

Relation d'analogie et distance sur un alphabet défini par des traits

Laurent Miclet, Arnaud Delhay

► **To cite this version:**

Laurent Miclet, Arnaud Delhay. Relation d'analogie et distance sur un alphabet défini par des traits. [Rapport de recherche] RR-5244, INRIA. 2004, pp.17. inria-00070754

HAL Id: inria-00070754

<https://hal.inria.fr/inria-00070754>

Submitted on 19 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

***Relation d'analogie et distance sur un alphabet
défini par des traits.***

Laurent Miclet et Arnaud Delhay

N° 5244

Juillet 2004

THÈME 3



***Rapport
de recherche***

Relation d'analogie et distance sur un alphabet défini par des traits.

Laurent Miclet* et Arnaud Delhay†

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Cordial

Rapport de recherche n° 5244 — Juillet 2004 — 17 pages

Résumé : Cet article rappelle d'abord ce qu'est une relation d'analogie entre quatre ensembles finis (ou alphabets) selon la définition de Lepage. Il étudie le cas où ils sont définis par des traits binaires et présente ensuite une distance sur ces alphabets, compatible avec la relation d'analogie. Il étudie aussi comment construire des alphabets définis par des traits binaires au sein desquels se trouvent des relations d'analogie systématiques et examine le nombre d'analogies possibles dans ces alphabets. Pour finir, il prouve l'équivalence entre la définition de l'analogie entre ensembles de Lepage et celle de Stroppa.

Mots-clés : Relation d'analogie, ensembles finis, alphabets, résolution d'équation analogique.

Merci à F. Yvon et N. Stroppa, de l'ENST Paris, pour leur collaboration. Ce rapport a été écrit dans le cadre de l'action ASILA du programme interdisciplinaire CTAN 2003 du CNRS.

* miclet@enssat.fr

† delhay@univ-rennes1.fr

Analogical relation and distance on alphabets defined by features.

Abstract: This article firstly recalls what is an analogical relation between four finite sets (alphabets), according to the definition by Lepage. It studies the case when the alphabet is defined by binary features and gives a distance which is compatible with the analogical relation. It studies also how to construct alphabets on binary features among which there are systematical analogical relations and gives insights on the possible number of analogies in alphabets. Finally, it proves the equivalence of the definition of analogy between sets (as given by Lepage) and of that given by Stroppa.

Key-words: Analogy, Finite Sets, Alphabets, Analogical Equation.

Table des matières

1	Présentation : relation d’analogie, équations analogiques.	4
1.1	Généralités.	4
1.2	Axiomes pour la relation d’analogie et les équations analogiques.	4
1.3	L’analogie entre ensembles finis, ou alphabets	5
1.4	L’analogie entre multi-ensembles	7
1.5	L’analogie entre ensembles flous.	7
2	Analogie sur un alphabet défini par des traits	7
2.1	Notions de base	7
2.2	Résolution d’équations analogiques sur des traits et équations analogiques complètes.	8
3	Une distance pour l’analogie sur un alphabet défini par des traits	9
3.1	Présentation	9
3.2	δ_{ds} est une distance	9
3.3	δ_{ds} est compatible avec l’analogie	10
4	Construction d’alphabets définis par des traits pour l’analogie	11
4.1	Retour sur la notation en traits	11
4.2	La plus petite équation analogique complète sur des ensembles.	11
4.3	Augmentation de la taille de l’alphabet.	12
4.4	Questions ouvertes	13
4.5	Distance dans les alphabets de $2n$ objets.	14
5	Cohérence avec la définition algébrique de l’analogie entre ensembles.	15
	Références	17

1 Présentation : relation d'analogie, équations analogiques.

1.1 Généralités.

D'une manière informelle, une relation d'analogie entre quatre objets a , b , c et d s'énonce comme : « a est à b comme c est à d » et s'écrit classiquement

$$a : b \doteq c : d$$

Il s'agit par exemple du rapport numérique de proportionnalité (ou *règle de trois*) :

$$\ll 6 \text{ est à } 3 \text{ comme } 16 \text{ est à } 8 \gg$$

ou d'oppositions de traits phonétiques :

$$\ll /p/ \text{ est à } /b/ \text{ comme } /t/ \text{ est à } /d/ \gg$$

Dans le domaine des séquences, une analogie sera par exemple

$$\ll \text{ exact est à inexacte comme fini est à infinie } \gg$$

ou

$$\ll \text{ ab est à aabb comme cde est à ccddee } \gg$$

L'analogie peut servir à raisonner, par exemple sous la forme d'une *équation analogique* :

$$\text{Trouver } X \text{ connaissant } A, B \text{ et } C \text{ et sachant que } A \text{ est à } B \text{ comme } C \text{ est à } X.$$

qui s'écrit :

$$A : B \doteq C : X \quad \text{ou} \quad A : B \doteq C : ?$$

Sa *résolution* consiste à trouver, si possible, une ou plusieurs valeurs correctes à X .

L'analogie peut aussi se voir comme une technique d'apprentissage supervisé qui généralise la méthode du plus proche voisin. Cette dernière peut s'énoncer sur un exemple :

« Je ne sais pas à quelle espèce appartient cet animal x que je n'ai jamais vu auparavant. Mais parmi ceux que je connais, celui qui lui ressemble le plus est un chat. Je conclus que x est un chat ».

Une analogie dans le même domaine serait :

« Je n'ai jamais encore vu cet animal x . Mais je connais des chats noirs, des chats blancs et des cygnes blancs et x est à un cygne blanc comme un chat noir est à un chat blanc. Donc, j'appellerai x un cygne noir. »

Plus formellement, si on cherche la valeur $f(x)$ de la fonction f sur un objet x et si on dispose d'un ensemble \mathcal{S} d'exemples qui sont des couples objet-valeur de f sur des objets: $\mathcal{S} = \{(a_1, f(a_1)), \dots, (a_m, f(a_m))\}$, il s'agit de trouver trois objets a , b et c de \mathcal{S} tels que $x : a \doteq b : c$. On peut alors calculer la valeur de f sur l'objet x en résolvant l'équation : $f(x) : f(a) \doteq f(b) : f(c)$

L'analogie a été étudiée en particulier en logique philosophique, en psychologie, en linguistique, en intelligence artificielle. Dans les séquences, sa définition et ses propriétés mathématiques, qui relèvent de la théorie des langages, sont encore en partie du domaine de la recherche. Une bibliographie très complète sur les divers aspects de l'analogie se trouve dans le document d'habilitation de Y. Lepage ([2]).

1.2 Axiomes pour la relation d'analogie et les équations analogiques.

Reprenons la notion de base: une *relation d'analogie* entre quatre objets s'exprime par « A est à B comme C est à D » et se note par la formule :

$$A : B \doteq C : D$$

Quelque soit le type des objets, et sans préciser la sémantique des relations « est à » et « comme » une définition formelle de l'analogie part des axiomes suivants ([2]) :

$$\text{Symétrie de « comme » : } C : D \doteq A : B$$

$$\text{Permutation des moyens : } A : C \doteq B : D$$

D'où découlent cinq conséquences

$$\text{Inversion des rapports : } B : A \doteq D : C$$

$$\text{Permutation des extrêmes : } D : B \doteq C : A$$

$$\text{Symétrie de la lecture : } D : C \doteq B : A$$

$$B : D \doteq A : C$$

$$C : A \doteq D : B$$

Au total, une relation analogique entre quatre objets a donc huit formes équivalentes.

A partir de quatre objets, 24 permutations sont possibles, qui se partitionnent en trois groupes autour des analogies canoniques :

$$A : B \doteq C : D$$

$$B : A \doteq C : D$$

$$A : B \doteq D : C$$

Par définition, *résoudre une équation analogique* consiste à trouver une ou plusieurs solutions à l'équation

$$A : B \doteq C : X$$

Un autre axiome (*le déterminisme*) pose qu'il existe une solution unique à l'une des équations triviales (l'autre s'en déduit) :

$$A : A \doteq B : X \quad \Rightarrow \quad X = B$$

$$A : B \doteq A : X \quad \Rightarrow \quad X = B$$

Pour résoudre une équation de manière cohérente avec les axiomes, on peut donner la définition suivante : X est une *solution correcte* à l'équation analogique

$$A : B \doteq C : X$$

si X est solution de cette équation et est aussi solution des deux autres équations :

$$C : X \doteq A : B$$

$$A : C \doteq B : X$$

1.3 L'analogie entre ensembles finis, ou alphabets

1.3.1 Le théorème de Lepage

Quand les objets sont des ensembles finis, encore appelés *alphabets*, on sans perdre de généralité les voir comme des parties d'un ensemble général \mathcal{P} . En prenant la relation « comme » comme l'égalité ensembliste, Lepage a donné une définition de l'analogie qui respecte les axiomes. La formule

$$A : B \doteq C : D$$

se traduit informellement par :

pour que A soit à B comme C à D , il faut transformer A en B et C en D en ajoutant et retranchant les mêmes éléments à A et à C .

C'est le cas, par exemple, des quatre ensembles $A = \{t_1, t_2, t_3, t_4\}$, $B = \{t_1, t_2, t_3, t_5\}$ et $C = \{t_1, t_4, t_6, t_7\}$, $D = \{t_1, t_5, t_6, t_7\}$

On a retranché t_4 et ajouté t_5 à A et C pour obtenir B et D .

En formalisant, on trouve que les conditions suivantes sont nécessaires pour satisfaire les axiomes de l'analogie :

$$A \subset B \cup C \quad \text{et} \quad A \supset B \cap C$$

Lepage ([2]) démontre le théorème suivant, sous les axiomes cités au paragraphe 1.2 :

Théorème 1 *Soient trois ensembles A , B et C . L'équation analogique $A : B \doteq C : D$ d'inconnue D a une solution si et seulement si les conditions d'inclusion suivantes sont vérifiées :*

$$A \subset B \cup C \quad \text{et} \quad A \supset B \cap C$$

La solution, unique, est alors :

$$D = ((B \cup C) \setminus A) \cup (B \cap C)$$

□

On peut récrire cette solution un peu différemment, en notant \bar{A} le complémentaire de A dans \mathcal{P} , comme l'union de trois ensembles disjoints deux à deux :

$$D = (B \cap \bar{A}) \cup (C \cap \bar{A}) \cup (A \cap B \cap C)$$

D'une manière générale, ce théorème montre que la résolution de l'analogie entre quatre ensembles $A : B \doteq C : D$, dans la définition de Lepage (qui est compatible avec les axiomes) n'est possible que lorsque les conditions d'inclusion $A \subset B \cup C$ et $A \supset B \cap C$ sont respectées. On peut alors calculer chaque ensemble à partir des trois autres.

Nous verrons au paragraphe 5 une autre définition de l'analogie entre ensembles et nous prouverons qu'elle est équivalente à celle de Lepage.

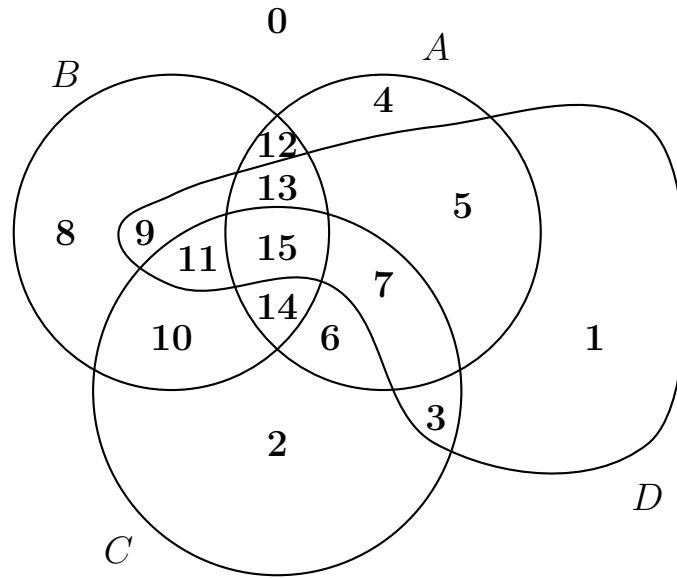


FIG. 1 – A, B, C et D décomposés en 16 sous ensembles disjoints deux à deux.

1.3.2 Une démonstration

La démonstration de Lepage est purement algébrique. Donnons-en une version plus « visuelle » en décomposant les quatre ensembles en 16 sous-ensembles disjoints deux à deux, comme sur la figure 1.

Des huit versions équivalentes de l'équation analogique $A : B \doteq C : D$, déduites des axiomes, on infère que les deux conditions d'inclusion :

$$A \subset B \cup C \text{ et } A \supset B \cap C$$

en impliquent six autres :

$$B \subset A \cup D \text{ et } B \supset A \cap D$$

$$C \subset A \cup D \text{ et } C \supset A \cap D$$

$$D \subset B \cup C \text{ et } D \supset B \cap C$$

En écrivant les ensembles comme l'union des sous-ensembles disjoints sous la forme :

$$A = \{4,5,6,7,12,13,14,15\}$$

$$B = \{8,9,10,11,12,13,14,15\}$$

$$C = \{2,3,6,7,10,11,14,15\}$$

$$D = \{1,3,5,7,9,11,13,15\}$$

Les 8 conditions d'inclusions impliquent que seuls les sous-ensembles notés 3, 6, 9, 12 et 15 sont non-vides. D'où finalement :

$$A = \{6,12,15\}$$

$$B = \{9,12,15\}$$

$$C = \{3,6,15\}$$

$$D = \{3,9,15\}$$

Ce qui donne par exemple :

$$D = (B \cap \bar{A}) \cup (C \cap \bar{A}) \cup (A \cap B \cap C)$$

Cette situation est résumée sur la figure 2 : des 16 sous-ensembles de départ, seuls 6 sont non vides. Ils sont notés E_0 à E_5 sur cette figure, avec la correspondance suivante avec la notation précédente :

E_0	E_1	E_2	E_3	E_4	E_5
0	3	12	9	6	15

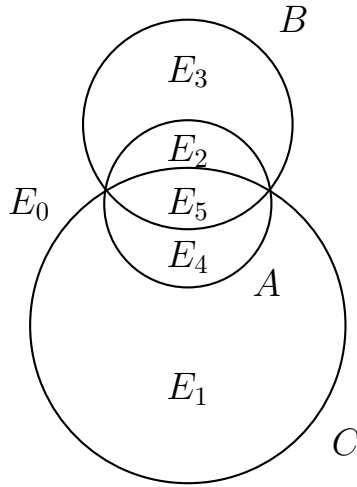


FIG. 2 – Quatre ensembles A , B , C et D en relation d'analogie $A : B \doteq C : D$ ont forcément la structure ci-dessus. On a : $D = E_1 \cup E_3 \cup E_5$

1.4 L'analogie entre multi-ensembles

On peut noter un multi-ensemble $\mathbf{A} = \{(a, f_A(a))\}$, où a est un élément de l'ensemble A et $f_A(a)$ son nombre d'occurrences. $f_A(a) = 0$ équivaut à $a \notin A$. Soit A , B , C et D quatre sous-ensembles d'un ensemble \mathcal{P} .

Lepage démontre que l'équation analogique entre multi-ensembles $\mathbf{A} : \mathbf{B} \doteq \mathbf{C} : \mathbf{X}$ se résout conformément aux axiomes si pour tous les éléments a de \mathcal{P} , on a :

$$f_A(a) \leq f_B(a) + f_C(a)$$

La solution unique est alors composée de tous les éléments a de \mathcal{P} avec pour nombre d'occurrences : $f_B(a) + f_C(a) - f_A(a)$.

Cette définition généralise parfaitement la précédente, si on voit un ensemble comme un multi-ensemble où les nombres d'occurrences valent 0 ou 1.

1.5 L'analogie entre ensembles flous.

Rappelons que le théorème de Lepage montre que l'analogie entre quatre ensembles $A : B \doteq C : D$ n'est possible que lorsque les conditions d'inclusion $A \subset B \cup C$ et $A \supset B \cap C$ sont respectées. On peut alors calculer chaque ensemble à partir des trois autres.

Si les ensembles considérés sont des sous-ensembles flous, on peut repartir des conditions d'inclusion en utilisant le concept d'implication floue et étudier¹ comment la résolution de l'équation par la formule

$$D = (B \cap \bar{A}) \cup (C \cap \bar{A}) \cup (A \cap B \cap C)$$

se transforme en terme de norme et de co-norme floue. Les axiomes de l'analogie doivent être préservés.

2 Analogie sur un alphabet défini par des traits

2.1 Notions de base

Soit un ensemble fini, que nous appellerons maintenant *alphabet*, composés d'éléments que nous appellerons des *objets*. Donnons-nous un ensemble de traits binaires, de telle sorte que chaque objet ait la valeur 0 ou 1 sur chaque trait. Nous pouvons assimiler un objet à l'ensemble de traits qui valent 1 pour lui et l'ensemble des objets peut être vu comme un *alphabet défini par des traits*.

1. Travail en cours.

En d'autres termes, un alphabet P est vu comme une partie d'un certain ensemble \mathcal{P} maximal, chaque trait correspond à un élément de \mathcal{P} et avoir un trait à 1 signifie pour P qu'il possède l'élément correspondant de \mathcal{P} .

Pour définir les objets de l'alphabet, nous aurons donc un tableau comme le suivant :

Objet	Valeurs des traits			
	t_1	t_2	...	t_m
P_1	1	0	...	1
P_1	1	1	...	0
P_3	1	0	...	0
			...	
P_p	0	1	...	1

Nous avons défini dans des publications précédentes ([3], [1]) ce que nous entendons par une distance *cohérente avec une relation d'analogie*. Nous avons introduit cette notion pour la construction d'un algorithme capable de calculer la solution d'une équation analogique sur des *séquences* d'objets. Elle s'appuie sur l'idée que si quatre objets sont en analogie, l'ensemble des valeurs des distances entre leur couples doit construire un parallélogramme.

Plus formellement, nous avons la définition suivante :

Définition 1 La distance δ est cohérente avec l'analogie si pour tout quadruplet P_a, P_b, P_c et P_d en analogie sous la forme :

$$P_a : P_b \doteq P_c : P_d$$

alors δ vérifie :

$$\delta(P_a, P_b) = \delta(P_c, P_d) \quad \text{et} \quad \delta(P_a, P_c) = \delta(P_b, P_d)$$

□

2.2 Résolution d'équations analogiques sur des traits et équations analogiques complètes.

Prenons, par exemple, les trois ensembles $A = \{t_1, t_2, t_3, t_4\}$, $B = \{t_1, t_2, t_3\}$ et $C = \{t_1, t_4, t_5, t_6, t_7, t_8, t_9\}$. Calculons la solution donnée par la formule de Lepage au théorème 1.

$$B \cup C = \{t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8, t_9\}$$

$$B \cap C = \{t_1\}$$

$$(B \cup C) \setminus A = \{t_5, t_6, t_7, t_8, t_9\}$$

alors

$$D = \{t_1, t_5, t_6, t_7, t_8, t_9\}$$

Comme nous nous intéressons à des ensembles définis par des traits, nous pouvons écrire l'exemple précédent par un tableau de traits :

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
A	1	1	1	1	0	0	0	0	0
B	1	1	1	0	0	0	0	0	0
C	1	0	0	1	1	1	1	1	1
D	1	0	0	0	1	1	1	1	1

Ceci revient à considérer que A, B et C sont des sous-ensembles d'un ensemble \mathcal{P} qui comprend en particulier les éléments t_1 à t_9 . On pourrait pour compléter, quand $A \cup B \cup C \neq \mathcal{P}$, ajouter les traits représentant les éléments de $\mathcal{P} \setminus (A \cup B \cup C)$ sous la forme de colonnes de quatre 0.

Il est clair, comme Lepage le remarque aussi, que l'ordre dans lesquels on considère les traits n'a pas d'importance et que l'on peut voir la résolution de l'équation $A : B \doteq C : D$ comme neuf équations analogiques élémentaires indépendantes, une par colonne.

Prenons donc ces équations élémentaires, qui correspondent à des ensembles définis par un seul trait binaire, c'est à dire réduits à un seul élément. Sur les huit équations analogiques élémentaires *a priori* possibles, seules six sont compatibles avec les conditions d'inclusion du théorème 1 : deux d'entre elles n'ont pas de solution : $1 : 0 \doteq 0 : X$ et $0 : 1 \doteq 1 : X$.

Le tableau suivant résume la situation (les équations sont écrites en colonne) :

$$\begin{array}{cccccccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & & 1 & 1 & 0 & \end{array}$$

Ce tableau permet donc de résoudre toute équation analogique à la manière de Lepage sur des objets définis par des traits : il suffit de résoudre colonne par colonne pour soit trouver une valeur unique, soit constater qu'il n'y a pas de solution. Il est clair que, comme tout ensemble fini peut se voir comme un alphabet défini par des traits et réciproquement, cette technique n'est qu'une réécriture équivalente du théorème de Lepage.

Nous proposons maintenant la définition suivante, pour caractériser les équations analogiques les plus intéressantes :

Définition 2 Une équation analogique entre ensembles définis par des traits est appelée complète quand elle a une solution qui utilise les six équations analogiques binaires ayant elles-mêmes une solution. \square

3 Une distance pour l'analogie sur un alphabet défini par des traits

3.1 Présentation

Nous proposons maintenant d'utiliser comme distance entre deux objets définis comme des ensembles de traits le cardinal de la différence symétrique entre ces deux objets.

Nous allons d'abord montrer que c'est effectivement une distance puis établir qu'elle est compatible avec l'analogie entre ensembles, dans un sens que nous précisons.

Définition 3 La différence symétrique de deux ensembles A et B est définie par : $A\Delta B = (A \cup B) \setminus (A \cap B) = (A \cup B) \cap (\overline{A \cap B})$. \square

Nous proposons donc de prendre comme distance sur les ensembles la quantité $\delta_{ds}(A, B) = |(A\Delta B)|$. Cette distance mesure le nombre d'éléments qui diffèrent entre deux ensembles. Avec cette définition, nous obtenons pour l'exemple précédent : $\delta_{ds}(A, C) = 7 = \delta_{ds}(B, D)$ et $\delta_{ds}(A, B) = 1 = \delta_{ds}(C, D)$.

3.2 δ_{ds} est une distance

Nous faisons ici la preuve que la distance que nous avons choisie est bien une distance au sens classique, c'est à dire qu'elle est définie positive, symétrique, et qu'elle vérifie l'inégalité triangulaire.

Vérifions que δ_{ds} sur l'ensemble des objets, noté \mathcal{P} , a ces propriétés :

- $\forall \mathcal{P}_1, \mathcal{P}_1 \in \mathcal{P}, \delta_{ds}(\mathcal{P}_1, \mathcal{P}_1) \geq 0$ (positive)

C'est toujours vrai car un cardinal est positif ou nul.

- $\delta_{ds}(\mathcal{P}_1, \mathcal{P}_1) = 0 \iff \mathcal{P}_1 = \mathcal{P}_1$ (définie)

Si $\delta_{ds}(\mathcal{P}_1, \mathcal{P}_1) = 0$ alors $(\mathcal{P}_1 \cup \mathcal{P}_1) \setminus (\mathcal{P}_1 \cap \mathcal{P}_1) = \emptyset$. Cela implique que $\mathcal{P}_1 \cup \mathcal{P}_1 = \mathcal{P}_1 \cap \mathcal{P}_1$ et donc que $\mathcal{P}_1 = \mathcal{P}_1$.

- $\forall \mathcal{P}_1, \mathcal{P}_1 \in \mathcal{P}, \delta_{ds}(\mathcal{P}_1, \mathcal{P}_1) = \delta_{ds}(\mathcal{P}_1, \mathcal{P}_1)$ (symétrie)

C'est vrai grâce à la symétrie de l'intersection et de l'union de deux ensembles.

- $\forall \mathcal{P}_1, \mathcal{P}_1, \mathcal{P}_3 \in \mathcal{P}, \delta_{ds}(\mathcal{P}_1, \mathcal{P}_3) \leq \delta_{ds}(\mathcal{P}_1, \mathcal{P}_1) + \delta_{ds}(\mathcal{P}_1, \mathcal{P}_3)$ (inégalité triangulaire)

Pour cette démonstration, nous reprenons le principe de la celle en section 3 : découper les ensembles \mathcal{P}_1 , \mathcal{P}_2 et \mathcal{P}_3 en sous-ensemble distincts deux à deux. Le cas général est représenté en figure 3. Ces partitions seront notées E_1 jusqu'à E_7 , et donnent les relations :

$$\mathcal{P}_1 = E_1 \cup E_4 \cup E_6 \cup E_7$$

$$\mathcal{P}_2 = E_2 \cup E_5 \cup E_6 \cup E_7$$

$$\mathcal{P}_3 = E_2 \cup E_5 \cup E_6 \cup E_7$$

Après calcul, on peut établir que :

$$\delta_{ds}(\mathcal{P}_1, \mathcal{P}_3) = |(E_1 \cup E_3 \cup E_4 \cup E_5 \cup E_6 \cup E_7) \setminus (E_4 \cup E_7)|$$

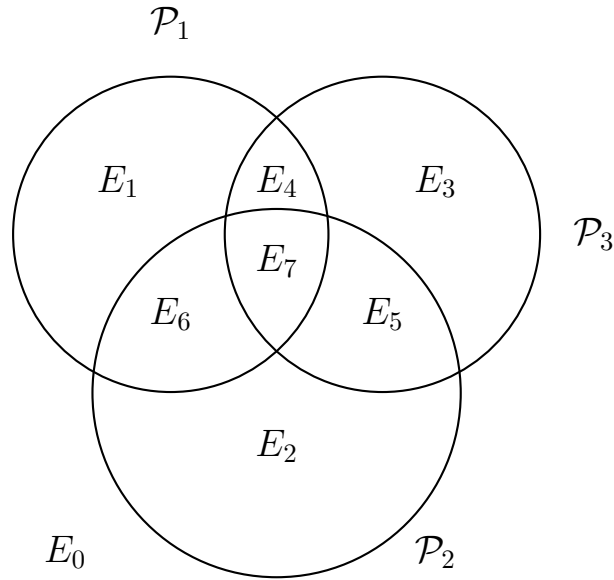


FIG. 3 – Représentation de \mathcal{P}_1 , \mathcal{P}_2 et \mathcal{P}_3 en intersection dans le cas général.

$$= |E_1 \cup E_3 \cup E_5 \cup E_6|$$

De la même façon, on a :

$$\delta_{ds}(\mathcal{P}_1, \mathcal{P}_1) = |E_1 \cup E_2 \cup E_4 \cup E_5|$$

$$\delta_{ds}(\mathcal{P}_1, \mathcal{P}_1) = |E_2 \cup E_3 \cup E_4 \cup E_6|$$

Comme les ensembles sont distincts, le cardinal de la somme est égal à la somme des cardinaux et on peut déduire que :

$$\delta_{ds}(\mathcal{P}_1, \mathcal{P}_1) + \delta_{ds}(\mathcal{P}_1, \mathcal{P}_3) = \delta_{ds}(\mathcal{P}_1, \mathcal{P}_3) + 2 \cdot |E_2| + 2 \cdot |E_4|$$

On a donc bien

$$\delta_{ds}(\mathcal{P}_1, \mathcal{P}_3) \leq \delta_{ds}(\mathcal{P}_1, \mathcal{P}_1) + \delta_{ds}(\mathcal{P}_1, \mathcal{P}_3)$$

Les ensembles \mathcal{P}_i ont été définis comme des parties de \mathcal{P} . On peut les représenter par un vecteur binaire dont les coordonnées suivent un ordre arbitraire sur les éléments de \mathcal{P} et valent 1 si l'élément correspondant leur appartient, 0 sinon. Pour \mathcal{P} fini, δ_{ds} est aussi la distance de Hamming des vecteurs binaires de taille $|\mathcal{P}|$

3.3 δ_{ds} est compatible avec l'analogie

La notion de distance cohérente avec une analogie a été définie au paragraphe 2.1. Nous allons démontrer la propriété suivante :

Théorème 2 *La distance δ_{ds} est cohérente avec l'analogie telle que définie par Lepage.*

Pour faire cette preuve, rappelons les conditions nécessaires au théorème de Lepage (théorème 1) : $A \subset B \cup C$ et $A \supset B \cap C$. Ces conditions font qu'il est possible de décomposer $B \cup C$ en 5 sous-ensembles comme représenté sur la figure 2. Ces 5 sous-ensembles sont notés E_1 à E_5 , ils sont en intersection vide deux à deux, et représentent le cas le plus général dès lors que chacun d'eux est non vide. On peut alors écrire que :

$$A = E_2 \cup E_4 \cup E_5 \quad \text{et} \quad B = E_2 \cup E_3 \cup E_5 \quad \text{et} \quad C = E_1 \cup E_4 \cup E_5$$

D'après le même théorème, D , la solution de l'équation analogique peut s'écrire :

$$D = ((B \cup C) \setminus A) \cup (B \cap C) = E_1 \cup E_3 \cup E_5$$

Nous avons posé que $\delta_{ds}(A, B) = |(A \cup B) \setminus (A \cap B)|$. Calculons donc

$$(A \cup B) \setminus (A \cap B) = E_3 \cup E_4$$

De la même façon, pour $\delta_{ds}(C,D)$,

$$(C \cup D) \setminus (C \cap D) = E_1 \cup E_4$$

Nous avons donc égalité des différences symétriques, et par conséquent de leur cardinal. Nous avons donc bien $\delta_{ds}(A,B) = \delta_{ds}(C,D)$ si $A : B \doteq C : D$. De la même façon, on pourrait prouver que $\delta_{ds}(A,C) = \delta_{ds}(B,D)$.

4 Construction d'alphabets définis par des traits pour l'analogie

4.1 Retour sur la notation en traits

Nous avons assimilé dans le paragraphe 2.2 les objets considérés aux ensembles des traits qu'il possèdent. Reprenons la figure 2 et remarquons que, du point de vue de distance comme de celui de la résolution d'équation analogiques, tous les éléments d'un même sous-ensemble E_i , pour $i = 1,5$, jouent le même rôle. L'exemple donné à ce paragraphe peut donc se simplifier de la manière suivante :

	t_1	t_2	t_4	t_5
A	1	1	1	0
B	1	1	0	0
C	1	0	1	1
D	1	0	0	1

en remarquant que le trait t_1 correspond à un seul élément, t_2 à deux, t_4 à un et t_5 à cinq.

Nous allons donc désormais changer légèrement la définition d'un *trait* en regroupant en un seul ceux qui sont fonctionnellement identiques mais en leur donnant une *cardinal* éventuellement différent de un.

En suivant la figure 2, seuls six traits sont nécessaires pour définir une analogie entre quatre objets. Nous allons modifier les notations pour les rendre plus faciles à généraliser. Notons désormais t_1 le trait binaire qui vaut *VRAI* ou 1 si un ensemble inclut le sous-ensemble E_3 , et *FAUX* ou 0 sinon, et nous faisons de même correspondre les cinq autres traits binaires aux cinq autres sous-ensembles par le tableau :

t_0	t_1	t_2	t_3	t_4	t_5
E_0	E_5	E_4	E_3	E_2	E_1

Nous définissons le *cardinal* $|t_i|$ d'un trait t_i comme celui du sous-ensemble qui lui correspond.

Renotons aussi A par A_1 , B par A'_1 , C par A_2 et D par A'_2 et appelons-les des *objets*, plutôt que des ensembles (de traits). L'ensemble de ces objets formera donc un *alphabet défini par des traits*. Nous pouvons maintenant réécrire l'équation analogique générale entre ensembles $A : B \doteq C : D$ comme

$$A_1 : A'_1 \doteq A_2 : A'_2$$

et la décrire par le tableau binaire suivant :

	t_0	t_1	t_2	t_3	t_4	t_5
A_1	0	1	1	0	1	0
A'_1	0	1	0	1	1	0
A_2	0	1	1	0	0	1
A'_2	0	1	0	1	0	1

4.2 La plus petite équation analogique complète sur des ensembles.

Nous cherchons ici à savoir combien il faut au minimum de traits pour définir une analogie complète entre quatre objets. La définition d'une analogie complète a été donnée au paragraphe 2.2.

Nous avons vu au même paragraphe qu'il y a six équations analogiques binaires ayant une solution, sur les huit possibles. Six traits au moins sont donc nécessaires si l'on veut qu'aucune paire parmi les quatre objets A_1 , A'_1 , A_2 et A'_2 ne soit en relation d'inclusion². Ceci se visualise sur la figure 2.

Nous pouvons donc dire que l'équation $A_1 : A'_1 \doteq A_2 : A'_2$ ainsi décrite comme le tableau binaire de six traits du paragraphe 4.1 est *la plus petite équation analogique complète sur les alphabets d'objets définis par des traits*.

2. Parmi les six, un trait joue un rôle particulier : celui dont la valeur est nulle sur les quatre objets. En interprétation ensembliste, il caractérise les éléments qui n'appartiennent à aucun des sous-ensembles A_1 , A'_1 , A_2 et A'_2 .

4.3 Augmentation de la taille de l'alphabet.

4.3.1 De quatre à cinq objets ; un résultat général.

Si on ajoute un cinquième objet A_3 aux quatre premiers, il est facile de constater qu'il ne peut être en analogie avec aucun triplet parmi ces quatre. Supposons d'abord qu'il soit construit sur les cinq mêmes traits. L'analogie $A_1 : A'_1 \doteq A_2 : A_3$, par exemple, impliquerait $A_3 = A'_2$.

Si on lui ajoute un trait "personnel" t_6 , qui vaut 0 partout et 1 seulement sur cet objet, on obtient le tableau suivant, où le symbole "?" est à remplacer par 0 ou 1.

	t_0	t_1	t_2	t_3	t_4	t_5	t_6
A_1	0	1	1	0	1	0	0
A'_1	0	1	0	1	1	0	0
A_2	0	1	1	0	0	1	0
A'_2	0	1	0	1	0	1	0
A_3	?	?	?	?	?	?	1

La sixième colonne donnera une équation binaire fausse, quelque soit le triplet choisi dans les quatre premiers objets. Elle serait tout aussi fausse si t_6 valait 0 sur A_3 et 1 partout ailleurs. Aucun trait différent ne peut être ajouté pour permettre à A_3 d'être en analogie avec un triplet pris parmi avec les quatre objets de base.

On ne peut donc avoir qu'un quadruplet en analogie dans un alphabet de cinq éléments³, quelque soit le nombre de traits.

En appliquant le même argument à un alphabet de taille quelconque, on trouve le résultat suivant :

Théorème 3 *Si un alphabet défini par des traits comprend p objets (avec $p \geq 5$) dont quatre sont en relation d'analogie (notons-les O_1, O_2, O_3, O_4 , on a donc : $O_1 : O_2 \doteq O_3 : O_4$), aucun objet parmi les $p - 4$ restants n'est en relation d'analogie avec un triplet pris parmi O_1, O_2, O_3, O_4 .*

□

4.3.2 De quatre à six objets.

On peut en revanche ajouter un sixième objet A'_3 de façon qu'aucun couple parmi les six objets ne soit en relation d'inclusion, et obtenir les relations analogiques suivantes :

$$\begin{aligned} A_1 : A'_1 &\doteq A_2 : A'_2 \\ A_1 : A'_1 &\doteq A_3 : A'_3 \\ A_2 : A'_2 &\doteq A_3 : A'_3 \end{aligned}$$

Cela se fait en définissant le trait t_6 de la manière suivante :

	t_0	t_1	t_2	t_3	t_4	t_5	t_6
A_1	0	1	1	0	1	0	0
A'_1	0	1	0	1	1	0	0
A_2	0	1	1	0	0	1	0
A'_2	0	1	0	1	0	1	0
A_3	0	1	1	0	0	0	1
A'_3	0	1	0	1	0	0	1

On a :

$$\begin{aligned} A_1 &\subset A'_1 \cup A_2 \\ A_1 &\supset A'_1 \cap A_2 \\ A'_2 &= ((A'_1 \cup A_2) \setminus A_1) \cup (A'_1 \cap A_2) \\ A_2 &\subset A'_2 \cup A_3 \\ A_2 &\supset A'_2 \cap A_3 \\ A'_3 &= ((A'_2 \cup A_3) \setminus A_2) \cup (A'_2 \cap A_3) \\ A_2 &\subset A'_2 \cup A_3 \\ A_2 &\supset A'_2 \cap A_3 \\ A'_3 &= ((A'_2 \cup A_3) \setminus A_2) \cup (A'_2 \cap A_3) \end{aligned}$$

3. Rappelons qu'un alphabet est un ensemble (fini, mais ça ne change rien pour cette remarque), donc que tous ses éléments sont distincts.

4.3.3 De six à $2n$ objets.

Pour passer à $2n$ objets en construisant un grand nombre d'analogies, il est suffisant de construire $n + 4$ traits, dont les valeurs binaires sont affectées selon les règles suivantes, pour tout i :

- t_0 à 0 pour A_i et A'_i ,
- t_1 à 1 pour A_i et A'_i ,
- t_2 à 1 pour A_i et à 0 pour A'_i ,
- t_3 à 0 pour A_i et à 1 pour A'_i ,
- t_{i+3} à 1 pour A_i et A'_i , à 0 sur tous les autres objets.

Cette construction est visualisée dans le tableau suivant :

	t_0	t_1	t_2	t_3	t_4	...	t_{i+2}	t_{i+3}	t_{i+4}	...	t_{j+2}	t_{j+3}	t_{j+4}	...	t_{n+3}
A_1	0	1	1	0	1	...	0	0	0	...	0	0	0	...	0
A'_1	0	1	0	1	1	...	0	0	0	...	0	0	0	...	0
...															
A_i	0	1	1	0	0	...	0	1	0	...	0	0	0	...	0
A'_i	0	1	0	0	0	...	0	1	0	...	0	0	0	...	0
...															
A_j	0	1	1	0	0	...	0	0	0	...	0	1	0	...	0
A'_j	0	1	0	1	0	...	0	0	0	...	0	1	0	...	0
...															
A_n	0	1	1	0	0	...	0	0	0	...	0	0	0	...	1
A'_n	0	1	0	1	0	...	0	0	0	...	0	0	0	...	1

On peut maintenant énoncer le théorème suivant :

Théorème 4 Dans un tel alphabet de $2n$ objets, aucun couple d'objets n'est en relation d'inclusion. On a de plus, pour tout i et j compris entre 1 et n :

$$A_i : A'_i \doteq A_j : A'_j$$

□

La démonstration consiste d'abord à vérifier que pour tout i et j , on a :

$$A_i \subset A'_i \cup A_j$$

$$A_i \supset A'_i \cap A_j$$

et

$$A'_j = ((A'_i \cup A_j) \setminus A_i) \cup (A'_i \cap A_j)$$

Le dernier point se constate en remarquant que :

$$\begin{aligned} A'_i \cup A_j &= \{t_1, t_2, t_3, t_{i+3}, t_{j+3}\} \\ A'_i \cap A_j &= \{t_1\} \\ (A'_i \cup A_j) \setminus A_i &= \{t_3, t_{i+3}, t_{j+3}\} \\ ((A'_i \cup A_j) \setminus A_i) \cup (A'_i \cap A_j) &= \{t_1, t_3, t_{j+3}\} = A'_j \end{aligned}$$

Pour un total de $2n$ objets, décrits de la sorte par $n + 4$ traits, on a donc $n(n - 1)/2$ analogies complètes ($i \neq j$), qui s'écrivent $A_i : A'_i \doteq A_j : A'_j$. Chacune de ces analogies en engendrant sept autres, on a donc au total $4n(n - 1)$ quadruplets d'objets en analogie dans l'alphabet qui en compte $2n$.

4.4 Questions ouvertes

Peut-il y avoir plus de $4n(n - 1)$ quadruplets d'objets en analogie complète dans un alphabet de $2n$ objets sur $n + 4$ traits, avec $n \geq 6$?

Beaucoup plus sans doute: il doit être possible, par exemple, d'ajouter $2n$ objets, correspondant à une permutation des colonnes des $2n$ premiers, en s'assurant qu'il n'y a pas deux objets identiques parmi les $4n$. Ca

ne doit pas être trop difficile, puisque les $2n$ nouveaux sont à choisir parmi $2^{n+4} - 2n$ objets possibles. Dans ce cas, on crée (au moins) $4n(n-1)$ analogies supplémentaires. Et on doit pouvoir recommencer... jusqu'où?

Combien y a-t-il de quadruplets d'objets en analogie complète dans un alphabet de 2^n objets définis sur n traits?

Pour $n = 6$, la réponse est assez facile. Il y a $2^6 = 64$ objets différents et on trouve $6! = 720$ quadruplets en analogie. Il suffit en effet de permuter les colonnes de la plus petite équation analogique et d'en déduire tous les quadruplets possibles.

Pour $n > 6$, la combinatoire est un peu plus complexe. La simulation exhaustive donne $21 \times 6!$ quadruplets pour $2^7 = 128$ objets et $276 \times 6!$ quadruplets pour $2^8 = 256$ objets. La formule de récurrence reste à trouver.

4.5 Distance dans les alphabets de $2n$ objets.

Reprenons les alphabets d'objets construits au paragraphe 4.3.3 pour y étudier la distance δ_{ds} définie au paragraphe 3. Rappelons qu'un trait correspond à un sous-ensemble (voir le paragraphe 4.1) et possède un cardinal. Si chaque trait est minimal, c'est à dire correspond à un sous-ensemble composé d'exactly un élément, on a les distances de différence symétrique suivantes :

$$\delta(A_i, A'_i) = \delta(A_i, A_j) = \delta(A'_i, A'_j) = \delta(A_j, A'_j) = 2$$

$$\delta(A_i, A'_j) = \delta(A'_i, A_j) = 4$$

Si les traits correspondent à un nombre d'éléments plus grand que un, les distances seront supérieures, mais on aura encore pour chaque quadruplet (A_i, A'_i, A_j, A'_j) un parallélogramme de distances dont la diagonale vaut la somme des côtés :

$$\begin{aligned} \delta(A_i, A'_i) &= \delta(A_j, A'_j) = |t_2| + |t_3| \\ \delta(A'_i, A_j) &= \delta(A'_i, A'_j) = |t_i| + |t_j| \\ \delta(A_i, A'_j) &= \delta(A'_i, A_j) = |t_2| + |t_3| + |t_i| + |t_j| \end{aligned}$$

Par exemple, pour les six objets construits sur les cinq traits t_1, t_2, t_3, t_4 et t_5 de cardinaux respectifs 2, 3, 1, 2 et 1, on a le tableau suivant ⁴

	t_1	t_2	t_3	t_4	t_5
A_1	1 1	1 1 1	0	0 0	1
A'_1	1 1	0 0 0	1	0 0	1
A_2	1 1	1 1 1	0	1 1	0
A'_2	1 1	0 0 0	0	1 1	1

La distance de différence symétrique vaut pour cet exemple :

$$\delta(A_1, A'_1) = \delta(A_2, A'_2) = |t_2| + |t_3| = 3 + 1 = 4$$

$$\delta(A_1, A_2) = \delta(A'_1, A'_2) = |t_4| + |t_5| = 2 + 1 = 3$$

$$\delta(A_1, A'_2) = \delta(A'_1, A_2) = 4 + 3 = 7$$

4. Le trait t_0 , dont on ignore le cardinal, n'a pas été mis. Quelque soit ce cardinal, le résultat est inchangé.

5 Cohérence avec la définition algébrique de l'analogie entre ensembles.

Stroppa ([4]) a étudié la relation d'analogie dans des ensembles munis de différentes lois de compositions. Dans le cas où les ensembles n'ont aucune propriété particulière, il donne la définition suivante :

Définition 4 *Quatre ensembles A, B, C et D sont en analogie $A : B \doteq C : D$ si et seulement si il existe quatre ensembles X, Y, Z et T tels que :*

$$A = X \cup Y$$

$$B = X \cup Z$$

$$C = T \cup Y$$

$$D = T \cup Z$$

□

Nous allons prouver dans ce paragraphe le résultat suivant :

Théorème 5 *La définition de l'analogie entre ensembles de Stroppa est équivalente, sous les contraintes d'inclusion de Lepage, à celle de l'analogie entre ensembles définis par des traits que nous avons donnée au paragraphe 2.2 et donc équivalente à celle de Lepage.*

□

Montrons d'abord que quatre ensembles en analogie entre ensembles au sens de Stroppa, sous les contraintes d'inclusion de Lepage, le sont aussi au sens de Lepage. Pour cela, nous allons utiliser notre description par traits.

Les ensembles X, Y, Z et T ne sont *a priori* pas disjoints deux à deux. Nous allons les décomposer en union d'ensembles disjoints deux à deux, en utilisant le cas général représenté sur la figure 4. Les 16 sous-ensembles disjoints sont encore ici notés par un simple numéro, de **0** à **15**.

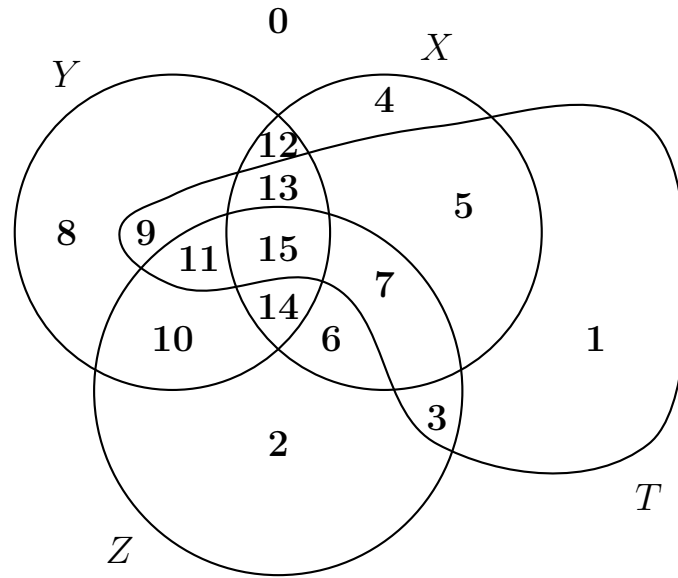


FIG. 4 – X, Y, Z et T décomposés en 16 sous ensembles disjoints deux à deux.

On a alors :

$$\begin{aligned} A &= X \cup Y = \{4,5,6,7,8,9,10,11,12,13,14,15\} \\ B &= X \cup Z = \{2,3,4,5,6,7,10,11,12,13,14,15\} \\ C &= T \cup Y = \{1,3,5,7,8,9,10,11,12,13,14,15\} \\ D &= T \cup Z = \{1,2,3,5,6,7,9,10,11,13,14,15\} \end{aligned}$$

Les contraintes d'intégrité données par Lepage :

$$\begin{aligned} B \cap C &\subset A \\ B \cup C &\supset A \end{aligned}$$

impliquent immédiatement :

$$\mathbf{3} = \emptyset$$

D'autre part, l'expression de D comme solution, sous la forme

$$D = ((B \cup C) \setminus A) \cup (B \cap C)$$

implique aussi que :

$$\mathbf{6} = \mathbf{9} = \mathbf{12} = \emptyset$$

On a donc nécessairement :

$$\begin{aligned} A &= X \cup Y = \{4,5,7,8,10,11,13,14,15\} \\ B &= X \cup Z = \{2,4,5,7,10,11,13,14,15\} \\ C &= T \cup Y = \{1,5,7,8,10,11,13,14,15\} \\ D &= T \cup Z = \{1,2,5,7,10,11,13,14,15\} \end{aligned}$$

Sous la forme de traits binaires, on trouve donc le tableau suivant, qui est (après réordonnement des colonnes) identique à la définition par traits de la plus petite analogie complète (voir le paragraphe 4.2).

	0	{5,7,10,11,13,14,15}	8	2	4	1
A	0	1	1	0	1	0
B	0	1	0	1	1	0
C	0	1	1	0	0	1
D	0	1	0	1	0	1

Ceci prouve donc, en utilisant la représentation par traits que nous avons étudiée, que quatre ensembles en analogie au sens de Yvon et Stroppa (sous les contraintes de Lepage) le sont aussi selon la définition de Lepage.

La réciproque s'établit en revenant à la démonstration que nous avons donnée du théorème de Lepage au paragraphe 1.3.1. En décomposant les ensembles A , B , C et D en 16 sous-ensembles disjoints deux à deux, on a, en tenant compte des contraintes d'inclusion :

$$\begin{aligned} A &= \{6,12,15\} \\ B &= \{9,12,15\} \\ C &= \{3,6,15\} \\ D &= \{3,9,15\} \end{aligned}$$

Pour prouver l'existence de quatre ensembles X , Y , Z et T , avec :

$$\begin{aligned} A &= X \cup Y \\ B &= X \cup Z \\ C &= T \cup Y \\ D &= T \cup Z \end{aligned}$$

il suffit de prendre $X = \{12,15\}$, $Y = \{6,15\}$, $Z = \{9,15\}$ et $T = \{3,15\}$.

Références

- [1] Arnaud DELHAY and Laurent MICLET. Solving analogical equations for learning by analogy with sequences. In *Proceedings of CAP-2004, Presses Universitaires de Grenoble*, pages 347–362, Montpellier, Juin 2004.
- [2] Yves LEPAGE. *De l'analogie rendant compte de la commutation en linguistique*. HDR, Université Joseph Fourier Grenoble I, may 2003.
- [3] Laurent MICLET and Arnaud DELHAY. Analogy on sequences: a definition and an algorithm. Technical Report 4969, INRIA, October 2003.
- [4] Nicolas STROPPA. Analogie et structures algébriques. Technical Report (to appear), ENST, 2004.



Unité de recherche INRIA Rennes
IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
<http://www.inria.fr>
ISSN 0249-6399