

## UNA CLASSIFICAZIONE GERARCHICA DI DATI GEOLINGUISTICI TRATTI DALL'AIS. SAGGIO DI DIALETTOMETRIA DENDROGRAFICA

Collaborazione informatica: Erasmus Langer (Università Politecnica  
di Vienna)<sup>1</sup>

### 1. ALCUNI PRINCIPI DELLA CLASSIFICAZIONE GERARCHICA

Lo strumento euristico classico della classificazione gerarchica è l'albero (chiamato anche in modo più specialistico: grafo arborescente). La struttura ramificata dell'albero con una molteplicità di foglie, ramoscelli e rami rispetto alla radice unica — poli tra i quali si inserisce uno spazio di dipendenze gerarchiche ben articolate — costituisce una rappresentazione metaforicamente molto valida per la descrizione dei risultati di qualsiasi processo *evolutivo*. Ora si sa che ogni *evoluzione* procede mediante ramificazioni, biforcazioni o frammentazioni consecutive e che nascono così vari raggruppamenti in classi (gruppi, unità ecc.), tra i quali esistono sempre determinate dipendenze o relazioni. La lista delle scienze che si sono avvalse dell'albero in quanto mezzo euristico è lunga: oltre la *zoologia*, la *paleontologia*, la *genetica* e tante altre discipline *biologiche* figurano anche discipline *antropiche* e *sociali*, tra cui anche la *linguistica*. E noto che l'albero genealogico viene adoperato in sede linguistica sin dai tempi di August SCHLEICHER (1863). L'uso dell'albero per scopi linguistici è sempre rimasto tuttavia alquanto problematico. In effetti, lo strumento euristico "albero" offre particolari condizioni di applicazione, che fino ad oggi non sono state sufficientemente discusse. E' così rimasta, tra la stragrande maggioranza dei linguisti, una più o meno diffusa sfiducia rispetto all'albero. Cf. a questo proposito la buona documentazione storica di STEWART 1976.

Oggi, con la comparsa di nuove possibilità euristiche (legate cioè all'uso generalizzato di computers per la costruzione di alberi genealogici in gran numero e di qualsiasi strutturazione), il dibattito al merito si è riaperto.

#### 1.1. La generazione di alberi mediante algoritmi tassometrici

La tassometria moderna (chiamata anche "Numerical Classification, Classification automatique, Analyse des données, Automatische Klassifikation" ecc.) di-

<sup>1</sup> Ringrazio cordialmente la dottoressa Adelaide Fiocchi-Baehr (Università di Salisburgo) della revisione stilistica del mio testo italiano.

spone ormai, nel suo arsenale metodico, di una folta serie di algoritmi classificatori destinati a generare, sempre in base a certe regole matematiche, strutture arbore-scenti ben definite. Nella manualistica tassometrica internazionale questi algoritmi vengono chiamati “Agglomerative Hierarchic Clustering Methods” (p. es. SNEATH/SOKAL 1973, 214 s.), (méthodes de) “classification hiérarchique” (p. es. CHANDON/PINSON 1981, 98 s.) o “Hierarchisch agglomerative Verfahren” (p. es. BOCK 1974, 356 s.). Siccome la stragrande maggioranza degli algoritmi dendrografici gerarchici descritti nei manuali tassometrici può esser eseguita agevolmente mediante i cosiddetti “software packages” statistici (di diffusione internazionale e utilizzabili tanto nei grossi centri universitari di calcolo quanto con singoli “personal computers”), nulla si oppone alla sperimentazione massiccia di alberi genealogici di qualsiasi strutturazione in sede di *geolinguistica*.

### 1.2. Tanto per ricordare: l'inevitabile molteplicità dei procedimenti (e risultati) classificatori

Per qualsiasi problema di classificazione *non* si dà l'*unicità* di soluzioni classificatorie. Ogni insieme di dati destinato ad essere classificato può esserlo *in vari modi*. Il classificatore deve dunque effettuare una serie di *scelte* o di *decisioni metodiche*, guidato in ciò ovviamente dalle sue posizioni *teoriche*. Deve scegliere tra l'altro il procedimento classificatorio secondo la finalità intrinseca del suo progetto (scartando eo ipso altri procedimenti apparentemente meno adatti) ed anche regolare l'impostazione teorica dell'interpretazione finale dei risultati su presupposti teorici da precisare in anticipo. Ogni classificazione rimane sempre, in quanto *metodo euristico*, l'*umile ancella* delle intenzioni *teoriche* del classificatore. Imitando la formula allitterativa ben nota di LASSWELL (“Who speaks what language...?”) pare lecito crearne un'altra (con risonanze allitterative latine in *qu-*): “*Quis facit qualem ordinem quot qualiumque classium quibus methodis ac per quales fines?*”.

Per ovvie ragioni di spazio questo articolo non contiene che il risultato iconico di una sola classificazione dendrografica. Essa è stata scelta, a scopo illustrativo, tra tante altre generate durante la mia ultima “campagna” di dialettometria dendrografica. In questi lavori mi sono avvalso della collaborazione efficace e preziosa del valente dottor ERASMUS LANGER, docente al Politecnico di Vienna. Al dottor LANGER sono stati affidati i calcoli numerici (con SPSS — x) e — problema molto più spinoso ancora — il disegno automatizzato (mediante stampante a raggi Laser) degli alberi (ciascuno con almeno 251 foglie e 250 ramificazioni). Che il dottor LANGER riceva in questa sede i miei più sentiti ringraziamenti per la sua preziosa collaborazione!

### 1.3. La logica della strutturazione degli alberi: binarietà e agglomerazione “dal basso in su”

1.3.1. L'albero famoso dello SCHLEICHER (1863) disponeva di ramificazioni tanto *binarie* quanto *ternarie*. In sede di tassometria le ramificazioni sono sempre binarie con esclusione di qualsiasi altra possibilità. Questa limitazione costituisce, ovviamente, un'*idealizzazione* ossia *riduzione* non poco problematica rispetto alla realtà spesso esuberante dell'evoluzione biologica o glottologica.

1.3.2. La generazione agglomerativa dell'albero comincia da parte delle foglie, cioè “dal basso in su” (si veda la Fig. 1). Ad ogni foglia corrisponde uno degli N elementi classificabili. Da questi N elementi (o foglie) l'algoritmo di classificazione comincia coll'agglomerarne certuni, a seconda delle sue virtù matematiche, riducendo così progressivamente (cioè con ogni agglomerazione o fusione binaria) il numero degli elementi rimanenti ed ancora da agglomerare (o fondere tra di loro). Un albero con N foglie dispone, ad agglomerazione compiuta, di N-1 ramificazioni (o biforcazioni). Ciascuna delle ramificazioni “ha il comando” di due classi (gruppi, unità, elementi ecc.) sottostanti, le quali, a seconda della loro posizione nella gerarchia dendrografica, dispongono di un diverso grado di eterogeneità interna (micro-, meso-, macro-gruppi). Più le classi sono “in alto”, cioè vicine alla radice (si veda la Fig. 1), più la loro eterogeneità interna (“intra-class-heterogeneity”) è alta. Per fare un esempio (si veda di nuovo la Fig. 1): le classi dipendenti dalle ramificazioni 5 e 6 dispongono di una eterogeneità interna minore di quelle sottostanti alle biforcazioni 1 e 2. Si osservi anche — sulla Fig. 1 — la freccia puntata in alto che indica la direzione delle N-1 agglomerazioni.

### 1.4. Una volta di più: l'orientamento *complessivo* del pensiero dialettometrico

Di fronte ad alcuni equivoci e fraintendimenti relativi alla finalità primaria del pensiero dialettometrico, espressi soprattutto da alcuni dei miei amici italiani, ritengo utile ribadire qui due tra i principi elementari di qualsiasi impostazione dialettometrica:

1. La dialettometria si applica soltanto ai dati di *atlanti linguistici*: “extra atlantes linguisticos nulla salus dialectometrica”.
2. Alla dialettometria spetta, in quanto procedimento *induttivo* (che parte dal *particolare* per arrivare al *generale*), la considerazione *complessiva* dei dati di un atlante linguistico. Non va dimenticato però che una considerazione complessiva senza indicazioni precise delle coordinate del pensiero di colui che pratica questa considerazione, non è possibile.

Considerazione *isolata* di singole carte di un atlante linguistico o considerazione *complessiva* di molte carte: si tratta di due facce di una sola medaglia. Non c'è

la minima traccia di una qualsiasi concorrenza e contraddizione. Già nello JABERG maturava, 67 anni or sono, l'idea di una considerazione complessiva dei dati della raccolta che stava allora preparando (AIS): "Fin d'ora la grande varietà dei limiti che presentano le nostre carte fa prevedere che non ci sarà un'unica soluzione giusta sempre e dappertutto. Bisognerà cominciare collo studio particolareggiato di ogni carta ed elevarsi a grado a grado a sintesi più alte." (JABERG 1923, 1191—1192).

## 2. BREVE DESCRIZIONE DELLA CATENA DIALETTOMETRICA UTILIZZATA

### 2.1. Misurazione dei dati originali tratti dall'AIS

Per questo problema rimando ai miei lavori anteriori (1981, 352 s. 1982, 15 s.; 1984a I, 31 s.; 1984b, 11 s.): la matrice dei dati ivi descritta è quella qui utilizzata:

*volumi dell'AIS esauriti*: I, II, IV.

*variabilità geolinguistica considerata* rispettivamente *misurata su scala nominale* (politomica): variazione lessicale (ca. due terzi delle carte-AIS), variazione morfo-sintattica (ca. un terzo delle carte-AIS)

*dimensioni della matrice dei dati*:

N = 251 punti (o elementi)-AIS (con inclusione di un punto artificiale: P. 999 = italiano standard).

p = 696 carte "di lavoro" (o attributi).

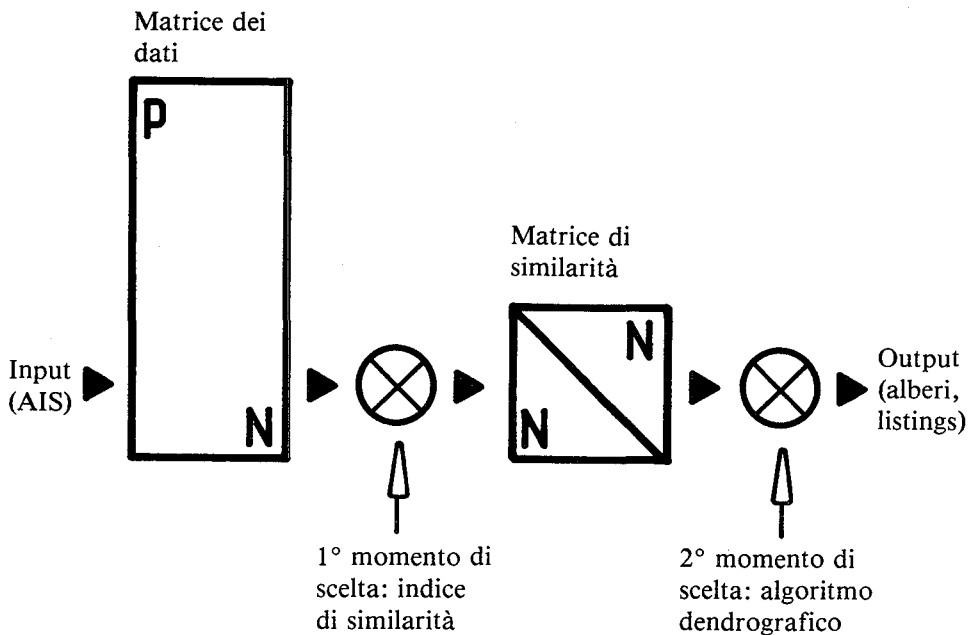
### 2.2. Dalla matrice dei *dati* alla matrice di *similarità*

Tra queste due matrici si colloca la scelta (e l'applicazione consecutiva) dell'*indice di similarità*. Questa scelta costituisce un punto cruciale della catena classificatoria. Ne ho ampiamente descritto i problemi nei miei scritti dialettometrici anteriori (cf. p. es. GOEBL 1984a I, 74 s.). L'indice qui utilizzato viene chiamato "Indice ponderato di identità" (IPI; in tedesco: "Gewichtender Identitätswert, GIW"). La sua particolarità matematica consiste, in breve, nell'attribuire, a tassati di piccola estensione geografica, un maggiore peso numerico rispetto a quelli più estesi e considerati quindi come "banali", meno significativi ecc. Si capisce subito che la ponderazione numerica dei tassati (o tipi lessicali) meno diffusi equivale ad una certa *impostazione teorica* in sede di geolinguistica, che d'altronde vanta non pochi fautori tanto fra *romanisti* che *germanisti*, *indoeuropeisti* (glottologi) ed altri. Nell'abbreviazione IPI(1) (o GIW(1)) la cifra 1 indica il valore numerico del *peso* ammesso da me nella formula complessiva dell'IPI (o GIW). Non potendo approfondire l'argomento in questa sede, rimando alle mie pubblicazioni del 1983 (in tedesco: 12 ss.) e del 1987 (in francese: 70 ss.) dove il lettore interessato troverà una ampia discussione degli aspetti matematici e linguistici del problema.

### 2.3. Scelta dell'algoritmo di classificazione gerarchica

Si tratta di nuovo, con questa scelta, di *modellizzare* la concezione linguistica dell'evoluzione dendrografica (meglio: dendrograficamente rappresentabile) dei dialetti in una formula matematica appositamente selezionata. A questo proposito riassumiamo brevemente i sommi capi dei rispettivi concetti dei linguisti. *L'opinio communis* ammette — non senza sfumature nei dettagli — che in una massa di dati geolinguistici inizialmente omogenei si siano operate, coll'andar del tempo, un certo numero di agglomerazioni o raggruppamenti (spesso chiamati, in modo semanticamente erroneo: frammentazioni, fragmentations, Ausgliederungen ecc.) che portano allo stabilirsi di una gerarchia dialettale, della quale si vedono ancora oggi le tracce. A questa concezione che, oltre ad esistere in tante varianti, non è priva di assunzioni problematiche, fanno eco non pochi algoritmi tassometrici, tra i quali, secondo le mie esperienze personali, gli algoritmi "complete linkage" e quello proposto da WARD sono molto validi per la ricerca geolinguistica. Qui viene adoperato, a mo' d'esempio, l'algoritmo "complete linkage". Per una descrizione dettagliata cf. GOEBL 1984a I, 172 e 1983, 17 s. o — molto meglio ancora — i capi rispettivi della letteratura manualistica: p. es. SNEATH/SOKAL 1973, 216 s. (inglese); BOCK 1974, 382 s. (tedesco) e CHANDON/PINSON 1982, 106 s. (francese).

Riassumiamo schematicamente l'articolazione della catena dialettometrica utilizzata:



L'output di un'analisi dendrografica comprende, a prescindere dal disegno dell'*albero genealogico*, anche *lunghe liste* numeriche coi valori delle N-1 fusioni adoperate, in modo tale che il classificatore possa controllare la genesi dell'albero nei più minuti dettagli. La riproduzione tanto dell'albero (con tutte le 251 foglie e le 250 ramificazioni) quanto dei "listings" attinentivi è praticamente impossibile per motivi di spazio. Qui ci contentiamo della presentazione di un disegno *mediamente semplificato* dell'albero calcolato (Fig. 1) e della sua *trasposizione* (cioè *spazializzazione*) *nella rete dell' AIS* (Fig. 2).

### 3. INTERPRETAZIONE LINGUISTICA DELLE FIGG. 1 E 2

Tra le molte possibilità d'interpretazione linguistica di classificazioni dendrografiche spiccano le ipotesi *geolinguistica* (e *sincronica*) e quella *glottocronologica* (e *diacronica*).

#### 3.1. Interpretazione *geolinguistica* delle Fig. 1 e 2

La Fig. 1 contiene l'albero al rovescio: la chioma verso il basso e la radice in alto. Mentre l'agglomerazione dell'albero comincia al livello delle N foglie, l'interpretazione dell'albero deve farsi in senso inverso: dall'alto in basso. I rami più grossi, vicini alla radice, sono i più interessanti. Sulla Fig. 1 sono contrassegnati colle cifre 1—6. Ogni albero gerarchico contiene diversi livelli di astrazione o — in termini fotografici — di risoluzione granulosa. Una risoluzione (molto grossolana) dell'intera rete-AIS in soli sei gruppi sarebbe utile, ma gioverebbe poco agli scopi illustrativi di questo articolo. Abbiamo dunque scelto un livello medio di risoluzione in 25 classi (gruppi, unità ecc.), contrassegnata ciascuna dalle lettere A-Y. L'interpretazione *geolinguistica* dell'albero diventa così molto più agevole. Il nostro proposito è di far vedere l'ammirevole precisione e compattezza, con cui vengono delineate, sulla Fig. 2, le aree dialettali dell'Alta Italia e della Ladinia intra alpina.

Sulle Fig. 1 e 2 si osservino soprattutto:

1. in una prospettiva più *grossolana*:  
la compattezza colla quale saltano fuori i cinque grandi complessi (macro-gruppi) dialettali seguenti:

- a. Ladinia occidentale e centrale
- b. Piemonte-Aosta
- c. Veneto-Friuli-Trentino
- d. Liguria-Toscana-Emilia-Romagna-Marche
- e. Lombardia-Svizzera italiana.

2. in una prospettiva più *dettagliata*:  
la plausibilità della *corologia* (o *corematica*)<sup>2</sup> dei 25 gruppi (o *coremi*) considerati.  
In linea di massima si tratta di suddivisioni corologiche ben conosciute:

- A—B Bipartizione della Ladinia grigionese.
- C Agglomerazione dei punti-AIS della Ladinia centrale (ad esclusione del P. 316, Cortina d'Ampezzo, piuttosto cadorino dal punto di vista *intra*-linguistico) e affiliazione di questa classe al ramo grigionese. Ne derivano fortissimi argomenti *a favore* della concezione ascoliana dell'“unità (= classe) ladina“.
- D—F La gerarchia dell'albero ci dice che dei tre gruppi classificati, i gruppi E (occitanico alpino e amfizione) e F (piemontese) sono più vicini tra di loro rispetto a D (Aosta con, stranamente, i PP. 140 e 150, ambedue occitanici).
- G—K Non stupisce il fatto che il gruppo G (Friuli) si stacca dal resto del macro-gruppo. Va detto tra parentesi che, sulla base di altri procedimenti gerarchici, il Friuli e la Ladinia centrale vengono accoppiati. Il Trentino (J) e il complesso bellunese-cadorino (K) si spostano un po' dal veneto centrale (I) e dal veneziano della Serenissima (P. 376) nonché dalle regioni di “là da mar“ (H).
- L—R Tra questi gruppi, P (Pavia, Oltrepò pavese) e Q (piacentino, cremonese, mantovano) fanno da amfizione fra tre maggiori poli di attrazione, e cioè: gli spazi lombardo, emiliano e ligure. I gruppi L (Toscana), M (Marche), N (Romagna) e O (Emilia con amfizione) combaciano perfettamente colla visione tradizionale di questo spazio dialettale.
- S—Y L'articolazione del macro-gruppo lombardo è molto interessante. Il lombardo occidentale (Y: imperniato su Milano) ed il lombardo orientale (W: imperniato su Brescia e Bergamo) costituiscono il nucleo di questo macro-gruppo. Da notare la satellizzazione del gruppo X (Valtellina) da parte del lombardo occidentale (Y) e non di quello orientale (W) geograficamente tuttavia più vicino. Al meso-gruppo (W, X, Y) si annettono, via via, i microgruppi U (Val d'Ossola) e V (provincia di Novara), mentre le parlate italo-svizzere (S: Poschiavo e Val Bregaglia; T: Ticino settentrionale = Sopraceneri) si scostano un po'.

E' ovvio che un'analisi comparativa di molti alberi analoghi sarebbe molto più interessante ed anche proficua per la nostra comprensione degli ingranaggi *inter- e intra*-dialettali che non la prospettiva isolata testè presentata.

Prima di chiudere questo paragrafo vorrei sottolineare ancora due punti:

1. la sorprendente *coerenza geografica* all'interno dei 25 gruppi (o coremi) qui esaminati. Alla vicinanza *uni*-dimensionale all'interno dei 25 gruppi della Fig.

<sup>2</sup> I termini “corema” e “corematica” sono stati conati (in francese) dalla Scuola cartografica di Montpellier capeggiata da R. BRUNET (cf. BRUNET 1987, 191 e 211).

*Proiectum:* LEX-ITALIEN

*Fons:* AIS vol. I, II, IV

*Criter. metrolog.:* Variabil. lexic. et morphosyntact.

*Matrix dat.:* N=251 Elem., p= 696 Attr. (scala nomin. polytom.)

*S<sub>jk</sub>:* GIW(1)<sub>jk</sub>/IPI(1)<sub>jk</sub>

*Algor. agglom. hierarch.:* COMPLETE LINKAGE

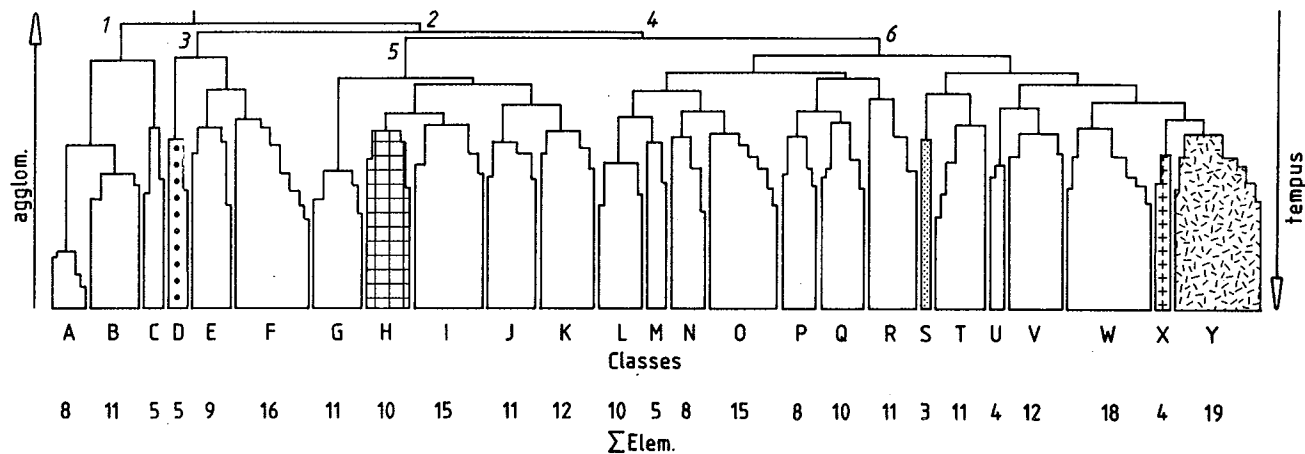


Fig. 1: Classificazione dendrografica di 251 elementi dialettali secondo il metodo "complete linkage".

Per una spazializzazione dei gruppi A—Y si veda la Fig. 2.



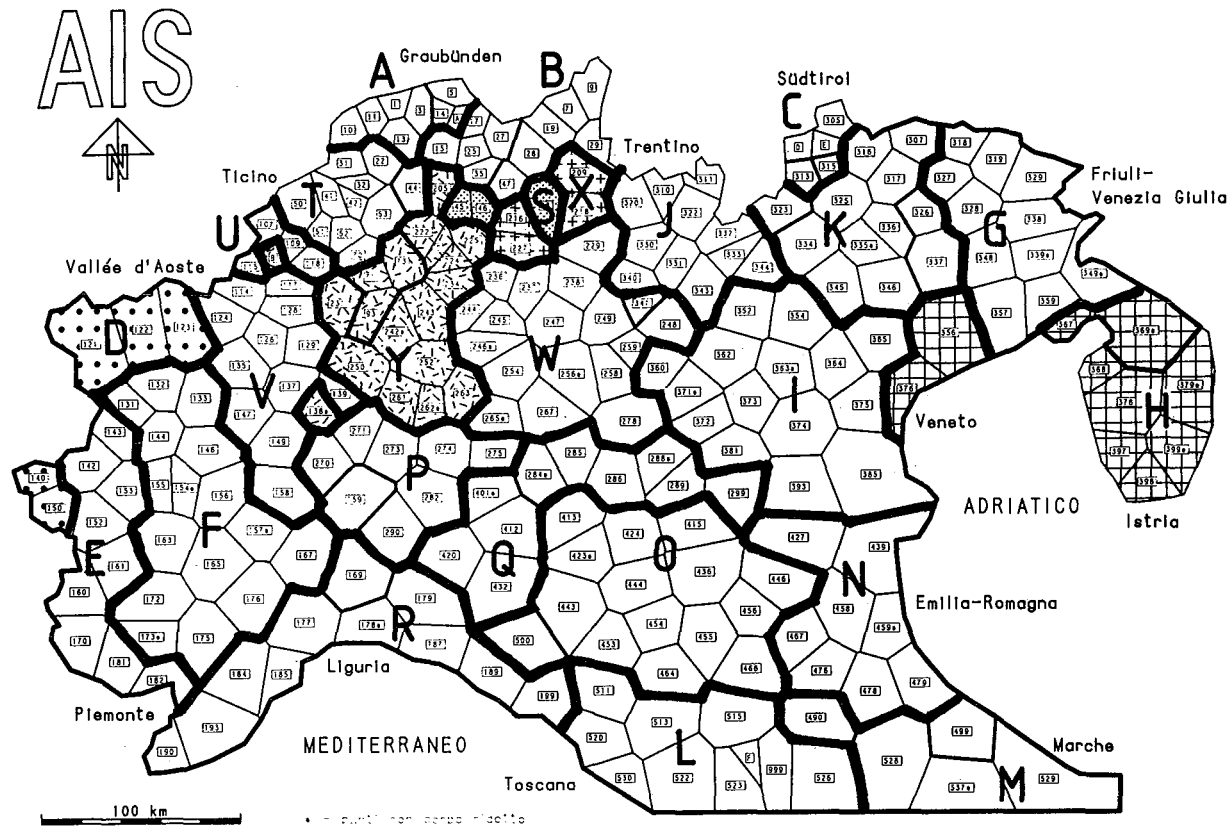


Fig. 2: Spazializzazione del grafo arborescente ("albero") della Fig. 1.

Segle per alcuni punti-AIS:

- |        |        |
|--------|--------|
| A 16   | D 312  |
| B 116* | E 314* |
| C 223* | F 524  |

Il fondo della carta di THIESSEN corrisponde perfettamente a quello utilizzato nei miei studi dialettometrici anteriori (relativi all'AIS; cf. 4. Bibliografia).

1 corrisponde una contiguità (ossia compattezza) *bi*-dimensionale (o geografica) all'interno delle 25 aree (o coremi) corrispondenti della Fig. 2. Rare sono le discontiguità: p. es. i PP. 116 e 138 che fanno da propaggini geografiche del gruppo Y, ed i PP. 140 e 150 che sono prolungamenti territoriali del gruppo D (Aosta). La bipartizione geografica del gruppo H corrisponde però a fatti di dislocazione geografica storicamente accertati (diffusione transadriatica della parlata veneziana).

2. Anche all'*interno* dei 25 gruppi qui presentati ci sono strutturazioni gerarchiche interessanti che, per ragioni di spazio, non possiamo esaminare in questa sede.

### 3.2. Interpretazione *diacronica* delle Figg. 1 e 2

Si tratta dell'applicazione del pensiero *glotto-* (o *lessico-*)*cronologico*. La glottocronologia parte, com'è noto, dal presupposto teorico che la parentela quantitativa tra due idiomi (definiti come tali mediante un certo numero di lessemi: cf. la lista famosa di SWADESH) permette, all'interno di un modello logaritmico, di calcolare — sempre con dati lessicali — il tempo svoltosi sin dalla separazione dei due idiomi, che all'inizio dovevano ovviamente costituire un'entità comune. Pur scartando il problema annoso del calcolo della data delle rispettive frantumazioni, possiamo dire che i presupposti teorici e matematici della glottocronologia su basi lessicali sono molto simili a quelli della dialettometria dendrografica. Visto il parallelismo formale, pare lecito interpretare l'albero della Fig. 1 anche dal punto di vista diacronico. Ciò facendo dobbiamo seguire di nuovo la direzione indicata dalla freccia a destra dell'albero ("tempus"). Ecco dunque la sceneggiatura delle frantumazioni:

1. Dall'unità originaria della rete-AIS si stacca in un primo momento la Ladina occidentale e centrale; risultato: la rete-AIS è bipartita (ramificazioni 1 e 2).

2. Coll'ulteriore andar del tempo la rimanente unità (= classe) *non-ladina* si scinde in due parti; risultato globale: tripartizione (ramificazioni 1, 3 e 4).

3. L'unità (= classe) *non-ladina* e *non-piemontese* (o galloromanza) (= gruppi G—Y) si spezza — grosso modo — lungo l'Adige: risultato: quadripartizione (ramificazioni 1, 3, 5 e 6). E così via.

Le Fig. 1 e 2 consentono un'agevole continuazione di questa sceneggiatura. Senza voler condividere le aspre critiche rivolte dal COSERIU alla glottocronologia nel 1965 — critiche che a distanza di oltre 25 anni appaiono alquanto esagerate e, soprattutto, teoricamente e metodicamente scadute — è ovvio che bisogna essere cauti<sup>3</sup>. Delle due interpretazioni presentate qui sopra la prima mi pare essere molto

<sup>3</sup> Questo vale soprattutto per l'assunzione di date molto precise per le ramificazioni diacroniche; cf. a questo proposito la differenza metodica che corre tra gli scritti di CORTELAZZO (1985) (e con lui di H. GUITER) da una parte, e di CAVALLI-SFORZA/WANG (1986) dall'altra.

più valida ed euristicamente fertile che non la seconda. E possibile però dedurre dalla sceneggiatura diacronica alcuni cenni classificatori importanti. Fra le tante possibilità inerenti vorrei, giunto al termine di questo piccolo dono natalizio, sottolineare soltanto la posizione *eccentrica* della Ladinia occidentale e centrale all'interno della rete-AIS esaminata. Questo fatto (e tanti altri ancora della stessa coniazione dialettometrica; cf. GOEBL 1988) potrebbe ravvivare le discussioni intorno alla "questione ladina" deviate, negli ultimi anni, in acque metodicamente forse troppo calme.

#### 4. BIBLIOGRAFIA

- AIS: Sprach- und Sachatlas Italiens und der Südschweiz, Jaberg, K./Jud, J. (eds.), Zofingen 1928—1940, 8 voll.
- Bock, H.H.: Automatische Klassifikation. Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten (Cluster-Analyse), Göttingen 1974.
- Brunet, R.: La carte. Mode d'emploi, Parigi, Montpellier 1987.
- Cavalli-Sforza, L.L./Wang, W. S—Y.: Spatial Distance and Lexical Replacement, in: Language 62 (1986) 38—55.
- Chandon, J.—L./Pinson, S.: Analyse typologique. Théories et applications, Parigi ecc. 1981.
- Cortelazzo, M.: Sonziaco e gradese: un confronto lessicostatistico, in: Ce fastu? 61 (1985) 31—35.
- Coseriu, E.: Critique de la glottochronologie appliquée aux langues romanes, in: Actes du X<sup>e</sup> Congrès international de linguistique et philologie romanes (Strasburgo 1962), Straka, G. (ed.), Parigi 1965, 87—96.
- Goebel, H.: Éléments d'analyse dialectométrique (avec application à l'AIS), in: Revue de linguistique romane 45 (1981) 349—420.
- Goebel, H.: Dialektometrie. Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie, Vienna 1982 (Denkschriften der Österr. Akademie der Wissenschaften, phil.-hist. Klasse, vol. 157).
- Goebel, H.: "Stammbaum" und "Welle". Vergleichende Betrachtungen aus numerisch-taxonomischer Sicht, in: Zeitschrift für Sprachwissenschaft 2 (1983) 3—44.
- Goebel, H.: Dialektometrische Studien. Anhand italo-romanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF, Tübingen 1984a, 3 voll.
- Goebel, H.: Lineamenti di dialettometria (con applicazione all'AIS), in: Guida ai dialetti veneti 6 (1984b) 7—53.
- Goebel, H.: Points chauds de l'analyse dialectométrique: pondération et visualisation, in: Revue de linguistique romane 51 (1987) 63—118.
- Goebel, H.: Considerazioni dialettometriche sul problema dell'"unità retoromanza (ladina)", in: Mondo ladino 12 (1988) 39—59.

- Jaberg, K.: Il turismo dei dialetti. Caratteristiche e curiosità dell'atlante linguistico-etnografico svizzero-italiano, in: *Le vie d'Italia* 29 (1923) 1185—1195.
- Schleicher, A.: Die Darwinsche Theorie und die Sprachwissenschaft. Offenes Sendschreiben an Herrn Dr. Ernst Häckel, a.o. Professor der Zoologie und Director des zoologischen Museums an der Universität Jena, Weimar 1863.
- Sneath, P.H.A./Sokal, R.R.: Numerical Taxonomy. The Principles and Practice of Numerical Classification, San Francisco 1973.
- Stewart, A.H.: Graphic Representation of Models in Linguistic Theory, Bloomington 1976.

#### Povzetek

#### VREDNOTENJE PODATKOV JEZIKOVNEGA ATLASA AIS. POSKUS KLASIFIKACIJE NAREČJA Z ANALIZO GENEALOŠKEGA DREVEŠA

Opisuje se uporaba genealoškega drevesa pri vrednotenju jezikovnih podatkov, ki omogočajo pravilno klasifikacijo govorov. Tako ovrednoteni podatki so nabrani v gradivu, ki ga nudijo I., II. in IV. zvezek AIS (Jaberg in Jud, Sprach- und Sachatlas Italiens und der Südschweiz, Zofingen, 1928—1940). Podatki se nanašajo na 251 obdelanih točk (samo iz severnega in srednjega dela mreže atlasa), in sicer za 696 pojmov, s čimer je mogoče dobiti dokaj verno sliko lokalnega govora in preko te, s pomočjo primerjave podatkov, podobo dialekta. Genealoško drevo (skica 1) se ustvari s pomočjo taksometričnega algoritma. Nato se razvejana struktura uporabi na mreži atlasa (skica 2). Tako nastane podoba narečne razcepljenosti za severno italijanščino in retoromanščino (ladinščino). Ocenjujemo jo tako s sinhronega vidika, torej geolingvistično, ali pa z diahronega, kar bi lahko imenovali statistiko besedišča ali celo glotohronijo, torej pogled na razvoj jezika.

S tem se potrdi umestnost in praktična korist rezultatov, dobljenih s pomočjo opazovanja dialekta, ki izhaja iz primerjave razvejanosti genealoškega drevesa.