Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 00 (2016) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

21th International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France

Linguistic classification: T-norms, fuzzy distances and fuzzy distinguishabilities

Laura Franzoi[a,c,*], Andrea Sgarro[b,c]

[a] *Faculty of Mathematics and Computer Science, University of Bucharest, Romania*
[b] *Department of Mathematics and Geosciences, University of Trieste, Italy*
[c] *Human Language Technologies Research Center, University of Bucharest, Romania*

Abstract

Back in 1967 the linguist Ž. Muljačić used an additive distance between ill-defined linguistic features which is a forerunner of the fuzzy Hamming distance between strings of truth values in standard fuzzy logic. Here we show that if the logical frame is changed one obtains additive distances which are either sorely inadequate, as in the Łukasiewicz or probabilistic case, or coincide with the distance originally envisaged by Muljačić, as happens with a whole class of T-norms (abstract logical conjunctions) which includes the nilpotent minimum. All this strengthens the role of Muljačić distances in linguistic clustering and of Muljačić distinguishabilities (a notion subtly different from distances, but quite inalienable) in linguistic evolution. As a preliminary example we re-take and re-examine Muljačić original data.

Keywords: string distance, fuzzy distance, string distinguishability, T-norm, linguistic classification, linguistic evolution

1. Introduction

Back in 1967[14] the Croat linguist Ž. Muljačić introduced what appears to us as a natural *fuzzy* generalization of crisp Hamming distances between binary strings of fixed length n, called henceforth *Muljačić distance*; he wanted to show that Dalmatic, now an extinct language, is a bridge between the Western group of Romance languages and the Eastern group, mainly Romanian and its variants. The situation is the following: Romance languages $L, \Lambda, \ldots$ are each described by means of n features, which can be present or absent, and so are encoded by a string $s(L) = \underline{x} = x_i \ldots x_n$, where x_i is the truth value of the proposition *feature i is present in language L*; however, presence/absence can be ill-defined and so each $x = x_i$ is rather a truth value $x \in [0, 1]$ in a multi-valued logic; $x = x_i$ is a *crisp* value only when $x = 0 = false = absent$, or $x = 1 = true = present$, else is a *strictly fuzzy* value. In the sequel we set $x \wedge y \doteq \min[x, y]$, $x \vee y \doteq \max[x, y]$ and $\overline{x} \doteq 1 - x$; these are the truth values of conjunction AND, disjunction OR and negation NOT w.r. to propositions with truth values x and y in *standard fuzzy logic*, a popular form of multi-valued

* Corresponding author. Tel.: +39 348 7054366.
E-mail address: laura.franzoi@gmail.com

1877-0509

logic, while e.g. Łukasiewicz logic, cf. below, has the same negation but different conjunctions and disjunctions. What happens if one changes the logical framework, and *minima* and *maxima* are replaced by another T-norm (conjunction) and the corresponding T-conorm (disjunction)? Section 2 shows that the alternatives are either sorely inadequate, as in the Łukasiewicz and probabilistic case, or give back the very same distance, as in the nilpotent-minimum case: thus, our results enhance the relevance of Muljačić distances and Muljačić distinguishabilities in computational linguistics (the coding-theoretic notion of *distinguishability*, as opposed to distance, was introduced by Cl. Shannon back in the fifties [17], cf. Section 2).

While Section 2 is rather technical, Section 3, which is based on Muljačić' original data of 1967, is a description of the use one might make of our tools in computational linguistics, a task we are beginning to tackle in ongoing research [3] on extensive data within the activities of the Human Language Technologies Research Center, Bucharest University, cf. also [10]. To the best knowledge of these authors, the implications of Muljačić research as far as fuzziness is concerned have been overlooked in literature, possibly because in his years fuzziness was still an emerging and not yet well understood notion, cf. [9]. We are confident that Muljačić distances, coupled with Muljačić distinguishabilities, may prove to be a useful tool not only, in the wake of Muljačić, in linguistic classification and linguistic evolution, cf. Section 3, but also in other domains, coding theory as already dealt with in [9], or even bioinformatics, cf. the concluding section.

2. Methods: fuzzy distances and T-norms

For the moment being we stick to the logical operators which are standard in fuzzy logic, *maxima* and *minima*. We move to *additive* distances between logical strings and in particular between two truth values x and $y \in [0,1]$, i.e. between strings of length 1, x and y being the truth-values of propositions P_x and P_y, respectively. Mimicking usual Hamming distances, we set $d(x,y) \doteq \big[(P_x \text{ true}) \text{ AND } (P_y \text{ false})\big] \text{ OR } \big[(P_x \text{ false}) \text{ AND } (P_y \text{ true})\big]$, i.e. $d(x,y) = \big[x \wedge (1-y)\big] \vee \big[(1-x) \wedge y\big]$.
Two truth values x and y are *consonant* if $x \vee y \le \frac{1}{2}$ or $x \wedge y \ge \frac{1}{2}$, else they are *dissonant*; the *fuzziness* of the truth value x, is $f(x) \doteq x \text{ AND } \overline{x} = x \wedge (1-x)$. It has been proved [9] that $d(x,y)$ can be equivalently expressed as $d(x,y) = f(x) \vee f(y)$ or $1 - f(x) \vee f(y)$, according whether x and y are consonant or dissonant, or also [9,15], as $d(x,y) = |x-y| + f(x) \wedge f(y)$. Additively, the first equivalence gives an expression for string distances which stresses its relation to the crisp Hamming situation ($\mathcal{D}$ and C denote dissonant and consonant positions, respectively):

$$d(\underline{x},\underline{y}) \doteq \sum_{i \in \mathcal{D}} \Big[1 - [f(x_i) \vee f(y_i)]\Big] + \sum_{i \in C} \Big[f(x_i) \vee f(y_i)\Big] \tag{1}$$

The second equivalence makes clear the relation with taxicab distances, or Minkowski distances. We stress that (1) is a distance between strings of truth values, rather than a distance between the corresponding languages, cf. Section 3.

Back in 1956 Claude Shannon introduced [11,17] into information and coding theory the notion of *distinguishability*, later generalized to broader contexts as ours in [1,2,3]:

$$\delta(x,y) \doteq \min_z [d(x,z) \vee d(y,z)] \tag{2}$$

In [9] Muljačić distinguishability or fuzzy Hamming distinguishability has been computed to be:

$$\delta(\underline{x},\underline{y}) \doteq \sum_{i \in \mathcal{D}} \frac{1}{2} + \Big[\sum_{i \in C} f(x_i) \vee \sum_{i \in C} f(y_i)\Big] \tag{3}$$

In a general continuous context, *minima* in (2) should be replaced by *infima*, at least *a priori*; however, in our case minima are always attained, even if z is constrained to belong to the ternary alphabet $\{0, \frac{1}{2}, 1\}$ [9]. Both Muljačić distances (called Sgarro distances in [6]) and Muljačić distinguishabilities are *fuzzy metrics*, cf. the Appendix; self-distances $d(\underline{x},\underline{x})$ and self-distinguishabilities $\delta(\underline{x},\underline{x})$ are strictly *positive*, unless $\underline{x}$ is crisp, i.e. all its components x_i are crisp. We stress that, unlike the corresponding distances, Muljačić distinguishabilities do *not* have an *additive* nature, due to consonant positions. For metric distances, be they crisp or fuzzy, one soon proves the basic bounds, cf. e.g. [1,2,3,9]:

$$\frac{d(x,y)}{2} \leq \delta(x,y) \leq d(x,y) \tag{4}$$

Often distinguishabilities are equal to the lower bound, in case corrected to its integer ceiling $\left\lceil \frac{d(x,y)}{2} \right\rceil$ when distances, and so distinguishabilities, cf. (2), are constrained to be integers: this e.g. happens with crisp Hamming distances, edit distances or Kendall permutation distances, by this *trivializing* the notion of distinguishability. However, in our case, cf. Section 2 and Appendix B, Muljačić' distinguishabilities span the interval in (4), and equality of distances does *not* imply equality of distinguishabilities, cf. Section 3 and Appendix B. As we argue below in Section 3, distinguishabilities are essential when one moves from static linguistic classification and clustering to dynamic linguistic evolution.

While in a crisp setting it is quite clear what disjunctions and conjunctions should be, this is not so in a *fuzzy* context as ours is: one has a whole range of abstract conjunctions AND and disjunctions OR, called T-norms and T-conorms respectively (cf. Appendix A for definitions and basic properties), which compete with the standard choice of *minima* and *maxima*, and which give rise to alternative forms of multi-valued logics. Remarkable cases are Łukasiewicz, probabilistic and nilpotent-minimum as covered below (cf. again Appendix A for definitions); actually, in the paper additional options will be taken into account.

Given additivity (1), in this section we take $n = 1$, $x, y \in [0, 1]$; the generalization to n-length sequences is straightforward. Since all T-norms take the same values on the border $x \wedge y = 0$ or $x \vee y = 1$, and so do all the T-conorms, we shall operate mainly on the *open unit square*. Recall that for any T-norm $x \top y$ or T-conorm $x \perp y$, one has $x \top y \leq x \wedge y$, $x \perp y \geq x \vee y$, cf. Appendix A. We generalize *nilpotent mimima* to parametric β-nilpotent minima, where the inequality in Appendix A is replaced by

$$x + y < \beta\,, \quad 0 < \beta \leq 2$$

We add also the T-norm introduced in [8] for technical reasons of fuzzy arithmetic, but actually quite meaningful as we argue below, that is the *α-minima* T-norms and the corresponding dual T-conorms:

$$x \top y = 0 \ \text{ for } x \vee y < \alpha, \text{ else } \ x \top y = x \wedge y$$
$$x \perp y = 1 \ \text{ for } x \wedge y > 1 - \alpha, \text{ else } \ x \perp y = x \vee y; \quad 0 < \alpha \leq 1$$

Observe that the drastic T-norm as in Appendix A is re-obtained as a limit case both of parametrized β-nilpotent minima when $\beta = 2$, and of α-minima when $\alpha = 1$, while non-interactivity is re-found for $\alpha = \beta = 0$. Actually, we find it convenient to take more general $\mathcal{R}$-minima where $\mathcal{R}$ is a region of the closed unit square which verifies:

(i) $x, y \in \mathcal{R}$ implies $x \vee y \neq 1$
(ii) $x, y \in \mathcal{R}$ implies $y, x \in \mathcal{R}$
(iii) $x, y \in \mathcal{R}$, $u \leq x, v \leq y$ implies $u, v \in \mathcal{R}$

and define the $\mathcal{R}$-minimum T-norm by:

$$x \top y = 0 \ \text{ for } x, y \in \mathcal{R}, \ \text{else } x \top y = x \wedge y$$

The fact that any $\mathcal{R}$-*minimum* is actually a T-norm is soon checked: e.g. for the associative property take, say, $x, y \in \mathcal{R}$: one has $(x \top y) \top z = 0 \top z = 0$, and $x \top (y \top z) = x \top u = 0$ since either $u = 0$ or $u = y \wedge z \leq y$, and so $x, u \in \mathcal{R}$, use (iii) . Note that on the 1-complement or *negated* region $\mathcal{R}^N$, defined by $x, y \in \mathcal{R}$ iff (if and only if) $\overline{x}, \overline{y} \in \mathcal{R}^N$, the corresponding conorm is 1, else is $x \vee y$. The underlying idea is that when the truth values x and y of the propositions P_x and P_y are "too small", i.e. belong to $\mathcal{R}$, then the possibility (the truth value) of the disjunction $P_x \vee P_y$ is nil in practice, and so is set straightaway equal to 0. Our interpretation makes "large values" of the parameters α and β rather odd, as will be confirmed below when $\alpha > \frac{1}{2}$ or $\beta > 1$. Note that the nilpotent minimum is hybrid, having a Łukasiewicz disjunction $x + y$ which defines $\mathcal{R}$ versus a non-interactive conjunction outside $\mathcal{R}$, this being a reason why we prefer the option put forward in[8]; a Łukasiewicz conjunction outside $\{x, y : x + y < 1\}$ would give $x \top y = 1$, and correspondingly one would have to generalize nil-potent minima to an operator $x \tau y \in \{0, 1\}$ which is not even associative: $(x\tau y)\tau z = 0$, while $x\tau(y\tau z) = x$ as soon as $x + y < 1$, $y + z \geq 1$.

The region $\mathcal{R}$ is called *proper* when it has void intersection with $\mathcal{R}^N$. Equivalently, $\mathcal{R}$ is proper when $x, y \in \mathcal{R}$ implies $x + y < 1$: else by (iii) there would be an intersection $x, 1 - x$ with the segment $x + y = 1$, and by (ii) also $1 - x, x$ would belong to $\mathcal{R}$ which would then have a non-void intersection with its negation $\mathcal{R}^N$. Proper regions $\mathcal{R}$ are those of α-minima with $\alpha \leq \frac{1}{2}$ and β-nilpotent minima with $\beta \leq 1$, including non-interactivity. The largest *proper* region $\mathcal{R}$ corresponds to standard nilpotent minima ($\beta = 1$), the smallest to non-interactivity ($\alpha = \beta = 0$).

Given a T-norm $x \top y$ and the corresponding *dual* T-conorm $x \perp y$, the *fuzziness* of the truth value $x \in [0, 1]$ is defined as $f(x) \doteq x \top (1 - x) = x \top \overline{x}$. Of course $f(x) = f(1 - x)$, $f(x) = 0$, while $f(\frac{1}{2}) \leq \frac{1}{2} \wedge (1 - \frac{1}{2}) = \frac{1}{2}$. The fuzziness is called *proper* when:

⋆ $f(x)$ increases (possibly weakly) from $f(0) = 0$ to $f(\frac{1}{2}) > 0$

These authors deem that *a proper fuzziness*, i.e. the possibility to give a positive truth value to the conjunction of a proposition and its negation, *is an inalienable property of multi-valued logical systems which aim at modeling fuzziness in a significant way*. In the example below of proper and improper fuzziness we assume $x \leq \frac{1}{2}$.

Proper cases:
minimum or non-interactive norm, standard nilpotent minimum, and more generally α-minima with $\alpha \leq \frac{1}{2}$, β-nilpotent minima with $\beta \leq 1$: $f(x) = x$
product or probabilistic: $f(x) = x - x^2$, $f(\frac{1}{2}) = \frac{1}{4}$

Improper cases:
Łukasiewicz, drastic, β-nilpotent minimum with $\beta > 1$: $f(x) = 0$
α-minimum with $\alpha \geq \frac{1}{2}$: $f(x) = x$ for $x < 1 - \alpha$ else $f(x) = 0$

We move to distances between truth values for a given T-norm, which one can soon extend to additive distances $d(\underline{x}, \underline{y})$ for strings $\underline{x}$ and $\underline{y}$ of length n:

$$d(x, y) = \Big[x \top (1 - y)\Big] \perp \Big[(1 - x) \top y\Big]$$

Whatever the T-norm the following properties are obviously verified:

[i] $0 \leq d(x, y) \leq 1$
[ii] $d(x, y) = d(y, x)$, $d(\overline{x}, \overline{y}) = d(x, y)$
[iii] on the border $d(x, y) = |x - y|$

So there is symmetry and invariance w.r. to negation of both arguments; [iii] explains why below the border is mainly ignored. We stress the following obvious properties for *self-distances* $d(x, x)$ (recall that $x \perp y \geq x \vee y$):

Lemma 1. $d(x, x) \geq f(x)$, $d(x, x) = 0$ *iff* $f(x) = 0$

E.g. in the probabilistic case $d(x, x) = f(x)\,[2 - f(x)] > f(x)$ unless $x = 0$. Let us move to distances. For the proper-fuzziness case of Muljačić or fuzzy Hamming distances one has the fuzzy *metric* distance:

$$d_M(x, y) = 1 - f(x) \vee f(y) \quad \text{or} \quad f(x) \vee f(y)$$

according whether x, y are *dissonant* or *consonant*. Does one have other such well-behaved examples?

Łukasiewicz: A straightforward computation gives back the taxicab or Minkowski distance $d(x, y) = |x - y|$, "too crisp" a distance in a fuzzy context; recall that the Łukasiewicz fuzziness is improper, actually fuzziness is "ignored", $f(x) = 0$.

Product: The fuzziness $f(x) = x - x^2$ is proper, but, unfortunately, the fuzzy metric property $d(x, x) \leq d(x, y)$ falls. E.g. $d(\frac{1}{3}, \frac{1}{3}) = \frac{2}{3^2} \perp \frac{2}{3^2} = \frac{32}{3^4}$, while $d(\frac{1}{3}, \frac{1}{6}) = \frac{29}{3^4} < d(x, x)$. Observe that the metric property here violated is needed to have the basic inequality for distinguishabilities $\delta(x, y) \leq d(x, y)$, cf. (4).

Lemma 2. *If $d(x, z) = 0$ for any two distinct point x and z in the* open *unit square then the triangle inequality falls.*

Proof. Just take y on the border, the inequality would write $d(x, z) + |y - z| = 0 + |y - z| \geq |y - x|$ whatever y, which is absurd, just take y on the border strictly nearer to z than to x in the Euclidean or taxicab geometry (the two coincide for $n = 1$).

Lemma 3. *If $\mathcal{R}$ is not proper, then $d(x, y) = 0$ whenever $x, \overline{y} \in \mathcal{R} \cap \mathcal{R}^N$.*

To check the lemma, recall the definition of $\mathcal{R}^N$ and just observe that $x, \overline{y} \in \mathcal{R} \cap \mathcal{R}^N$ iff $\overline{x}, y \in \mathcal{R}^N \cap \mathcal{R} = \mathcal{R} \cap \mathcal{R}^N$. However straightforward, this lemma is enough to show the inadequateness of α-minima with $\alpha > \frac{1}{2}$ and β-nilpotent minima with $\beta > 1$:

Drastic, α-minima with $\alpha > \frac{1}{2}$, β-nilpotent minima with $\beta > 1$: the intersection $\mathcal{R} \cap \mathcal{R}^N$ has size > 2, e.g. in the drastic limit case one has $d(x, y) = 0$ over the whole of the *open* unit square; so the triangular property falls, e.g. in the drastic case with $y = 0$, $0 < z < x < 1$ one has $d(x, z) + d(y, z) = |y - z| = z < d(x, y) = |x - y| = x$.

In the special case of α-minima with $\alpha > \frac{1}{2}$ the intersection of $\mathcal{R}$ and $\mathcal{R}^N$ is the open square of sides $]1 - \alpha, \alpha[$, and so the condition $x, \overline{y} \in \mathcal{R} \cap \mathcal{R}^N$ in Lemma 2 is equivalent to $x, y \in \mathcal{R} \cap \mathcal{R}^N$, without the negation; this is not so with β-nilpotent minima, $1 \leq \beta < 2$, as soon checked.

The *proper* region $\mathcal{R}$ is as large as possible with the standard nilpotent minimum, $\beta = 1$, while in the non-interactive case it is as small as possible, actually void in the *open* unit square. From now on we stick to *proper* $\mathcal{R}$-minimum norms; we show in the theorem and in the corollary that fuzziness and self distance w.r. to any proper $\mathcal{R}$, $f_{\mathcal{R}}(x)$ and $d_{\mathcal{R}}(x, x)$, or w.r. to non-interactivity as in the Muljačić case, $f_M(x)$ and $d_M(x, x)$, all coincide:

Theorem 1. *Whichever $\mathcal{R}$-mimimum T-norm with $\mathcal{R}$ proper gives back Muljačić distance: $d_{\mathcal{R}}(x, y) = d_M(x, y)$.*

Proof. Given the invariance of T-norm distances w.r. to complementation of the arguments, we deal only with the case $x + y \leq 1$, and, given symmetry, we assume also $x \leq y$, which implies $x \leq \frac{1}{2} \wedge y$. One has $x + y \leq 1$ i.e. $y \leq 1 - x$, and so one soon checks that $x \wedge \overline{y} \leq \overline{x} \wedge y$. We move to our T-norms; notice that the couple $1 - x, y$ can*not* belong to $\mathcal{R}$, because this would imply $1 - x + y < 1$, i.e. $x > y$. This proves that $d_M(x, y) = \overline{x} \wedge y$; observe that $d_{\mathcal{R}}(x, y) = \overline{x} \wedge y = d_M(x, y)$ whenever $x, \overline{y} \in \mathcal{R}$, $x \top \overline{y} = 0$. If instead $x, 1 - y$ lies outside $\mathcal{R}$ one might have either $d_{\mathcal{R}}(x, y) = (x \wedge \overline{y}) \perp (\overline{x} \wedge y) = (x \wedge \overline{y}) \vee (\overline{x} \wedge y) = \overline{x} \wedge y = d_M(x, y)$ or $d(x, y) = (x \wedge \overline{y}) \perp (\overline{x} \wedge y) = 1$; in the latter case, however, the couple $x \wedge \overline{y}$, $\overline{x} \wedge y$ should belong to $\mathcal{R}^N$, i.e., using De Morgan rules, one should have $(\overline{x} \vee y, x \vee \overline{y}) \in \mathcal{R}$. Using (iii) above to diminish the coordinates, this implies that y, x and $\overline{x}, \overline{y}$ both belong to $\mathcal{R}$, but then, using symmetry, also x, y and $\overline{x}, \overline{y}$ would both belong to $\mathcal{R}$, which would then intersect $\mathcal{R}^N$, which is absurd.

Corollary 1. *If $\mathcal{R}$ is proper $f_{\mathcal{R}}(x) = d_{\mathcal{R}}(x, x) = f_M(x) = d_M(x, x) = x \wedge (1 - x)$*

Proof. The couple $x, 1 - x$ cannot belong to $\mathcal{R}$, because then also $1 - x, x$ would belong to it, and $x, 1 - x$ would belong to both $\mathcal{R}$ and $\mathcal{R}^N$: thus $f_{\mathcal{R}}(x) \doteq x \top (1 - x) = x \wedge (1 - x)$. Now use the theorem.

Comment: Take $n \geq 1$; additive distances as found above are either inappropriate or coincide with Muljačić distance $d_M(\underline{x}, \underline{y})$ as in (1), in which case the corresponding distinguishabilities coincide with Muljačić distinguishability $\delta_M(\underline{x}, \underline{y})$ as in (3). We deem that the present paper enhances the appropriateness of choosing the logical operators of standard fuzzy logic, *minimum* and *maximum*, i.e. of choosing the non-interactive norms. In particular, the possibly most popular norms apart from interactivity, i.e. Łukasiewicz and probabilistic, give an improper fuzziness or violate basic metric requirements. In next section our tools are applied, as a preliminary example, to Muljačić' original data; they are being used in ongoing research on computational dialectology[4], cf. also[10].

In clustering and distance-based methods one would prefer a *crisp* metric distance for languages. If to each language L we associate its string $s(L)$ of n fuzzy truth values $\big(s(L) = s(\Lambda)$ does *not* imply $L = \Lambda\big)$, let us set $d^*(L,\Lambda) = d\big(s(L), s(\Lambda)\big)$ for distinct languages $L \neq \Lambda$, else $d^*(L,L) = 0$. One has a crisp metric for equivalence classes each made up of languages at distance 0 from each other. We deem that it is precisely here that the advantages of having a proper fuzziness show up: if one takes usual distances for strings in the unit n-cube, the Euclidean distance, say, or the taxicab distance as derived from Łukasiewicz norms (both give back crisp Hamming distances in the special case when the arguments are crisp), two languages L and Λ are declared equivalent whenever they have the same string, be it crisp or fuzzy. Instead, with a proper fuzziness as in the Muljačić case, equivalence requires that the common string be also *crisp*, since Muljačić self-distance is zero only when its argument is crisp. As for language distinguishabilities $\delta^*(L,\Lambda)$ and the corresponding equivalence classes, dealt with in the same way, they coincide with those for the associated strings, as soon checked, unless the two associated strings are equal *and* crisp, when the distinguishability is "forced" to be zero. *Languages which are ill-defined are* never *put into the same equivalence class*, even when the two associated strings are equal.

3. Experimental data: back to Muljačić and a re-start

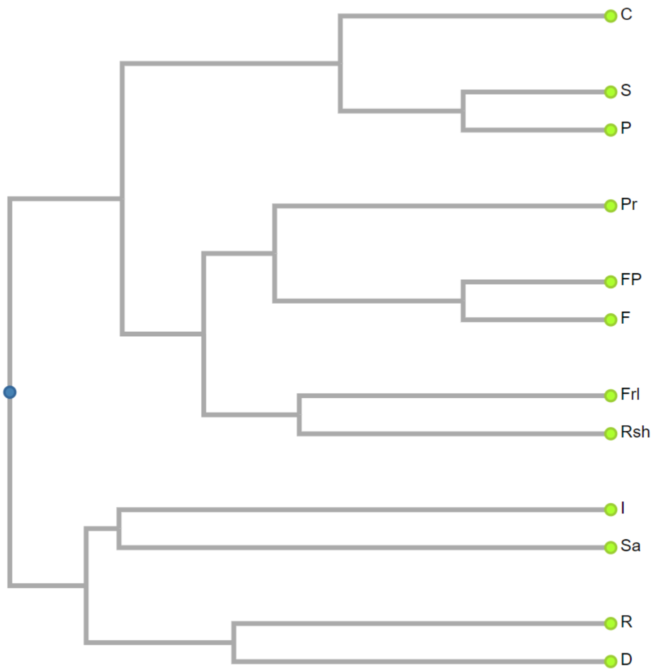
As a preliminary example, we go back to Muljačić original data[14]; he uses 12 ternary strings of length 40, and makes a very sparse use of fuzziness: the only strictly fuzzy value is .5, below replaced by a star $*$ to help readability. The 12 languages are, respectively: R=Romanian, D=Dalmatic, I=Italian, Sa=Sardinian, Frl=Frioulan, spoken in North-East Italy, Rsh=Romansh, the 4th language of Switzerland, Pr=Provençal, FP=Francoprovençal spoken across Italian and French Alps, F=French, C=Catalan, S=Spanish (Castilian) and P=Portuguese.

R	1	0	1	1	0	0	0	0	0	0	1	1	1	0	1	0	1	1	1	0	0	1	1	0	1	0	0	1	0	1	1	0	0	0	1	1	0	1	1	1
D	1	1	1	1	0	0	0	0	1	0	0	0	1	0	0	0	1	1	1	0	1	1	1	0	1	0	1	0	1	1	0	0	*	0	1	1	1	1	1	1
I	1	1	1	0	0	0	1	0	1	0	0	0	1	*	0	0	1	1	1	1	1	0	1	0	1	0	1	1	1	1	1	1	1	1	0	1	0	0	1	0
Sa	1	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	0	1	1	*	1	0	1	1	0	0	0	1	0	1	1	0	0	1	0	1	1	1	1	1
Frl	1	1	1	1	1	0	0	1	1	0	0	0	1	0	0	0	1	1	1	1	*	0	1	0	1	1	1	0	0	0	1	1	0	0	1	0	1	1	0	0
Rsh	1	0	1	0	0	0	0	1	1	0	0	1	1	0	0	0	1	1	1	*	1	1	1	0	1	1	1	0	1	0	0	1	0	1	*	0	1	1	0	0
Pr	1	0	0	1	0	0	0	*	1	0	0	0	0	0	*	0	1	1	1	1	1	0	1	0	1	1	1	*	0	0	0	0	1	0	0	*	1	1	0	0
FP	1	0	1	0	1	1	0	*	1	0	0	0	*	0	0	*	1	0	1	1	0	0	1	0	1	1	1	0	1	0	0	0	1	0	0	0	1	1	0	0
F	0	0	0	0	1	1	0	*	1	0	0	0	0	0	0	*	1	0	0	1	0	0	1	0	1	1	1	0	1	0	0	0	1	0	0	0	1	1	0	0
C	1	1	1	0	0	0	0	1	1	*	0	0	1	1	1	0	1	1	1	1	0	1	1	0	1	1	1	0	0	0	0	0	1	0	0	1	1	1	0	0
S	1	1	1	1	0	0	0	1	1	0	0	1	1	1	1	0	0	1	1	1	0	1	1	0	1	0	1	*	0	0	0	0	1	0	0	1	0	0	0	0
P	1	0	1	0	0	1	0	1	1	0	0	1	1	1	1	0	0	1	1	1	0	1	0	0	1	0	1	1	0	0	0	0	1	0	0	1	0	0	0	0

Classification. Using directly (1) and (3) to write a simple Python code, we re-compute distances, correcting a couple of slips, and compute the corresponding distinguishabilities for the first time; distances have been multiplied by 2 and distinguishabilities by 4 only to facilitate reading; Muljačić used strict fuzziness quite sparsely, and so 4δ is either equal to its crisp lower bound $2d$ or only slightly above it[9].

$2d/4\delta$	R	D	I	Sa	Frl	Rsh	Pr	FP	F	C	S	P
R												
D	23/24											
I	37/38	26/26										
Sa	33/34	32/32	30/30									
Frl	37/38	26/26	32/32	38/38								
Rsh	38/40	25/26	31/32	40/42	19/20							
Pr	**38/42**	27/30	33/36	41/44	21/24	24/26						
FP	47/50	32/34	36/38	52/54	22/24	23/24	20/22					
F	54/56	**39/40**	43/44	59/60	29/30	30/30	21/24	9/12				
C	37/38	26/26	32/32	38/38	22/22	23/24	19/22	22/24	29/30			
S	31/32	30/30	30/30	34/34	30/30	31/32	26/30	34/36	41/42	14/14		
P	34/34	**39/40**	33/34	41/42	39/40	32/34	30/34	31/34	38/40	19/20	9/10	

Following Muljačić, in the 12×12-matrix we have left blank the all-zero secondary diagonal and, by symmetry, the triangle entries above it; we have omitted the uninformative R-row and the P-column. Take e.g. the remarkable case of Provençal Pr, Romanian R, French F and Dalmatic D: one has $d(Pr, R) < d(F, D)$ but instead $\delta(Pr, R) > \delta(F, D)$; one has also, with Portuguese P, $d(Pr, R) < d(P, D)$ but instead $\delta(Pr, R) > \delta(P, D)$. Note also that we are implicitly assuming, as did Muljačić himself, that features are non-interactive (independent) and equally important: were we to perform clustering on larger real-world data we would have to resort to methods which bioinformatics has nowadays made popular, bootstrapping, say. Even if linguists might object to the use of outdated material, we append the UPGMA tree of distances. Unsurprisingly, given the sparse use of fuzziness, the tree for distinguishabilities is virtually the same.



Evolution. We move to the relation between *decoding*, *distances* and *distinguishabilities* in language evolution. Assume we have k possible ancestor languages $A_1, \ldots, A_k$ for language L: which is the correct ancestor? A basic principle of decoding, in its generalized form as given in [1,2,3], tells that *decoding by minimum distance*, i.e. selecting a language A_u which minimizes in j the distance $d^*(A_j, L)$, is certainly successful whenever the distinguishabilities

$\delta^*(A_v, A_w)$ are sufficiently high: more precisely, if one assumes that the "corruption" due to time evolution cannot be $> w$, i.e. that the distance between the ancestor language and the "output language" L is $\leq w$, one is safe when the distinguishabilities between possible ancestors are all $> w$. In case of ties, one either guesses (*hard decoding*) or declares a detected error (*soft decoding*); in the latter case the reliability criterion is accordingly modified[1]; alternatively, one may resort to *list decoding* and rather provide the whole list of minimizing languages[5,13].

In language evolution one has to resort to a new tool, namely distinguishability: we stress once more that, unlike what happens with most metric string distances of practical use, Muljačić distinguishabilities are *not* trivial and are *explicitly needed*, cf. above Section 2. Let us ask the (politically incorrect and linguistically untenable) question: is Dalmatic a dialect of Italian or of Romanian? A minimum-distance decoder points to Romanian, $d(R, D) = 11.5 < d(I, D) = 13$, an unreliable verdict since the distinguishability between Italian and Romanian is only $\delta(R, I) = 9.5$. With even less political correctness, let us ask whether Provençal is a dialect of French or of Italian, taking for granted that it *must* be a dialect of the two: the verdict is French, $d(F, Pr) = 10.5 < d(I, Pr) = 16.5$, and now the distinguishability is high enough, $\delta(F, I) = 11$. Notice that we are ignoring the effect of possible homoplasies, unavoidably so in lack of a well-established evolutionary model comparable to those of bioinformatics. We are confident that the notion of distinguishability as opposed to distance may prove useful to build such models also in a linguistic context.

4. Conclusions

The results of this paper support use of Muljačić distance and Muljačić distinguishability as based on the standard logical operators of fuzzy logic. We did not present new material of direct linguistic interest, but rather put to work old and new tools for linguistic classification and linguistic evolution on the basis of historical data, in the hope that Muljačić' ideas might be successfully revived, extended and applied to up-to-date material. In[10] the first author is employing *four* variants of fuzzy distances to classify linguistic data, while the purport of distinguishabilities on linguistic evolution has still to be assessed[4]. The notion of fuzzy distinguishability has been already applied[9] to coding theory (error-correction and error-detection), unsurprisingly so because it is precisely in coding theory that string distinguishabilities first arose[17]; actually, the subtle difference between distance and distinguishability might prove useful also in bioinformatics, where one deals with long DNA strings, with possibly some fuzziness, rather than comparatively short strings of ill defined linguistic features.

Acknowledgements

The second author thanks FRA 2014 on *Learning Specification*, Trieste University, for partial support.

Appendix A. T-norms

T-norms $x \top y$ are "abstract" logical conjunctions and are ruled by the axioms:

(i) $x \top y = y \top x$, $x \top (y \top z) = (x \top y) \top z$ (commutativity and associativity)
(ii) $u \leq x, v \leq y$ implies $u \top v \leq x \top y$ (monotony)
(iii) $x \top 1 = x$, $x \top 0 = 0$ (neuter element and nullific)

Actually these axioms are slightly redundant, because $x \top 0 = 0$ can be derived from the rest. T-conorms $x \perp y$ are "abstract" logical conjunctions, c.f. e.g.[7], and are ruled by the same axioms, only replacing (iii) by (iii bis) $x \perp 1 = 1$, $x \perp 0 = x$ (nullific and neuter element). Once a T-norm is given its *dual* T-conorm is obtained by De Morgan's rule: $x \perp y \doteq \overline{\overline{x} \wedge \overline{y}}$. Remarkable examples of T-norms are:

minimum, non-interactivity: $x \top y = x \wedge y$
Łukasievicz: $x \top y = 0 \vee (1 - x - y)$
probabilistic, product: $x \top y = xy$
nilpotent minimum: $x \top y = 0$ if $x + y < 1$, else $x \top y = x \wedge y$

drastic: $x \top y = 0$ on the open unit square

The corresponding dual T-conorms are, respectively: $x \perp y = x \vee y$; $x \perp y = 1 \wedge (x + y)$; $x \perp y = x + y - xy$; $x \perp y = 1$ if $x + y > 1$, else $x \vee y$; $x \perp y = 1$ on the open unit square. In fuzzy arithmetic for fuzzy quantities *non-interactivity* is considered to be an adequate anolog of probabilistic independence [12] for random variables. Other examples are given in the body of the paper. Non-interactivity gives the largest T-norm and the smallest T-conorm, while drasticity gives the smallest T-norm and the largest T-conorm: so, whatever the T-norm and T-conorm one has $x \top y \leq x \wedge y$, $x \perp y \geq x \vee y$.

Appendix B. Fuzzy metric distances

The axioms for fuzzy metric distances $d(x, y)$ are:

(i) $0 \leq d(x, x) \leq d(x, y)$,
(ii) $d(x, y) = d(y, x)$ (symmetry)
(iii) $d(x, z) + d(z, y) \geq d(x, y)$ (triangular inequality)

Crisp objects x are those for which the self-distance $d(x, x)$ is zero, else they are (strictly) fuzzy. In our case Muljačić distance and Muljačić distinguishability are both fuzzy metrics, and the opposition crisp/fuzzy is the usual one. Note that a fuzzy metric can be "defuzzified" by simply imposing that self-distances should be all zero and by "gluing together" languages at distance 0: the approach taken in Section 3 above, however, is subtler and more respectful of fuzziness. Note also that fuzzy metric spaces as found usually in the literature are somehow at variance with our choice, which is however quite simple and quite "natural" in the present context.
To stress why the notion of distinguishability is *not* trivial take the "artificial" but meaningful example:

$d(x, y)$	a	b	c	d	e
a		1	1	1	1/2
b			1/4	1/2	1
c				1/4	3/4
d					1/2
e					

which is soon checked to be a crisp metric distance on $x, y \in \{a, b, c, d, e\}$; in particular $d(x, z) + d(z, y) \geq d(x, y)$ (here and below the all-0 secondary diagonal and the triangle below it have been left blank only to facilitate reading). Distinguishabilities can be computed by an exhaustive search:

$d(x, y)$	a	b	c	d	e
a		1	3/4	1/2	1/2
b			1/4	1/4	1/2
c				1/4	1/2
d					1/2
e					

The triangular inequality *falls*: $\delta(a, d) + \delta(d, b) = \frac{3}{4} < \delta(a, b) = 1$. As for the bounds (4), interval couples which achieve the lower bound are: $(a, d), (b, d), (b, e)$, the upper bound: $(a, b), (a, e), (b, c), (c, d), (d, e)$, while (c, e) and (a, c) give the intermediate values $\delta(c, e) = \frac{2}{3} d(c, e)$ and $\delta(a, c) = \frac{3}{4} d(a, c)$.

References

1. Bortolussi, L., Sgarro, A.: Possibilistic coding: error detection vs. error correction. Combining Soft Computing and Statistical Methods in Data Analysis, ser. Advances in Intelligent and Soft Computing, C. B. et al., Ed., vol. 77, pp. 41–48 (2010) Springer Verlag
2. Bortolussi, L., Dinu, L. P., Sgarro, A.: Spearman permutation distances and Shannon's distinguishability. Fundamenta Informaticae, vol. 118, no. 3, pp. 245–252 (2012)
3. Bortolussi, L., Dinu, L. P., Franzoi, L., Sgarro, A.: Coding theory: a general framework and two inverse problems. Fundamenta Informaticaes, 141, pp.297–310 (2015)
4. Ciobanu, A., Dinu, A. Dinu, L., Franzoi, L., Sgarro, A.: An evolutionary channel for ill-defined linguistic features. (2017) - work in progress
5. Csiszár, I., Kőrner,J.: Information Theory. Coding Theorems for Discrete Memoryless Systems. Cambridge Univeristy Press (2011)
6. Deza, M. M., Deza, E.: Dictionary of Distances. Elsevier B. V. (2006)
7. Dubois, D., Prade, H.: Fundamentals of Fuzzy Sets. Kluwer Academic Publishers (2000)
8. Franzoi, L.: (Ir)relevance in incomplete fuzzy arithmetic. SYNASC, pp. 287–291, (2016)
9. Franzoi, L., Sgarro, A.: Fuzzy Hamming distinguishability. FUZZ-IEEE (2017) - in press IEEE
10. Franzoi, L.: Jaccard-like fuzzy distances for computational linguistics. Submitted to SYNASC 2017
11. Kőrner, J., Orlitsky, A.: Zero-Error Information Theory. IEEE Transactions on Information Theory, vol. IT-44, pp. 11–32, (2002)
12. Klir, G. J., Folger, T. A.: Fuzzy Sets, Uncertainty and Information. Prentice Hall Int. (1988)
13. Lint, J. H. v.: Introduction to Coding Theory. Springer - Verlag Berlin Heidelberg (1999)
14. Muljačić, Z.: Die Klassifikation der romanischen Sprachen. RJbuch, XVIII, pp. 23–37 (1967)
15. Sgarro, A.: A fuzzy Hamming distance. Bulletin Math. de la Soc. Sci. Math. de la R. S. de Romanie, vol. 69, no. 1-2, pp. 137–144 (1977)
16. Sgarro, A.: Possibilistic Information theory: a coding theoretic approach. Fuzzy Sets and Systems, vol. 132, pp. 11-32 (2002)
17. Shannon, C.: The zero error capacity of noisy channel. IRE Transactions on Information Theory, vol. 2, pp. 8–19 (1956)