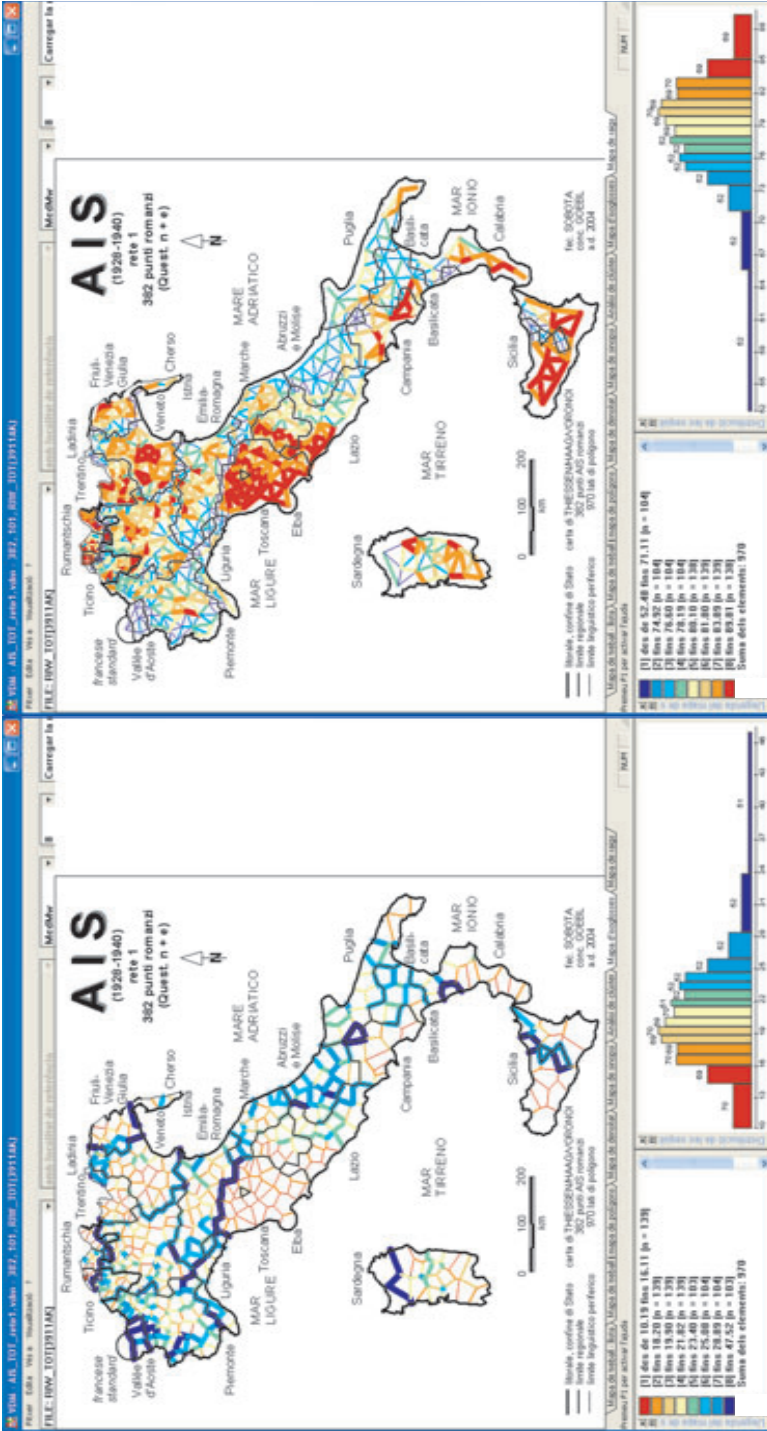
Mapa 18. Mapa de parámetros: sinopsis de 382 coeficientes de asimetría de Fisher (CAF)

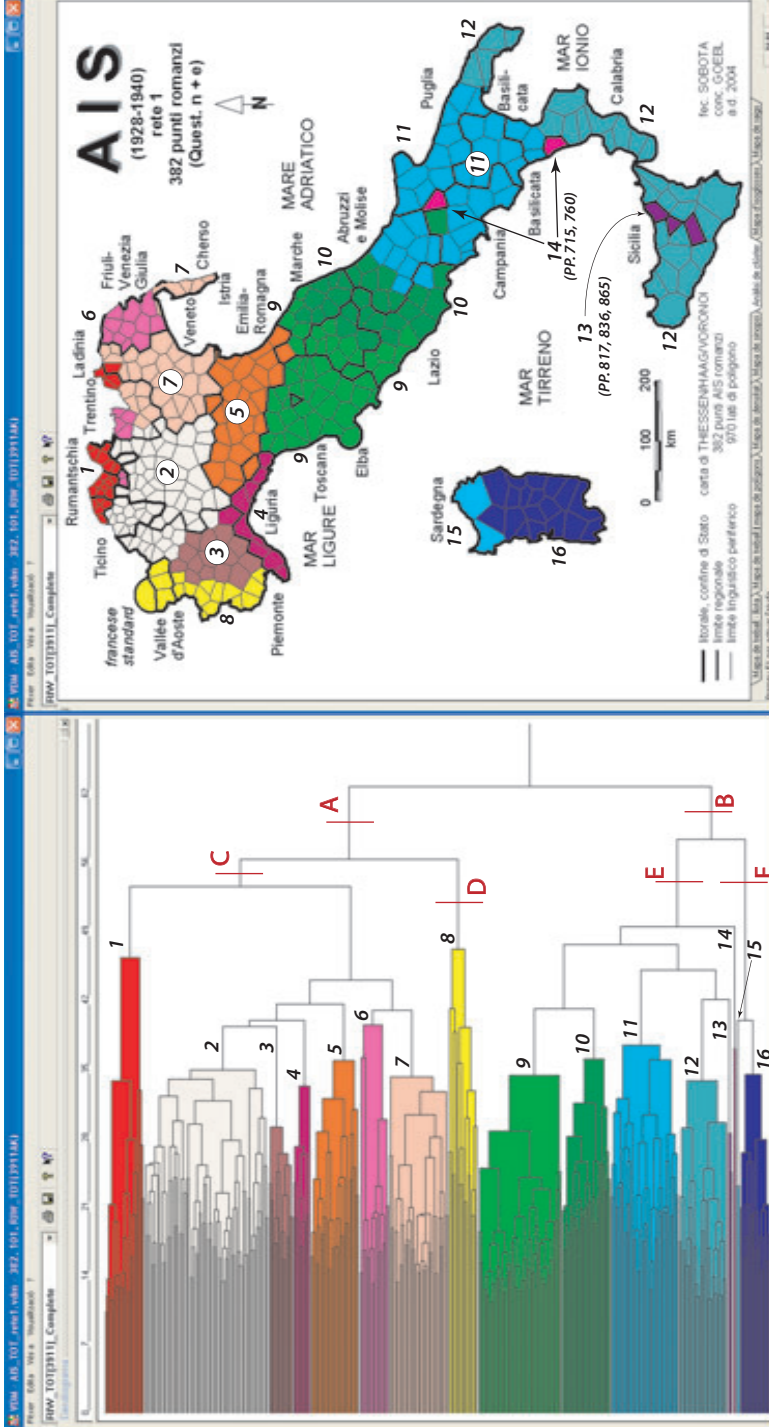
Algoritmo de visualización: MEDMW 6-tuplo  
 Medición de similitud:  $IRI_{jk}$   
 Corpus: subcorpus léxico (1.225 mapas de trabajo)

Mapa 17. Mapa de parámetros: sinopsis de 382 coeficientes de asimetría de Fisher (CAF)

Algoritmo de visualización: MEDMW 6-tuplo  
 Medición de similitud:  $IRI_{jk}$   
 Corpus: subcorpus fonético (1.766 mapas de trabajo)



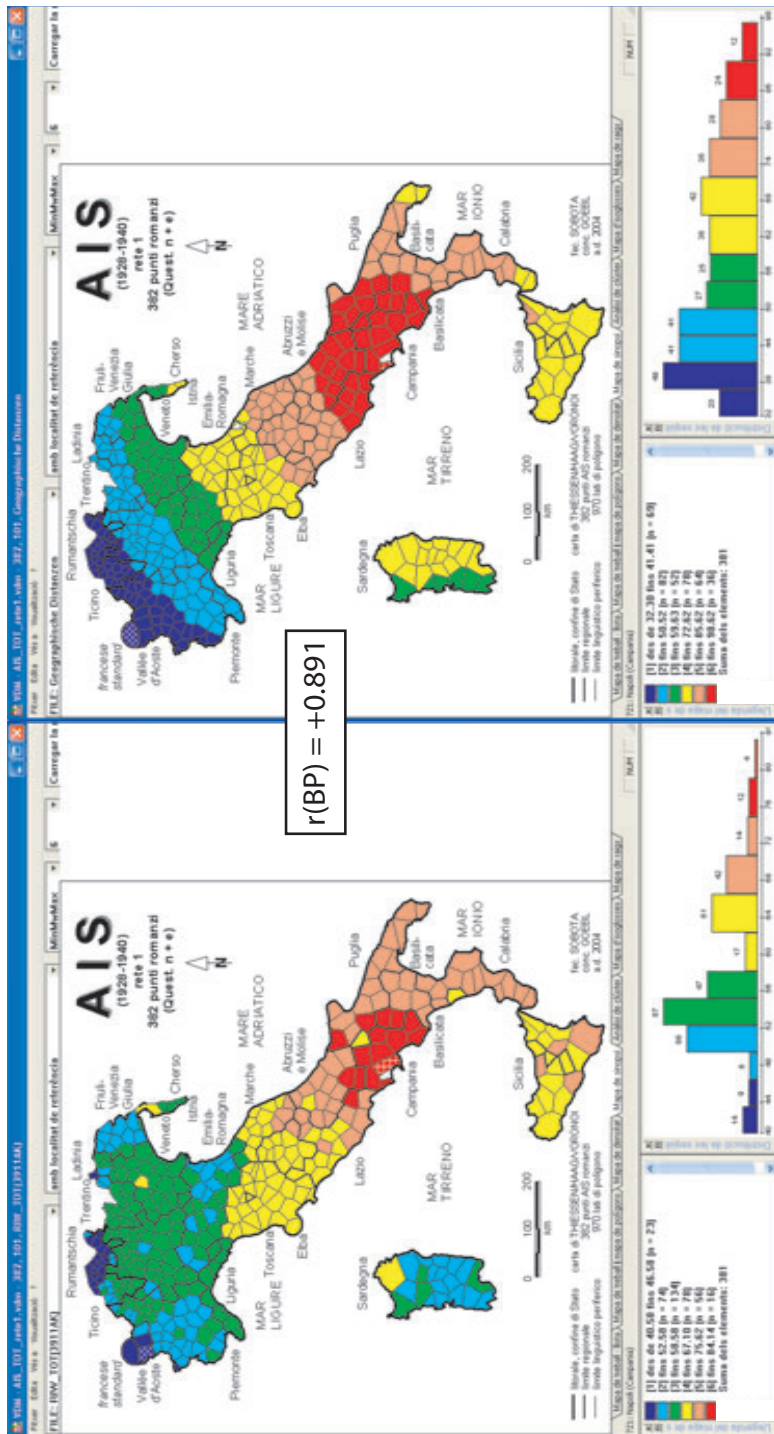




Mapa 21. Clasificación jerárquica aglomerativa: árbol genealógico

Algoritmo de clasificación: Complete Linkage  
 Medición de similitud:  $IR_{jk}$   
 Corpus: corpus integral (3,911 mapas de trabajo)

Mapa 22. Clasificación jerárquica aglomerativa: espacialización del mapa 21

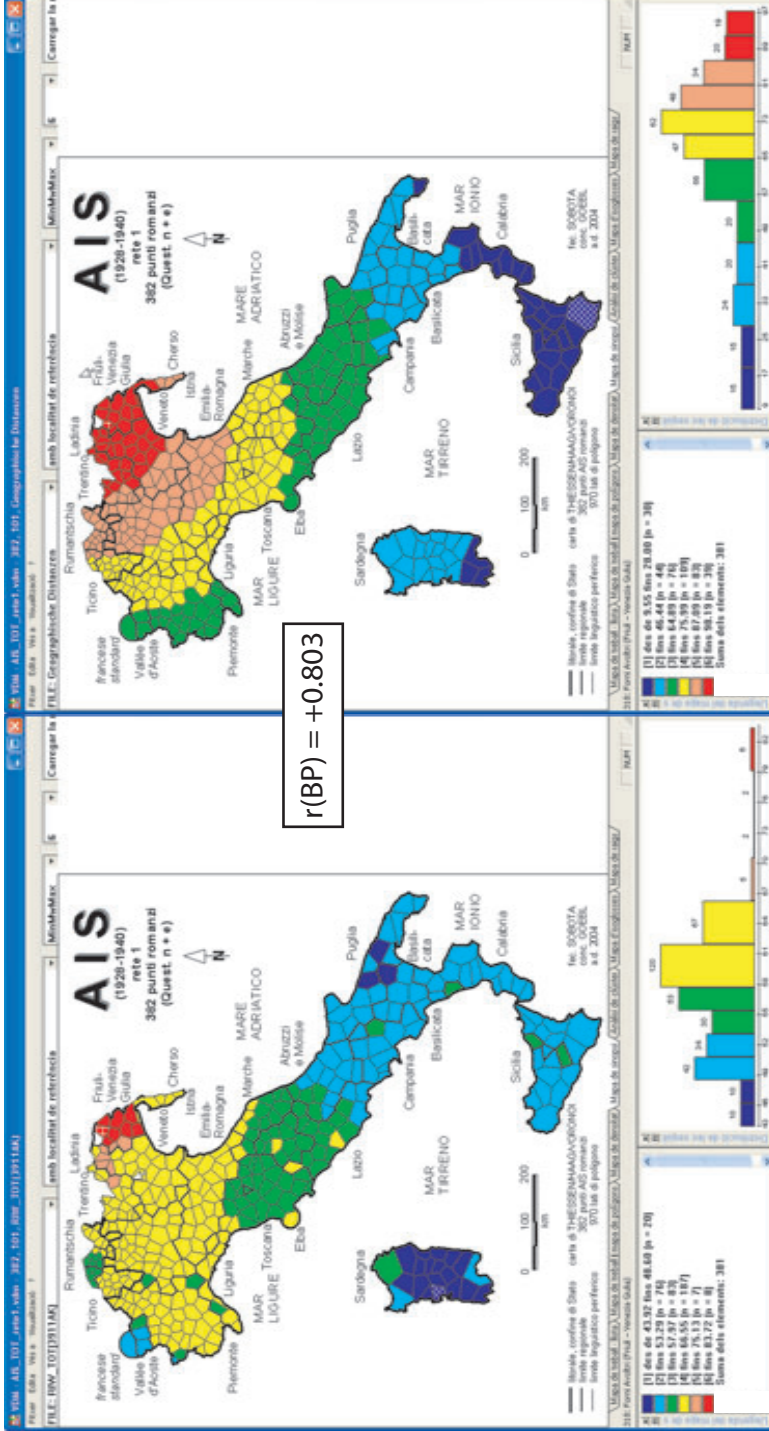


Mapa 23. Mapa de similitud relativo al P-AIS 721 Nápoles (Campania)

Algoritmo de visualización: MINMWMAX 6-tuplo  
 Medición de similitud:  $IR_{j,k}$   
 Corpus: corpus integral (3.911 mapas de trabajo)

Mapa 24. Mapa de proximidad relativo al P-AIS 721 Nápoles (Campania)

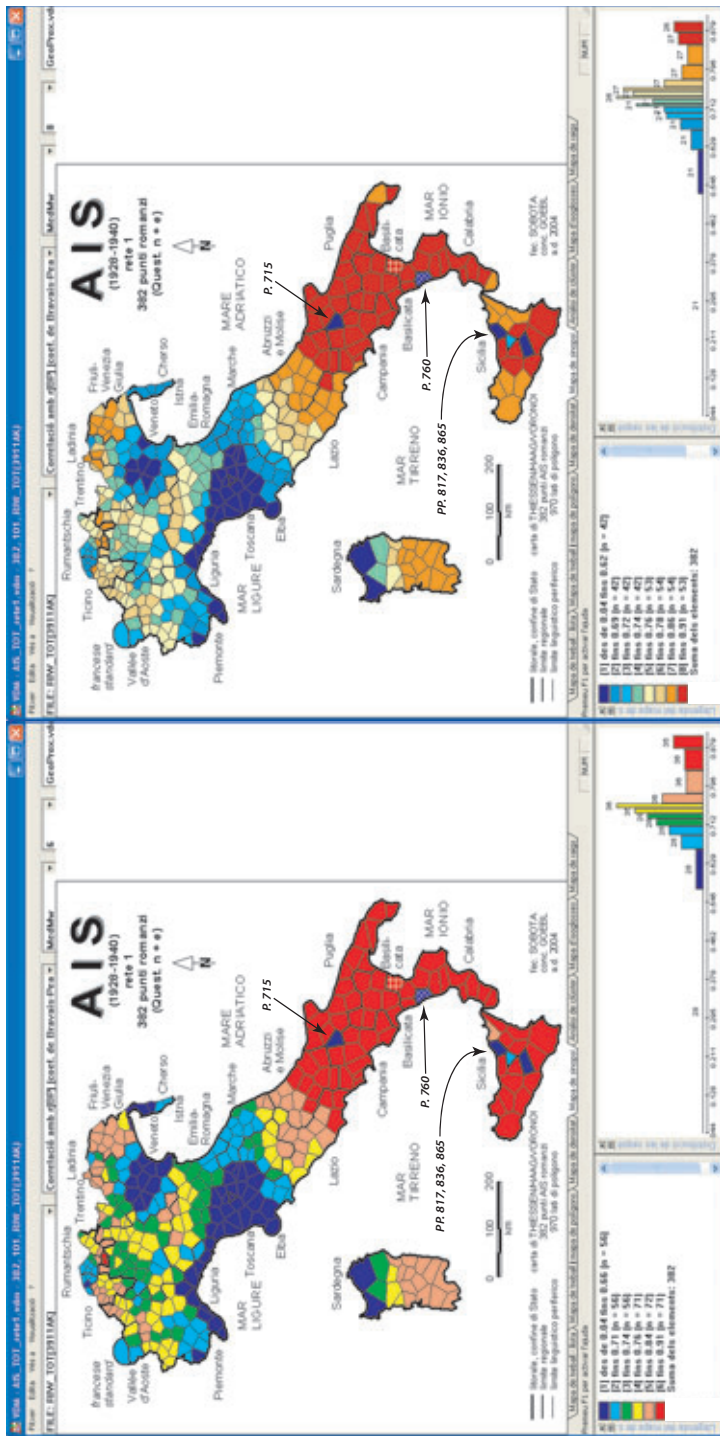
Algoritmo de visualización: MINMWMAX 6-tuplo  
 Medición de proximidad: teorema de Pitágoras



Mapa 25. Mapa de similitud relative al P-AIS 318 Forni Avoltri (Friuli)  
 Mapa 26. Mapa de proximitad relative al P-AIS 318 Forni Avoltri (Friuli)

Algoritmo de visualizaci3n: MINMWMAX 6-tuplo  
 Medici3n de proximitad: teorema de Pit3goras

Algoritmo de visualizaci3n: MINMWMAX 6-tuplo  
 Medici3n de similitud:  $IRI_{jk}$   
 Corpus: corpus integral (3.911 mapas de trabajo)

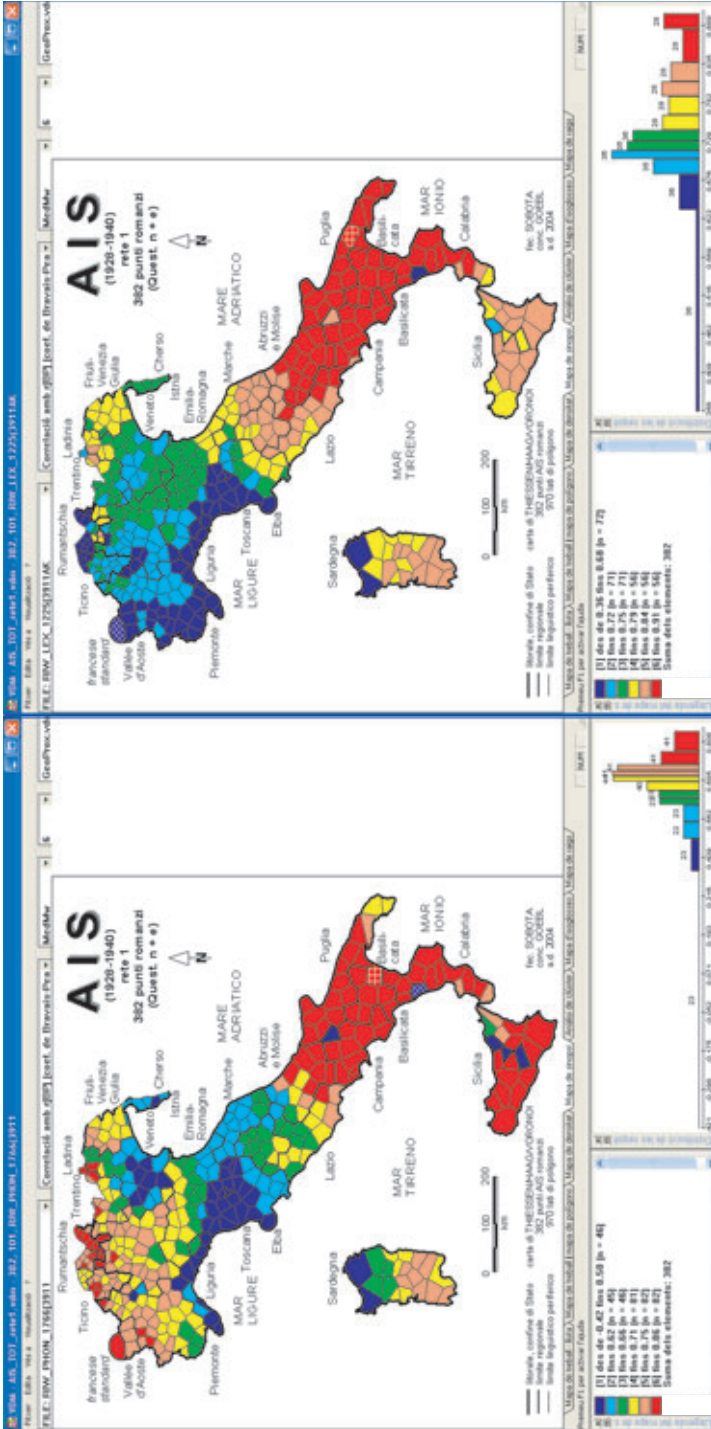


Mapa 27. Mapa de correlaciones: similitudes lingüísticas y proximidades geográficas

Algoritmo de visualización: MEDMW 6-tuplo  
 Medición de similitud:  $IRI_{jk}$   
 Corpus: corpus integral (3.911 mapas de trabajo)  
 Medición de proximidad: teorema de Pitágoras

Mapa 28. Mapa de correlaciones: similitudes lingüísticas y proximidades geográficas

Algoritmo de visualización: MEDMW 8-tuplo  
 Medición de similitud:  $IRI_{jk}$   
 Corpus: corpus integral (3.911 mapas de trabajo)  
 Medición de proximidad: teorema de Pitágoras

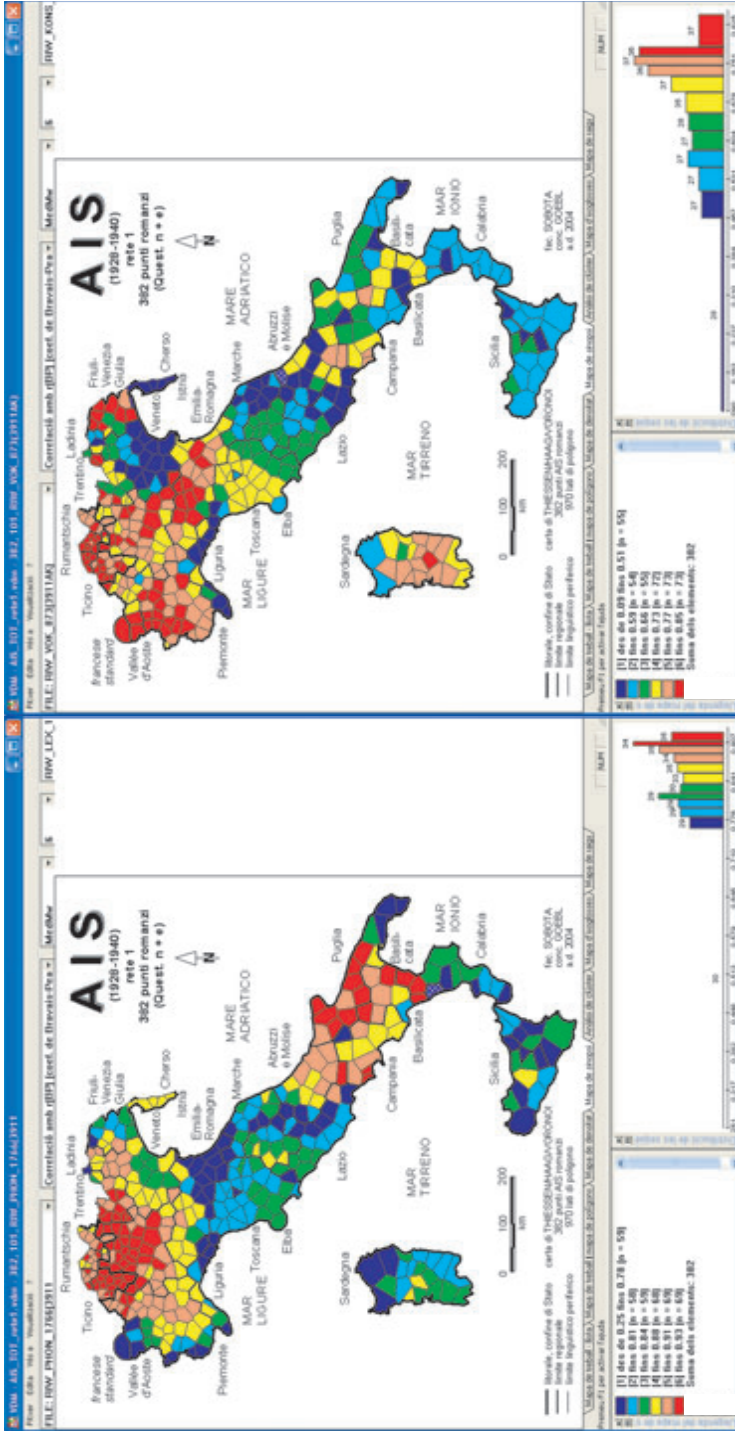


Mapa 29. Mapa de correlaciones: similitudes fonéticas y proximidades geográficas

Algoritmo de visualización: MEDMW 6-tuplo  
 Medición de similitud:  $IRI_{jk}$   
 Corpus: subcorpus fonético (1.766 mapas de trabajo)  
 Medición de proximidad: teorema de Pitágoras

Mapa 30. Mapa de correlaciones: similitudes léxicas y proximidades geográficas

Algoritmo de visualización: MEDMW 6-tuplo  
 Medición de similitud:  $IRI_{jk}$   
 Corpus: subcorpus léxico (1.225 mapas de trabajo)  
 Medición de proximidad: teorema de Pitágoras



Mapa 31. Mapa de correlaciones: similitudes fonéticas y léxicas

Algoritmo de visualización: MEDMW 6-tuplo  
 Medición de similitud:  $IR_{j,k}$   
 Subcuerpo lingüísticos: fonética: 1.766 mapas de trabajo léxico: 1.225 mapas de trabajo

Mapa 32. Mapa de correlaciones: similitudes vocálicas y consonánticas

Algoritmo de visualización: MEDMW 6-tuplo  
 Medición de similitud:  $IR_{j,k}$   
 Subcuerpo lingüísticos: vocalismo: 873 mapas de trabajo consonantismo: 805 mapas de trabajo