DOI: 10.13137/2039-8646/37279

Filogenesi linguistica e distanze metriche[*]

A Solomon Marcus (1925-2016)

Andrea Sgarro[**]
Dipartimento di Matematica, Informatica e Geoscienze
Università di Trieste
sgarro@units.it

Anca Dinu[**]
Faculty of Foreign Languages and Literatures
Universitatea din București (Romania)
ancaddinu@gmail.com

Liviu P. Dinu[**]
Department of Computer Science
Universitatea din București (Romania)
ldinu@fmi.unibuc.ro

Laura Franzoi
Human Language Technologies Research Center
Universitatea din București (Romania)
laura.franzoi@gmail.com

Paolo Marega
Dipartimento di Matematica, Informatica e Geoscienze
Università di Trieste
paolo.marega@studenti.units.it

Abstract

In computational linguistics, and more specifically in linguistic phylogeny, mathematical distances play a basic role. We introduce a class of distances not dealt with in literature which verify the triangle inequality and which we shall call α-distances*; we successfully apply them to the historical data of Žarko Muljačić on Romance languages, aiming to extend their use to the sintactical data of Giuseppe Longobardi relative to about one hundred languages of the Old World, data which are leading to results of extraordinary interest in linguistic phylogeny.*

[*] *Title: Linguistic phylogeny and metric distances / Titlul: Filogenie lingvistică și distanțe metrice.*
[**] Ulteriore afferenza: *Human Language Technologies Research Center*, Bucarest (Romania).

PAROLE CHIAVE

DISTANZE METRICHE / METRIC DISTANCES / DISTANȚE METRICE; LOGICA FUZZY / FUZZY LOGIC / LOGICĂ FUZZY; FILOGENESI LINGUISTICA / LINGUISTIC PHYLOGENY / FILOGENIE LINGVISTICĂ; LINGUA DALMATICA / DALMATIC LANGUAGE / LIMBĂ DALMATICĂ.

1. LINGUISTICA MATEMATICA (O COMPUTAZIONALE) E DISTANZE (QUASI-)METRICHE

Ricorre nel 2025 il centenario, patrocinato dall'UNESCO, della nascita di Solomon Marcus, centenario celebrato con solennità nel suo Paese, la Romania, ma forse trascurato al di fuori dei confini romeni.

Solomon Marcus è uno dei padri fondatori della *linguistica matematica* o *linguistica computazionale*, un ramo poco conosciuto dai non-specialisti, anche se il dominio scientifico cui appartiene, il *natural language processing*, è uno dei pilastri dell'Intelligenza Artificiale: basti pensare alla traduzione automatica o alla possibilità di conversare con la *machina sapiens*[1] servendosi ad esempio di ChatGPT o di DeepSeek.

Qui di seguito, riprendendo quanto in precedenza esposto[2], vogliamo dare un'idea di quale possa essere uno degli oggetti di cui la linguistica computazionale si occupa, centrando il nostro discorso sulla nozione squisitamente matematica di *distanza* e più nello specifico di *distanza metrica*. Il concetto di *distanza*[3], onnipresente in matematica e nelle sue molteplici applicazioni[4], può fornire spunti ai docenti per ampliare l'orizzonte culturale degli studenti e aprire una discussione in classe sul ruolo della matematica.

Le *distanze metriche* sono distanze che i matematici amano perché soddisfano a una serie molto esigente di assiomi che si rivelano straordinariamente produttivi, anche se la vita reale non è sempre generosa per cui buona parte delle distanze descritte in DEZA, DEZA (2016)[5] non sono affatto metriche (per una carrellata su distanze poco disciplinate si veda SGARRO, FRANZOI (2023)[6].

---

[1] Cfr. CRISTIANINI 2024.
[2] Cfr. SGARRO, FRANZOI 2023, pp. 44-54.
[3] Cfr. DEZA, DEZA 2016.
[4] Spigolando dai titoli dei capitoli dell'opera citata in nota 3: probabilità e statistica, codici, ingegneria e controlli, immagini e audio, internet, biologia, fisica, chimica, geografia, geofisica, astronomia, cosmologia, teoria della relatività, medicina, criminologia.
[5] Cfr. DEZA, DEZA 2016.
[6] Cfr. SGARRO, FRANZOI 2023, pp. 44-54.

Si parte da un insieme, o se volete da uno *spazio S*, i cui oggetti (i cui elementi), in numero finito o infinito, denoteremo con lettere come x, y o z (per evitare banalità converrà pensare che gli oggetti distinti siano almeno tre, anche se i matematici più diligenti si accontenterebbero di imporre che S non sia vuoto). Ad ogni coppia (x, y) di oggetti distinti o anche coincidenti (potrebbe ben essere $x = y$) viene associato un numero non negativo $d(x, y)$ che è appunto la distanza fra x e y.
Inutile dire che per poter usare un termine impegnativo come "distanza" bisogna richiedere qualche proprietà di regolarità in modo da escludere "comportamenti inammissibili". La distanza d si dice *metrica* se, qualunque siano gli oggetti x, y e z, distinti o coincidenti, valgono i seguenti tre *assiomi*:

$d(x, y) = 0$ se e solo se $x = y$ (solo le *auto-distanze* sono nulle);
$d(x, y) = d(y, x)$ (simmetria della distanza);
$d(x, z) + d(z, y) \geq d(x, y)$ (disuguaglianza triangolare).

I simboli $\geq$ e $\leq$ significano "maggiore o uguale a" e rispettivamente "minore o uguale a". La *disuguaglianza triangolare* è del tutto "naturale": nel primo membro si va da x a y passando per z e c'è dunque un *vincolo* da rispettare, nel secondo di vincoli non c'è traccia; inutile osservare che la normalissima *distanza euclidea* nel piano o nello spazio della geometria cartesiana è appunto metrica. Una generalizzazione molto mite consente di passare alle distanze *quasi-metriche*: il primo assioma viene indebolito a

i′) $0 \leq d(x, x) \leq d(x, y)$ (le auto-distanze potrebbero essere strettamente positive, ma comunque non possono superare le *etero-distanze*).

La lettrice o il lettore potrebbero obiettare che l'assioma più artificioso è casomai il *ii*), quello della simmetria; basti pensare a un percorso di andata/ritorno in auto in centro città con i sensi vietati, dove andare da x a y non è affatto lo stesso che fare il viaggio di ritorno, ma tant'è: conserveremo la simmetria e rinunceremo alle autodistanze necessariamente nulle.
La generalizzazione è mite, per trasformare una distanza quasi-metrica $d(x, y)$ in una

metrica rigorosa $d^*(x, y)$ basterebbe "imporre" che tutte le auto-distanze siano nulle anche se in partenza non lo erano, e il gioco sarebbe fatto com'è facile verificare. Non lo faremo: nel nostro contesto linguistico, e lo si vedrà, la possibilità di avere auto-distanze strettamente positive è significativa.

2. Filogenie linguistiche: la distanze di Muljačić e la lingua dalmatica

Per un linguista un problema cruciale è quello di classificare le lingue e scoprire le eventuali parentele che lascino intravedere l'esistenza di un progenitore comune: per dire, è del tutto evidente che italiano e francese siano apparentati e non stupisce nessuno l'esistenza di un progenitore comune neppure molto lontano nel tempo, in questo caso il latino. Ma il russo e l'italiano sono apparentati, ciò che è evidente agli specialisti? E il giapponese e il coreano lo sono? qui persino lo specialista avrebbe forti incertezze a rispondere; il progenitore comune, ammesso esista, è di certo assai remoto nel tempo.

Per affrontare i dilemmi della *filogenesi* linguistica potremmo ricorrere alle distanze di cui parla la matematica, e non sempre a distanze metriche. Scelti n parametri linguistici (lessicali, sintattici, fonetici) che il linguista giudica significativi, a una lingua L potremmo associare una *stringa* di n valori logici vero/falso (vero codificato con 1 e falso codificato con 0): ad esempio se i parametri sono 4 e solo i primi due sono veri nella lingua L, la stringa *binaria* associata a L sarebbe 1100 (la *lunghezza* della stringa di 4 bit è 4).

La distanza più semplice e ovvia fra due stringhe binarie è di certo la *distanza di Hamming* che si limita a contare il numero delle posizioni in cui le due stringhe sono diverse; ad esempio $d_H(1100, 1010) = 2$ perché le due stringhe di lunghezza 4 differiscono in due posizioni, la seconda e la terza; è facile verificare che la distanza di Hamming è metrica. A quest'idea hanno fatto ricorso sia linguisti ormai "storici", come Žarko Muljačić[7] o "sulla breccia" come Giuseppe Longobardi e la sua scuola[8]. La sorpresa è che sia le stringhe di Muljačić sia quelle di Longobardi sono ternarie e non binarie: c'è un terzo valore logico diverso da 0 e diverso da 1.

[7] Cfr. Muljačić 1967, pp. 23-37.
[8] Cfr. Ceolin *et al.* 2021, pp. 1-10.

L'impostazione di Longobardi porta a risultati straordinari, ad esempio rafforza l'ipotesi che il giapponese e il coreano siano parenti sia pur lontani, ma è "tecnicamente" più impegnativa per cui non la affronteremo in maniera diretta: la natura del terzo simbolo è delicata nel caso di Longobardi poiché riflette una situazione di *non-informatività*[9] più che una situazione di ambiguità logica del tipo "un po' sì un po' no", "non è chiaro se sì o se no", come quelle che sa ben gestire la logica *fuzzy*[10] e che è pienamente appropriata nel caso di Muljačić.

La natura atipica del terzo parametro porta Longobardi a usare varianti della distanza di Hamming che non sono metriche e violano la disuguaglianza triangolare, lasciando un po' di amaro in bocca ai matematici, che hanno cercato di rimediare ricorrendo a distanze alternative che rispettassero la disuguaglianza triangolare[11].

Tutto ciò rischia di farci affondare in tecnicismi linguistici, matematici e logici, per cui ci accontenteremo di approfondire il caso ormai storico di Muljačić, con ciò conservando tutta l'essenza del problema senza bisogno di precisazioni e di deviazioni pesanti; nel paragrafo 5 accenneremo al caso di Longobardi, di interesse attualissimo negli ambienti della ricerca linguistica di punta. Ritornando a Muljačić, la natura *fuzzy* del terzo valore logico, "un po' sì un po' no", "non è chiaro se sì o se no", è del tutto evidente: ricorreremo alla codifica numerica $\frac{1}{2}$, a metà strada fra 0 e 1.

3. Distanze *fuzzy*

Scelto un valore α fra 0 e $\frac{1}{2}$ (estremi compresi) la distanza che proporremo[12] fra stringhe di *n* simboli ternari (e quindi fra le lingue che da quelle stringhe sono descritte) è di tipo additivo; ha *n* addendi che si sommano, uno per ogni posizione

[9] Per intenderci, anche se siamo costretti a semplificare e a chiedere scusa ai linguisti: se al parametro "In questa lingua esiste la declinazione dei sostantivi" abbiamo risposto di no è inutile chiederci, come previsto da un parametro successivo, se genitivo e dativo possano avere desinenze distinte, come in tedesco, cui spetterebbe l'1, mentre in romeno, cui spetterebbe lo 0, sono sempre coincidenti (i sostantivi tedeschi e romeni si declinano).

[10] Cfr. e.g. Sgarro 2019, pp. 37-45; Sgarro, Franzoi 2021, pp. 50-62.

[11] Cfr. Franzoi, Sgarro 2017, pp. 1168-1177; Franzoi *et al.* 2019, pp. 345-352.

[12] Valori di α superiori a $\frac{1}{2}$ porterebbero a violare l'assioma i'); cfr. l'Appendice.

delle stringhe dalla prima all'ultima:

$$d(x, y) = d(x_1, y_1) + d(x_2, y_2) + \ldots + d(x_{n-1}, y_{n-1}) + d(x_n, y_n)$$

dove x_i e analogamente y_i indicano i due valori logici che occupano la *i*-esima posizione nelle due stringhe *x* e *y*, $1 \leq i \leq n$. Basterà dunque definire la distanza fra due valori logici singoli ("monodimensionali") *u* e *v* (la distanza fra due stringhe di lunghezza $n = 1$). La formula che abbiamo scelto, nuova in letteratura anche se legata a quanto fatto in FRANZOI, SGARRO (2017) e FRANZOI *et al.* (2019)[13], da argomentare e da giustificare, è

$$d_\alpha(u, v) = \alpha \text{ se } u = v = \tfrac{1}{2}, \; = |u - v| \text{ altrimenti} \qquad (1)$$

Le barre verticali indicano il valore assoluto ("ammazzano" il segno -); la quasi-metricità è provata nell'Appendice. Se le due stringhe sono *nitide* (*crisp*, binarie, si compongono di soli 0 e 1) si ritrova la distanza di Hamming, che ha anch'essa natura additiva.

Ad esempio $d(110\,\tfrac{1}{2}, \tfrac{1}{2}\,10\,\tfrac{1}{2}) = d(1, \tfrac{1}{2}) + d(1, 1) + d(0, 0) + d(\tfrac{1}{2}, \tfrac{1}{2}) = \tfrac{1}{2} + 0 + 0 + \alpha = \tfrac{1}{2} + \alpha$.

La logica *fuzzy*, o logica *sfocata, sfumata*, è stata oggetto di contributi apparsi in questa stessa rivista[14]: il valore logico di una proposizione *P* può essere *qualunque* numero compreso fra 0 e 1, ad esempio $\frac{1}{2}$ nei casi di totale ambiguità. Come comporre due valori logici *u* e *v* mediante gli operatori logici della logica *fuzzy* ($0 \leq u, v \leq 1$)? Le risposte sono varie; per la negazione NOT di fatto si usa solo la complementazione a 1, $\neg u$ = NOT $u = 1 - u$, per cui, se $u = ½$ indica l'ambiguità totale, $\neg u = u$. Più delicata è la situazione nei casi degli operatori AND ("e anche", congiunzione) e OR ("oppure non esclusivo", disgiunzione); seguono le tre risposte più comuni:

u OR $v = \max[u, v]$, u AND $v = \min[u, v]$, operatori standard o di Zadeh;

u OR $v = \min[u + v, 1]$, u AND $v = \max[u + v - 1, 0]$, operatori di Łukasiewicz;

u OR $v = u + v - u\,v$, u AND $v = uv$, operatori probabilistici;

(min e max stanno per minimo e massimo dei due numeri in parentesi quadra).

Jan Łukasiewicz (1878-1956) e Lotfi Zadeh (1921-2017) sono considerati, rispettivamente,

[13] Cfr. FRANZOI, SGARRO 2017, pp. 1168-1177; FRANZOI *et al.* 2019, pp. 345-352.
[14] Cfr. SGARRO 2019, pp. 37-45; SGARRO, FRANZOI 2021, pp. 50-62.

i padri della logica moderna "multivalente", a più valori logici, e della logica *fuzzy* nello specifico. Per quanto riguarda l'ultima scelta degli operatori, il lettore che sa di probabilità riconoscerà due celebri regole del calcolo probabilistico per eventi indipendenti. Pensando alla distanza di Hamming è naturale introdurre come distanze fra due valori logici u e v:

$$d(u, v) = (u \text{ AND } \neg v) \text{ OR } (\neg u \text{ AND } v) \qquad (2)$$

(il parametro è presente nella prima lingua e assente nella seconda, oppure è assente nella prima lingua e presente nella seconda); in tal modo generalizziamo la distanza di Hamming a logiche più flessibili di quella rigidamente binaria, no/sì, 0/1.

Con gli operatori scelti, nel caso che a noi interessa, quello ternario dei tre valori logici 0, 1 e $\frac{1}{2}$, i calcoli mostrano che le distanze, da estendere additivamente per stringhe di lunghezza n, diventano proprio ed esattamente α-distanze con α rispettivamente uguale a 0 (Łukasiewicz), $\frac{7}{16}$ (probabilistico) e $\frac{1}{2}$ (Zadeh), cui aggiungeremo come valore di riferimento anche $\alpha = \frac{1}{4}$ (gli operatore logici da cui esso deriva non fanno parte del breve elenco da noi esibito; cfr. l'Appendice, a cui sono relegati argomenti più "tecnici"). Lo ribadiamo: nel caso di due stringhe di n valori logici la loro distanza si ottiene sommando gli n contributi relativi alle singole posizioni, dalla prima all'ultima; l'additività conserva l'eventuale metricità o quasi-metricità delle stringhe monodimensionali, com'è immediato verificare.

Dal punto di vista metrico le distanze *fuzzy* possono essere molto indisciplinate[15], ma nel caso ternario delle α-distanze si tratta sempre e comunque di distanze quasi-metriche (rigorosamente metriche nel caso di Łukasiewicz $\alpha = 0$); si confronti l'Appendice.

Una breve nota finale sulle auto-distanze $d(x, x)$: possono esser viste come un'adeguata e significativa misura di *fuzziness* della stringa x, e in questo senso appare un peccato che agli operatori di Łukasiewicz corrisponda auto-distanza nulla per tutte le stringhe, e non soltanto per le stringhe *crisp*, nitide, composte dei soli valori logici 0 e 1.

[15] Cfr. SGARRO, FRANZOI 2023, pp. 44-54.

4. Risultati linguistici e alberi filogenetici

La lingua dalmatica si è estinta alla fine dell'Ottocento: era una lingua neolatina che a Ragusa-Dubrovnik fu usata per redigere gli statuti del 1272; l'ultimo parlante, Tuone Udaina o Antonio Udina, un minatore, morì a Veglia-Krk nel 1898 dopo essere stato intervistato dal linguista istriano Matteo Bartole, per cui sul dalmatico possediamo anche informazioni fonetiche.

L'idea di Žarko Muljačić era che il dalmatico sia un ponte, ormai scomparso, fra le lingue neolatine occidentali, come l'italiano o il francese, e quelle orientali, essenzialmente il romeno (cui si potrebbero aggiungere le sue varianti, come l'istroromeno parlato nell'Istria interna, ormai pressoché estinto). Le lingue considerate in Muljačić (1967)[16] sono romeno (R), dalmatico (**D**), italiano (I), sardo (Sa), friulano (Frl), romancio (Rsh), provenzale (Pr), francoprovenzale (FP), francese (F), catalano (C), spagnolo (S) e portoghese (P), mentre i parametri, parte lessicali e parte sintattici, sono in tutto 40. A partire dalle distanze fra le lingue calcolate con la (1) possiamo costruire degli alberi o *dendrogrammi* (cfr. Figure 1, 2, 3 e 4), ad esempio tramite l'algoritmo di classificazione UPGMA (*unweighted pair group method with arithmetic mean*) disponibile in rete. Gli α che abbiamo privilegiato danno quattro alberi sostanzialmente identici che convalidano tutti la tesi di Muljačić.

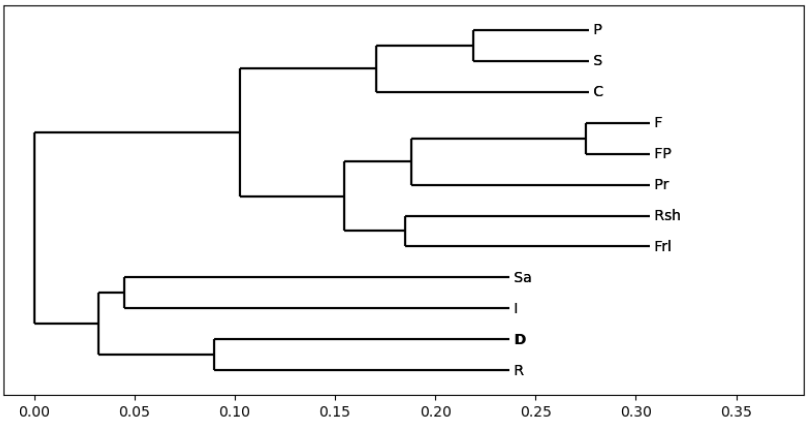

Figura 1. Albero con α=0

[16] Cfr. Muljačić 1967, pp. 23-37.

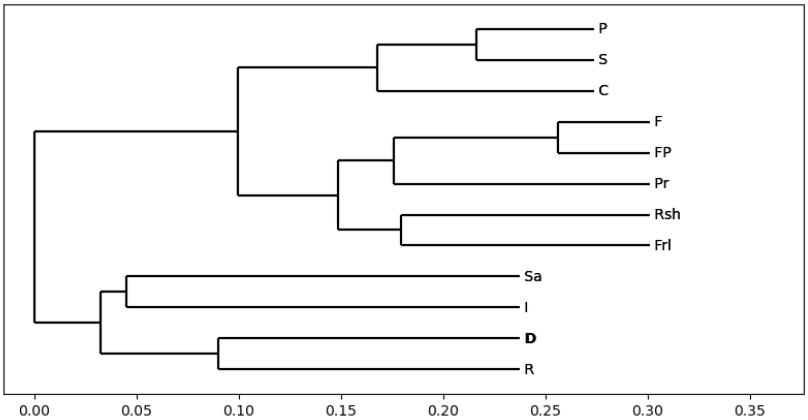

Figura 2. Albero con $\alpha=\frac{1}{2}$

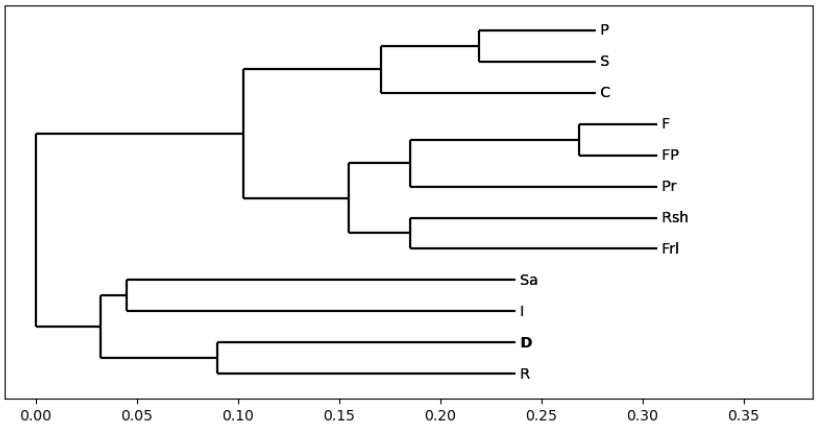

Figura 3. Albero con $\alpha=\frac{1}{4}$

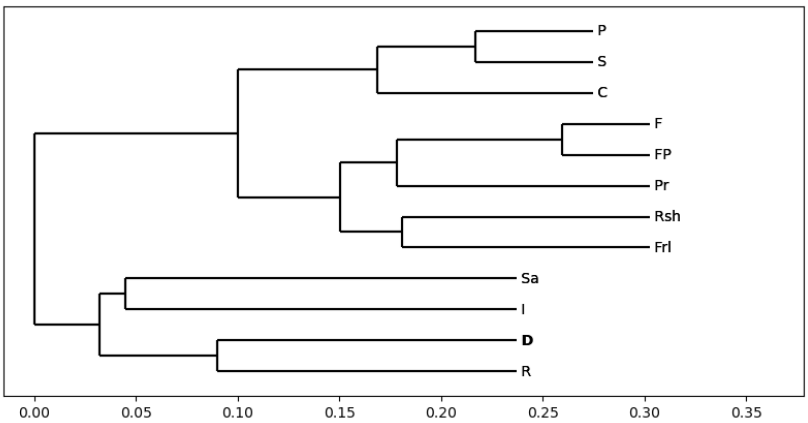

Figura 4. Albero con $\alpha=\frac{7}{16}$

5. Uno sguardo sulla ricerca oggi

I dati ternari di Longobardi sono relativi a 58 lingue del Vecchio Mondo, Europa, Asia e Africa, e a 98 parametri esclusivamente sintattici: l'idea, che si rivela assai fruttuosa, è

che la *sintassi* evolve più lentamente del *lessico*, come di uso tradizionale nella filogenesi linguistica, e può dunque darci informazioni profonde sulla "preistoria" delle lingue[17]. Mentre gli operatori logici di Łukasiewicz e di Zadeh sono già stati usati in letteratura[18], sia pur senza usare il formalismo unificante delle α-distanze, quelli corrispondenti a $\alpha = \frac{1}{4}$ e $\frac{7}{16}$ sono affatto nuovi. La natura problematica del terzo simbolo richiede tuttavia l'uso di uno strumento estremamente tecnico, la *trasformata di Steinhaus*[19]. Gli esperimenti preliminari con le α-distanze e con i dati di Longobardi si avvalgono della trasformata di Steinhaus e sono del tutto soddisfacenti. Il valore che appare il migliore, anche raffrontato ai risultati originali dell'articolo[20], è proprio $\alpha = \frac{1}{4}$; migliore nel senso che l'albero ottenuto è il più vicino al dendrogramma dell'articolo[21] che i linguisti considerano un "*benchmark*" (ossia un parametro di riferimento). Su questi argomenti la ricerca è in corso.

6. Appendice: Divagazioni logico-matematiche

Gli operatori logici della logica *fuzzy* come quelli che abbiamo citato devono soddisfare una serie di assiomi, per i quali si veda ad esempio la voce di Wikipedia *T-norms*. Basti qui osservare che gli assiomi vincolano in maniera univoca i risultati se almeno uno dei due valori logici coinvolti è nitido, 0 o 1; di conseguenza le distanze *fuzzy*[22], a prescindere dagli operatori logici scelti, danno lo stesso identico risultato se le due stringhe coinvolte sono prive di posizioni strettamente *fuzzy* dove non compaia né lo 0 né l'1. Data l'additività delle distanze fra stringhe, basterà occuparsi della distanza fra due singoli valori logici u e v (fra due numeri u e v compresi fra 0 e 1, $0 \leq u, v \leq 1$, o se si vuole fra due stringhe di lunghezza $n = 1$); i calcoli coinvolti non sono impegnativi.

Muljačić: $d(u, v) = \max\,[\min(u, 1 - v), \min(1 - u, v)]$, $d(\frac{1}{2}, \frac{1}{2}) = \frac{1}{2}$

[17] Cfr. Ceolin *et al.* 2021, pp. 1-10.
[18] Cfr. Franzoi, Sgarro 2017, pp. 1168-1177; Franzoi *et al.* 2019, pp. 345-352.
[19] Per dare un'idea, la *trasformata di Steinhaus* applicata alla distanza di Hamming produce la distanza di Jaccard che "cancella" le posizioni zero-zero; nei contributi di nota 18 la trasformata viene invece usata per "alleggerire" il peso delle posizioni non-informative. Tale esigenza è affrontata in maniera diversa da Longobardi, ad esempio nel contributo di nota 17, a sacrificio tuttavia della metricità, mentre la trasformata di Steinhaus se applicata a distanze (quasi) metriche produce distanze ancora (quasi) metriche. Tutto ciò è molto tecnico; la lettrice o il lettore interessati sono indirizzati a Franzoi, Sgarro 2017, pp. 1168-1177 e a Franzoi *et al.* 2019, pp. 345-352.
[20] Cfr. Ceolin *et al.* 2021, pp. 1-10.
[21] Cfr. Gray, Atkinson 2003, pp. 435-439.
[22] Cfr. Sgarro, Franzoi 2023, pp. 44-54.

Łukasiewicz: $d(u, v) = |u - v|$, $d(\frac{1}{2}, \frac{1}{2}) = 0$

(Nel calcolo a partire dalla (2) converrà distinguere i due casi $u \leq v$ e $u \geq v$.)

caso probabilistico: $d(u, v) = u + v - 3uv + u^2v + uv^2 - u^2v^2$, $d(\frac{1}{2}, \frac{1}{2}) = \frac{7}{16}$

caso misto: $d(u, v) = \max(u, v) - uv$, $d(\frac{1}{2}, \frac{1}{2}) = \frac{1}{4}$

Nel caso "misto" abbiamo scelto la congiunzione prodotto (probabilistica) e la disgiunzione standard: le ragioni per preferirlo al caso probabilistico "puro" verranno discusse fra poco.

Queste distanze, come si potrebbe dimostrare[23], verificano tutte la disuguaglianza triangolare; ovviamente quella di Łukasiewicz altro non è se non la celebre metrica di Minkowski o *taxi-distance*. La triangolarità, tuttavia, non è sempre garantita per le distanze[24]: ci appoggeremo a una situazione "estrema", quella degli operatori *drastici* o di Weber; se u e v sono entrambi strettamente *fuzzy* (né 0 né 1), se dunque la coppia (u, v) è interna al quadrato unitario:

u OR $v = 1$, u AND $v = 0$, operatori drastici o di Weber

Come si verifica facilmente, all'interno del quadrato unitario la distanza fra due valori logici è sempre 0. Se prendiamo i tre valori logici $u = 1$, $v = \frac{1}{4}$, $w = \frac{3}{4}$ si ha $d(u, w) + d(w, v) = \frac{1}{4} + 0 < d(u, v) = \frac{3}{4}$ in palese violazione della disuguaglianza triangolare; di passaggio si noti che l'α di Weber e quello di Łukasiewicz sono entrambi uguali a 0. Altrettanto problematica è la distanza probabilistica "pura": vale sì la disuguaglianza triangolare, ma purtroppo l'auto-distanza di $\frac{1}{3}$ supera l'etero-distanza fra 0 e $\frac{1}{3}$ ($\frac{32}{81} > \frac{1}{3}$), il che è inaccettabile; viceversa, com'è agevole controllare, la distanza "mista" è quasi-metrica.

Tutto ciò è notevole dal punto di vista matematico perché dimostra che, al di fuori del caso delle nostre stringhe ternarie con le relative α-distanze, il modello di distanza (2) è troppo generale, per cui i controlli metrici vanno fatti di volta in volta su operatori logici specifici.

Veniamo alla quasi metricità delle α-distanze per stringhe ternarie sull'*alfabeto* $\{0, \frac{1}{2}, 1\}$. Stante l'additività delle distanze, basterà occuparsi della disuguaglianza triangolare per tre valori logici arbitrari u, v e w, escludendo le situazioni in cui non compare l'autodistanza α, visto che, pensando al bordo del quadrato unitario, la metricità della *taxi-distance* è ben nota. Le posizioni miste nitido-*fuzzy* contribuiscono per $\frac{1}{2}$, quelle *fuzzy-fuzzy* per α. Se il membro di destra $d(u, v)$ vale α, e dunque

[23] Cfr. MAREGA 2024.
[24] Cfr. SGARRO, FRANZOI 2023, pp. 44-54.

$u = v = \frac{1}{2}$, la somma dei due termini di sinistra vale almeno $2\alpha \geq \alpha$; se il membro di destra vale $\frac{1}{2}$ la somma di sinistra vale almeno $\frac{1}{2}$ (senza restrizione effettiva possiamo supporre $u = 1$ e $v = \frac{1}{2}$ e poi controllare i tre casi possibili per z), se il membro di destra vale 0 oppure 1 il controllo è superfluo (non ci sono coppie *fuzzy-fuzzy*). Qualunque sia α si ha $d(0, \frac{1}{2}) = d(1, \frac{1}{2}) = \frac{1}{2} \geq \alpha = d(\frac{1}{2}, \frac{1}{2})$, in accordo con l'assioma i′).

BIBLIOGRAFIA

CRISTIANINI N.
2024, *Machina sapiens*, Bologna, il Mulino.

CEOLIN A., GUARDIANO C., LONGOBARDI G., IRIMIA M. A., BORTOLUSSI L., SGARRO A.
2021, «At the boundaries of syntactic prehistory», *Philosophical Transactions of the Royal Society B,* 376, pp. 1-10.

DEZA M. M., DEZA E.
2016, *Encyclopedia of distances*, Berlin, Heidelberg, Springer.

FRANZOI L., SGARRO A.
2017*, Linguistic classification: T-norms, fuzzy distances and fuzzy distinguishabilities*, Marseille, KES, pp. 1168-1177.

FRANZOI L., SGARRO A., DINU A., DINU L. P.
2019, *Linguistic classification: dealing jointly with irrelevance and inconsistency*, Varna, RANLP, pp. 345-352.

GRAY R. D., ATKINSON Q. D.
2003, «Language tree divergence times support the Anatolian theory of Indo-European origin», *Nature,* 426, pp. 435–439.

MAREGA P.
2024, *Distanze metriche e non metriche in Linguistica*, Tesi di Laurea, Università di Trieste - Dipartimento MIGe.

MULJAČIĆ Z.
1967, «Die Klassifikation der romanischen Sprachen», *Romanistisches Jahrbuch*, 18, pp. 23-37.

SGARRO A.
2019, «Come contraddirsi rimanendo coerenti: il caso della logica fuzzy», *Quaderni CIRD*, 19, pp. 37-45.

SGARRO A., FRANZOI L.
2021, «Sillogismi sfocati: il modus ponens», *QuaderniCIRD*, 22, pp. 50-62.
2023, «Distanze, lingue e parole», *QuaderniCIRD*, 27 (2023), pp. 44-54.