L'ordine delle parole

Andrea Sgarro, Liviu P. Dinu*

Università di Trieste
Università di Bucarest

Abstract

We put forward two mathematical tools, called the rank distance and the bubble distance, to compare the ordinal structure of natural languages. Traditionally, ordinal structures have only been studied in a very basic and "abstract" context, as for example, the six idealised arrangements for the subject-verb-object sequence. Our tools allow one to get an overall picture of the ordinal distance between a text and its translation. Thinking of future work, availability of web resources will make it possible to conduct large-scale experiments, which will account for the difficulties experienced by simultaneous interpreters, say between Italian and German, or even more so between Italian and Turkish. However, if one wants to get a realistic picture of these differences, one has to abandon the traditional approach to mathematics, based on binary logics (true vs. false, black vs. white), and introduce the grey nuances made available by soft logic; in other words one has to relinquish the esprit de géometrie *and adopt the* esprit de finesse *which is typical of soft computing. It is no coincidence that soft logic has been largely inspired by the flexible logical structures which are possessed by natural languages, rather than those of "hard" mathematics. A crisp mathematical approach will do for simple and abstract structures such as subject-verb-object, but it is wholly insufficient if one moves to the overall word order in actual texts.*

* Andrea Sgarro, DMUI, Università di Trieste.
Liviu P. Dinu, Facoltà di Matematica, Università di Bucarest.

Che l'ordine delle parole in tedesco sia diverso da quello in italiano è una circostanza risaputa di cui qualunque interprete simultaneo ha fatto le spese in cabina. Più facile (da questo punto di vista) è il compito del collega italiano che interpreta in simultanea dall'inglese, ma la soluzione del problema di Cipro porterà a Bruxelles gli interpreti dal turco, e le loro lamentele supereranno quelle dei colleghi di tedesco. Dunque, la percezione che abbiamo è: l'ordine delle parole in inglese e in italiano è abbastanza simile, lo è di meno fra tedesco e italiano, turco e italiano poi sono decisamente distanti dal punto di vista *ordinale*. È possibile rendere queste affermazioni più precise, più "matematiche"? Le lingue naturali sembrano essere troppo sfuggenti, con elementi di vaghezza o persino di ambiguità, per prestarsi a venir matematizzate secondo i principi dell'*esprit de géometrie* di cui parlava Pascal. In realtà, per merito soprattutto dell'intelligenza artificiale e dei sistemi esperti, sono ormai decenni che l'*esprit de finesse* ha fatto il suo ingresso in matematica: la computazione flessibile, o *soft computing*, rispetto a quella tradizionale, *rigida* o *hard*, consente di trattare in maniera più adeguata strutture non ben definite, senza doverle rinchiudere nei ceppi della matematica tradizionale. Del resto molti dei concetti delle nuove logiche *soft* provengono proprio dallo studio delle lingue naturali e delle strutture logiche che ne sono alla base, strutture che sono ben lontane dalla durezza binaria del sì o no, del bianco o nero senza gradazioni di grigio, senza sfumature o *sfocatezza* (*fuzziness*), per adoperare un termine tecnico del *soft computing*.

Torniamo al nostro tema e partiamo da una questione che è ben trattata nella tipologia linguistica, quella dell'ordine di soggetto S, oggetto O e verbo V nelle frasi affermative non marcate, per intenderci nelle principali quando sono enunciate senza particolari connotazioni emotive. Al latino *mater filium amat* si contrappone l'italiano *la madre ama il figlio*, si passa insomma dall'ordine non marcato SOV del latino a quello SVO dell'italiano e delle grandi lingue europee moderne (inglese, francese, spagnolo, russo, in una certa misura tedesco, ecc.). Alle porte dell'Europa troviamo un'altra grande lingua, il turco, che ha l'ordine latino SOV, accompagnata in questo dal fārsi, ossia dal persiano moderno, per citare una lingua indoeuropea dell'Asia. Lingue celtiche come l'irlandese o il gallese, ritornando in Europa, hanno l'ordine non marcato VSO. In questi tre esempi ordinali il soggetto S precede sempre l'oggetto O, ma non si tratta di una regola tipologica universale, "deterministica", al più lo è di tipo statistico: i tipologi hanno scovato tutti e sei gli ordinamenti possibili, anche se quelli in cui O precede S sono assai poco frequenti: l'ultima a resistere è la sequenza OSV, su cui non si hanno dati certi (Grandi 2003). Se confrontiamo l'ordine latino SOV e quello italiano SVO, oppure quello italiano SVO e quello irlandese VSO, notiamo che sono a distanza di un unico scambio fra posizioni consecutive l'una dall'altra: dal latino SOV passiamo a SVO permutando O e V, dall'irlandese VSO permutando V e S. Invece se confrontiamo direttamente il latino e l'irlandese gli scambi necessari salgono a due. In questo senso la distanza ordinale fra latino e italiano

oppure fra irlandese e italiano è pari a 1, mentre è 2 nel caso di latino e irlandese. A distanza 1 dall'italiano SVO stanno le sequenze VSO (irlandese) e SOV (latino, turco, persiano), a distanza 2 abbiamo la VOS e la contestata OSV, mentre a distanza 3 sta la sequenza OVS, tipica di lingue amerindie come lo *hixkaryana* brasiliano (Grandi 2003). Ma come comportarsi in presenza di strutture ordinali più complesse, che non si limitino a tener conto dei soli soggetto, oggetto e verbo? A servizio della tipologia linguistica porremo due DISTANZE ORDINALI, la *distanza di scambio* (*twiddle distance* o anche *bubble distance*) e la *distanza di rango* (*rank distance*), quest'ultima già collaudata nella linguistica computazionale (cfr. Dinu & Dinu 2005). Proporne due invece di una sola è di per sé un segno di *esprit de finesse*: una caratteristica tipologica di tipo ordinale verrà riconosciuta come significativa solo se è sufficientemente *stabile* rispetto a entrambe le due distanze (rispetto a tutti e due i modi di valutarla numericamente).

2. La matematica degli anagrammi

Per il momento, gli oggetti **x** e **y** di cui vogliamo calcolare la distanza, invece che frasi in due lingue, saranno stringhe di **n** simboli distinti che sono anagramma l'una dell'altra: le due stringhe **x** e **y** sono composte degli stessi **n** simboli, a distinguerle è solo l'ordine in cui i simboli compaiono nelle stringhe.[1] Partiamo dalla distanza di scambio **dS**. Potremmo ad esempio avere **x** = ROMA e **y** = AMOR. La distanza di scambio **dS**(**x**,**y**) = **dS**(ROMA , AMOR) è per definizione il minimo numero di scambi fra posizioni consecutive (*twiddles*) necessari a "riportare" la seconda stringa alla prima. Nel caso nostro: AMOR → MAOR → MOAR → OMAR → OMRA → ORMA → ROMA, per cui **dS**(AMOR , ROMA) = 6. Invece **dS**(ROMA , ORMA) = 1: basta infatti scambiare di posto i primi due simboli.

Veniamo ora alla distanza di rango **dR**, la cui definizione è più delicata.[2] Prendiamo due stringhe anagrammate di lunghezza **n**, per esempio di nuovo ROMA e AMOR. Passando da ROMA ad AMOR, la R si sposta di 3 posizioni in

1 Più che di anagrammi, i matematici preferiscono parlare di *permutazioni*. Il numero delle permutazioni di **n** simboli è **n**! (**n** fattoriale), ossia il prodotto di tutti i numeri interi da 1 a **n**. Al crescere del parametro **n** la crescita del fattoriale **n**! è rapidissima, perfino più di quella che viene considerata la tipica crescita esplosiva, ossia quella esponenziale. Si noti che nelle nostre stringhe stiamo escludendo la possibilità di avere simboli ripetuti, ma le distanze ordinali possono venir estese a questa situazione. Si pensi alle stringhe di DNA che arrivano a lunghezze di centinaia di milioni di simboli pur essendo "scritte" con sole quattro "lettere", A = adenina, C = citosina, G = guanina, T = timina (cfr. ad es. Dinu & Sgarro 2006a).

2 Beninteso il calcolo delle due distanze si può meccanizzare attuandolo in *software* mediante opportuni *algoritmi*. Se la distanza di scambio **dS** è più facile da spiegare, essa è più difficile da computare concretamente, nel senso che i tempi di calcolo sono più lunghi (in termini tecnici dalla complessità *lineare* si sale a quella *quadratica*); i vantaggi di tipo computazionale della *rank distance* possono portare a differenze drammatiche se le stringhe di simboli sono molto lunghe, come succede con stringhe di DNA. Per inciso, il curioso nome di *bubble distance* è legato a un ben noto algoritmo di ordinamento o *sorting* che viene spiegato in classe ricorrendo alla metafora di bollicine che salgono nell'acqua.

avanti, la O di una posizione, la M di una posizione indietro e la A di tre posizioni. Sommando questi quattro *scarti* si ottiene la distanza di rango fra le due stringhe: **dR**(ROMA , AMOR) = 3+1+1+3 = 8. È irrilevante se lo spostamento sia in avanti o indietro, conta solo la sua ampiezza. Altro esempio: **dR**(MARTE , TREMA) = 3+3+1+3+2 = 12, perché 3 è lo scarto per la M, per la A e per la T, 1 per la R e 2 per la E. Se prendiamo due stringhe che differiscono di un singolo *twiddle*, come ROMA e ORMA, troviamo **dR**(ROMA , ORMA) = 2: la distanza di rango dà il valore 2 allo scambio singolo, e ha dunque una *unità di misura* diversa rispetto a **dS**. Per tale ragione, *d'ora in avanti le distanze di rango verranno sempre divise per 2*: due stringhe a distanza unitaria differiranno per un singolo scambio qualunque delle due distanze si sia scelta, *bubble* o *rank*.

Diamo di seguito alcuni risultati matematici sulle due distanze (per le dimostrazioni cfr. Diaconis & Graham 1977). Essi si riferiscono, una volta fissata la lunghezza **n** delle stringhe, al *valore massimo* e al *valore casuale* (*random*); quest'ultimo è il valore della distanza statisticamente atteso (*expected distance*) che si ottiene permutando la prima stringa del tutto a caso:

rand **dS** (x,y) = (**n**² – **n**) / 4 , max **dS** (x,y) = (**n**² – **n**) / 2

rand **dR** (x,y) = (**n**² – 1) / 6 , max **dR** (x,y) = (**n**² [– 1]) / 4

(Il termine fra parentesi quadra va omesso dall'espressione del rango massimo quando **n** sia un numero pari).

Diamo i valori concreti nel caso di stringhe di lunghezza **n** da 3 a 10 (i puntini indicano che l'ultima cifra si ripete all'infinito):

n	3	4	5	6	7	8	9	10
rand **dS**	1,5	3	5	7,5	10,5	14	18	22,5
max **dS**	3	6	10	15	21	28	36	45
rand **dR**	1,3...	2,5	4	5,83...	8	10,5	13,3...	16,5
max **dR**	2	4	6	9	12	16	20	25

Come si vede i campi di variazione sono diversi e la *bubble distance* è decisamente più dilatata. In entrambi i casi il valore massimo corrisponde alla stringa rovesciata (letta da destra a sinistra), ma nel caso della *rank distance* ci sono anche altre stringhe che danno il valore massimo. Ad esempio **dS**(ROMA , AMOR) = 6 e **dR**(ROMA , AMOR) = 4 sono i valori massimi di scambio e di rango quando la lunghezza è **n**=4, ma anche **dR**(ROMA , AMRO) dà il valore massimo di rango 4 senza che Roma e Amro siano palindromi, mentre **dS**(ROMA , AMRO) è soltanto 5 e non 6. L'analisi mostra che la distanza di rango diventa poco discriminante di fronte a distanze alte, prossime al valore massimo, ciò che peraltro spiega come mai essa si calcoli così rapidamente. Spiega anche la curiosa anomalia del valore casuale, che nel caso della *bubble distance* sta esattamente a metà strada fra lo zero e il valor massimo, ma sta a circa due terzi di strada nel caso della *rank distance*.

Prima di procedere vogliamo fare un'osservazione importante. Tutto sommato non è poi così fuorviante che l'ebraico e l'arabo si scrivano "a rovescio", da destra a sinistra: l'immagine speculare non maschera né rimescola in maniera grave l'ordine originale. Bisogna insomma distinguere fra due stringhe a distanza

massima, immagine speculare l'una dell'altra, e due stringhe le cui strutture ordinali non sono *correlate* l'una rispetto all'altra, per cui l'ordine nella prima non c'entra o quasi con l'ordine della seconda. La situazione di *massima confusione* non corrisponde a quella di massima distanza, ma piuttosto a quella in cui la distanza ha un valore vicino al valore casuale, quello che si ottiene rimescolando "a vanvera" i simboli della stringa di partenza. Dunque: *valori vicini allo zero indicano che le due stringhe hanno grossomodo la stessa struttura ordinale, valori vicini al massimo indicano che le due stringhe hanno grossomodo la stessa struttura ordinale però ribaltata, mentre valori vicini al valore casuale indicano che le strutture ordinali delle due stringhe hanno ben poco a che vedere l'una con l'altra*.

Vogliamo concludere questa sezione matematica con un'osservazione che è dettata soltanto dal desiderio di rendere ben *leggibili* i dati numerici che via via si trovano. Purtroppo, non solo le due distanze hanno campi di variazione diversi, più ampio quello della *bubble distance*, ma l'estensione dei campi dipende da **n**, ossia dalla lunghezza delle stringhe (dal numero di simboli che le compongono). Per confrontare distanze relative a lunghezze diverse è allora conveniente riportarle tutte all'intervallo da 0 a 100. Trasformazioni di questo tipo sono molto comuni in matematica e si chiamano *normalizzazioni*. Dopo la normalizzazione, il valore casuale della *bubble distance*, stando proprio a metà strada fra 0 e il massimo, diventa 50. Nel caso della *rank distance* conviene ricorrere a una doppia normalizzazione, normalizzando i valori della distanza di rango fino al valore casuale sull'intervallo che va da 0 a 50 e i valori superiori a quello casuale sull'intervallo da 50 a 100.[3]

3. L'ordine delle parole: *l'esprit de géometrie*

Quanto abbiamo imparato sugli anagrammi può venir posto a buon fine per studiare strutture più complesse di quella della prima sezione. Pensiamo per esempio alla frase non marcata *la madre buona non ama il figlio cattivo* con l'ordinamento soggetto + attributo del soggetto + negazione + verbo + oggetto + attributo dell'oggetto; in simboli SANVOB, di lunghezza n=6 (A = attributo del soggetto, B = attributo dell'oggetto). La stessa struttura diventa ASNVBO in inglese (*a good mother does not like a bad son*), ASVBON in tedesco (*die gute Mutter liebt den bösen Sohn nicht*), SAOBNV in persiano (*mādar-e khub pesar-e bad rā dust nadārad*) e ASBOVN in turco (*iyi anne kötü oğlu sevmiyor*). Ciò consente di costruire le due matrici (tabelle) seguenti con le distanze non normalizzate, la prima per gli scambi e la seconda per i ranghi:

3 La normalizzazione è una trasformazione per proporzionalità: se l'intervallo [a,b] deve venir trasformato nell'intervallo [A,B], il corrispondente X compreso fra A e B di un qualunque valore x compreso fra a e b si trova con la formula: X = [B(x-a) – A(x-b)] / (b-a). In una tabella simile alla precedente ma con le distanze normalizzate, nelle due righe dei massimi ci sarebbe scritto sempre 100, e nelle due righe dei valori casuali sempre 50. Anche dopo la normalizzazione i valori concreti delle due distanze, *bubble* e *rank*, in generale *non* coincidono.

dS	italiano	inglese	tedesco	turco	persiano
italiano		2	5	7	4
inglese	2		3	5	6
tedesco	5	3		2	5
turco	7	5	2		3
persiano	4	6	5	3	

dR	italiano	inglese	tedesco	turco	persiano
italiano		2	4	5	4
inglese	2		3	4	5
tedesco	4	3		2	4
turco	5	4	2		3
persiano	4	5	4	3	

Le due *diagonali principali*, che dovrebbero essere piene di zeri, sono state lasciate vuote. Si osservi che, anche se i valori numerici sono talvolta diversi, le relazioni di ordinamento sono simili nelle due tabelle. Si confrontino i valori trovati con i valori casuali delle tabelle precedenti ponendo n=6, ossia 7.5 per gli scambi e 5.83 per i ranghi, valori quasi realizzati nel caso della distanza fra italiano e turco.

Emergono subito degli inconvenienti dovuti alla rigidezza dei nostri strumenti. In inglese, in turco e in persiano il verbo in realtà "avvolge" la negazione. Pensiamo all'ausilare *does* richiesto in inglese dalla forma negativa, o al fatto che la negazione *m[e]* in turco separa il tema *sev* (la parte semantica) dalla parte sintattica *iyor* che segnala la terza persona singolare del presente (*seviyor* = ama, *sevmiyor* = non ama); il verbo amare in persiano è un composto come succede spessissimo in questa lingua, *amare* = *dust d shtan* ossia *aver amico*: il sostantivo *dust* (amico) precede la negazione *na* (scritta attaccata al verbo, ma ciò è irrilevante). Nell'ultimo caso è stato facile dirimere il dubbio, sostituendo il verbo composto con uno semplice come *nemibinad* (non vede), ma è già evidente che l'*esprit de géometrie* crea problemi e tende a ingabbiare la "libertà" linguistica.[4]

Riprenderemo la difficoltà nella sezione seguente, prima vogliamo mostrare qualche modesto e del tutto preliminare risultato sperimentale non su strutture ma su testi. Una volta affinati gli strumenti, è sui testi che si dovrebbero condurre le sperimentazioni più ambiziose, quelle che porterebbero a numerizzare la sensazione dell'interprete in cabina di quanto profonda sia la differenza ordinale fra due lingue. Ci siamo serviti di una raccolta di racconti bilingui pubblicata dalla dtv per celebrare la sua serie di traduzioni in tedesco con il testo originale a lato (Niemeyer 1998). Nell'elaborazione che segue il tedesco funge da lingua di partenza, mentre le *lingue di arrivo* nella nostra scelta sono l'originale inglese, il francese, l'italiano, lo spagnolo, il russo e il turco. È stato estratto un paragrafo di una quindicina di righe da ciascun racconto, sono stati numerati nomi, aggettivi e verbi nell'ordine in cui compaiono nella traduzione tedesca – e non nell'originale, una scelta comoda ma di certo opinabile – e sono stati ricercati i corrispondenti negli originali. Per dare un esempio: dalla traduzione tedesca *Ein steiler* 1 *sandiger* 2 *Weg* 3 *lief* 4 *durch die Bäume* 5 *den Hugel* 6 *hinab zum Bucht* 7 si

4 Beninteso ciò succede già con strutture del tutto elementari come quella della prima sezione.

ritorna all'originale inglese *A steepy* 1 *sandy* 2 *road* 3 *ran* 4 *down the hill* 6 *to the bay* 7 *through the timber* 5, per poi calcolare la distanza di scambio e di rango della sequenza di numeri interi 1234675 (inglese) dalla sequenza di partenza 1234567 (tedesco) che vede gli interi da 1 a 7 nel loro ordine naturale. Poiché i paragrafi sono di lunghezza un po' diversa siamo ricorsi a distanze normalizzate fra 0 e 100 e arrotondate al secondo decimale (le distanze assolute sono state comunque scritte per completezza). Inutile dire che i risultati ottenuti non hanno nessuna significatività statistica, e servono più che altro a mettere alla prova il metodo, per poi capire dove vada ritoccato. L'approccio rigido ha in questo caso degli inconvenienti evidenti, cui accenneremo nel paragrafo successivo.

lingua	lunghezza	dR	dR norm.	dB	dB norm.
français	45	21	3,06	25	2,53
español	43	22	3,57	26	2,88
türkçe	39	19	3,75	21	2,83
russkiĭ	52	30	3,33	30	2,26
english	51	18	2,08	21	1,65
italiano	34	15	3,90	19	3,39

Distanze dalla traduzione in tedesco (sopra) e da quella in italiano (sotto).

lingua	lunghezza	dR	dR norm.	dB	dB norm.
français	45	11	1,63	11	1,11
español	43	0	0	0	0
türkçe	39	20	3,95	23	3,10
russkiĭ	52	21	2,31	24	1,81
english	51	5	0,06	5	0,39

Rispetto alle strutture schematiche di prima le nuove distanze sono decisamente più contenute: ciò fa supporre che l'ordine delle parole possa essere una caratteristica profonda "largamente" comune a qualsiasi lingua naturale. L'indicazione per la tipologia strutturale sarebbe allettante, ma purtroppo essa manca di una convalida statisticamente significativa.

4. L'ordine delle parole: *l'esprit de finesse*

Nel momento in cui applichiamo al materiale linguistico effettivo i nostri strumenti matematici nella loro forma nitida (*crisp*, *hard*) emergono chiaramente (almeno[5]) due inconvenienti:

5 Oltretutto non c'è una corrispondenza *crisp* fra le parti del discorso: si pensi a *during*, preposizione che nella traduzione diventa la locuzione preposizionale *im Lauf des*, *nel corso di*, per cui il sostantivo *Lauf* si presenta in uno solo dei due testi. Aggiungiamo anche che le due lingue non entrano in gioco in maniera simmetrica, per cui abbiamo a che fare con "distanze da" (distanze orientate) più che con "distanze fra" (distanze non orientate, simmetriche). Tutto ciò va pesato con cura prima di imbarcarsi in sperimentazioni statistiche su materiale esteso; ad ogni buon conto, lo strumentario *fuzzy* è abbastanza duttile per gestire tali inconvenienti.

1. a un unico elemento numerato nel testo di partenza (nel nostro caso quello tradotto) possono corrispondere più elementi nel testo di arrivo (nel nostro caso quello originale); oppure, simmetricamente:
2. a più elementi numerati del testo di partenza può corrispondere un unico elemento nel testo di arrivo.

Di conseguenza, se vogliamo evitare forzature eccessive:
1. Nel testo di arrivo più elementi possono venir contrassegnati con lo stesso indice.
2. Nel testo di arrivo un singolo elemento può essere contrassegnato da più indici diversi.

In Sgarro e Dinu (2006 b) si mostra come distanze *crisp* possano venir generalizzate a distanze di tipo *fuzzy*, in modo da poter gestire aspetti di vaghezza o sfocatezza (*fuzziness*) inevitabili in un contesto come quello linguistico, ma anche in quello biologico. Senza entrare in eccessivi dettagli tecnici e a costo di immiserire la portata dell'approccio *fuzzy*, vediamo come si possa ovviare ai due inconvenienti, partendo dal primo. Supponiamo di partire dalla frase *Questa sera andrò al cinema* e di arrivare al suo corrispondente tedesco *Diesen Abend werde ich ins Kino gehen*, e di voler marcare sostantivi e verbi. Dalla sequenza di partenza *sera* 1 *andrò* 2 *cinema* 3 si arriva a *Abend* 1 *werde* 2 *Kino* 3 *gehen* 2, ossia da 123 si arriva a 1232. In uno spirito *crisp*, dovremmo optare per uno solo dei due ordinamenti semplici compatibili con 1232, ossia 123 oppure 132, depennando la prima o rispettivamente la seconda presenza dell'indice 2. Invece in uno spirito *soft* o *fuzzy*, entrambi gli ordinamenti possono convivere, il che dà luogo a due distanze *rank* che sono 0 e 1, e a due distanze *bubble* dalla stringa di partenza, di nuovo 0 e 1. *La distanza fuzzy è sia* 0 *sia* 1; se, invece di 2 ordinamenti semplici, ce ne fossero **k** compatibili con l'ordinamento di arrivo, la distanza *fuzzy* avrebbe **k** valori tutti "veri" e tutti corretti.[6] In pratica sorge l'esigenza di *aggregare* questi **k** valori della distanza in un unico valore di consuntivo che potrebbe ben essere la loro media aritmetica (la loro somma divisa per **k**). Quest'operazione di "messa a fuoco" (*defuzzification*) porta nel nostro esempio alla distanza numerica ½, che è la media aritmetica fra 1 e 2.

In maniera analoga si risolve l'inconveniente numero 2. Passiamo dall'italiano *Questa mattina* 1 *sono* 2 *andato* 3 *in banca* 4 all'inglese *This morning 1 I went* (2,3) *to the bank* 4. Nella traduzione i due indici 2 e 3 sono stati scritti nel loro ordine naturale, ma avremmo anche potuto scrivere (3,2): le parentesi indicano proprio che al loro interno gli indici non vanno intesi come ordinati, anche se si è dovuto scriverli uno dopo l'altro a causa della natura lineare della scrittura. I due ordinamenti semplici compatibili con 1(2,3)4 sono 1234 e 1324, che portano alle due distanze "parziali" 0 e 1 (*rank* o *bubble*, indifferentemente): esse possono venir aggregate nel singolo valore di ½, come si è fatto prima.

6 Nella logica *fuzzy* il principio di non contraddizione e il *tertium non datur* cadono (Dubois & Prade 2000).

Per tornare al verbo turco che avvolge la negazione *sev-m[e]-iyor*, possiamo benissimo pensare a un ordinamento sfocato del tipo VNV, che congloba i due ordinamenti rigidi VN e NV, laddove nella sezione precedente avevamo conservato solo il primo. La distanza fra l'italiano SANVOB e il turco ASBOVNV (il verbo V è ripetuto) scenderebbe allora a 6,5 per gli scambi e 4,5 per i ranghi.

5. Per una tipologia linguistica delle strutture ordinali

Abbiamo proposto due distanze per lo studio delle strutture d'ordine nelle lingue naturali; è importante che questi strumenti matematici vengano resi sufficientemente flessibili, nello spirito del *soft computing*, per non introdurre illusoria precisione in un contesto che è per propria natura definito in maniera "fluida". Le strutture d'ordine trattate finora dalla tipologia linguistica sono semplici e astratte, com'è il caso della terna soggetto-verbo-oggetto SVO, ma strumenti più potenti permettono di affrontare obiettivi molto più ambiziosi, anche se, beninteso, converrà prima affinare la tecnica su testi di prova, e solo dopo sperimentare in grande. La disponibilità di dizionari e di traduttori automatici in rete consente di lavorare su masse di materiale linguistico praticamente illimitato. La sensazione di difficoltà più o meno grande che l'interprete simultaneo sperimenta istintivamente in cabina potrà venir ricondotta a distanze ordinali accuratamente valutate, dove l'avverbio *accuratamente* non implica l'introduzione dell'artificiosa esattezza che è inevitabile se si rimane chiusi nell'*esprit de géometrie*.

Riferimenti bibliografici

Diaconis P. & Graham R.L. (1977) "Spearman footrule as a measure of disarray", *Journal of the Royal Statistical Society*, Series B, 39:2, pp. 262-268.

Dinu L.P. & Dinu A. (2005) "On the syllabic similarity of romance languages", in *Lecture Notes in Computer Science*, Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Text Processing, CICLing 2005 (Mexico City, February 13-19 2005). Ed. by A. Gelbukh, pp. 785-788.

Dinu L.P. & Sgarro A. (2006a) "A low-complexity distance for DNA strings", *Fundamenta Informaticae*, 73:7, pp. 361-372.

Dinu L.P. & Sgarro A. (2006b) "Rank distance: a soft tool for comparison of DNA strings", in *Proceedings of 10th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, IPMU 2006 (Paris, July 2006), pp. 2791-2798.

Dubois D. & Prade H. (eds) (2000) *Fundamentals of Fuzzy Sets*, Dordrecht, Kluwer Academic.

Grandi N. (2003) *Fondamenti di tipologia linguistica*, Roma, Carocci.

Niemeyer H. (Hg.) (1998) *Liebe hat vielen Sprachen. Kurzgeschichten der Modernen Literatur, Original Texte und Übersetzungen*, München, dtv.