Analogy on text data

Yves LEPAGE

23rd of July 2022

Waseda University

The notion of analogy moved immediately from mathematics (natural numbers) to the study of the form of words and the study of their meaning. It was also used in a hidden form in comparative linguistics, and then explicitly at the birth of modern linguistics. In natural language processing, it has been used in computational morphology and distributional semantics. This talk will discuss the application of analogy to the processing of text data, be it characters, word forms, chunks, N-grams or sentences, in string or vector representations.

Thanks to the organisers of the IAMRL workshop
for inviting me to deliver this talk.

Analogy on Numbers and Words

Analogy on Natural Numbers (Vitrac, 1996; Schironi, 2007; Silva de Carvalho, 2013)

- Ἀναλογία δὲ ἡ τῶν λόγων ταυτότης (Euclid, V.3bis)
- Ἀναλογία δὲ ἐστιν ἡ τῶν λόγων ὁμοιότης (Euclid, V.7bis)
- *quae sit ratio pro proportione; ea Graece vocatur* ἀνα λόγον (Varro, *De ling. lat.*, X, 37)

Analogy on Natural Numbers (Vitrac, 1996; Schironi, 2007; Silva de Carvalho, 2013)

- ᾽Αναλογία δὲ ἡ τῶν λόγων ταυτότης (Euclid, V.3bis)
- ᾽Αναλογία δὲ ἐστιν ἡ τῶν λόγων ὁμοιότης (Euclid, V.7bis)
- *quae sit ratio pro proportione; ea Graece vocatur* ἀνα λόγον (Varro, *De ling. lat.*, X, 37)

Conformity of ratios between objects of the same type

Definitions: Continuous and Discrete Analogies (Michel, 1949; Lambert, 1996)

Theon of Smyrna (*Expositio*)

> ἀναλογία δ΄ ἐστὶ πλειόνων λόγων ὁμοιότης ἢ ταυτότης, τουτέστιν ἐν πλείοσιν ὅροις λόγων ὁμοιότης, ὅταν ὃν ἔχει λόγον ὁ πρῶτος πρὸς τὸν δεύτερον, τοῦτον ὁ δεύτερος πρὸς τὸν τρίτον ἢ ἄλλος τις πρὸς ἄλλον.

Aristotle (*Nic. Eth.*, 1131 a29)

> ἡ γὰρ ἀναλογία ἰσότης ἐστὶ λόγων, καὶ ἐν τέτταρσιν ἐλαχίστοις.
>
> ἡ μὲν οὖν διῃρημένη ὅτι ἐν τέτταρσιν,δῆλον.
>
> ἀλλὰ καὶ ἡ συνεχής [...] δὶς οὖν ἡ τοῦ β εἴρηται

Definitions: Continuous and Discrete Analogies (Michel, 1949; Lambert, 1996)

Theon of Smyrna (*Expositio*)

> ἀναλογία δ΄ ἐστὶ πλειόνων λόγων ὁμοιότης ἢ ταυτότης, τουτέστιν ἐν πλείοσιν ὅροις λόγων ὁμοιότης, ὅταν ὃν ἔχει λόγον ὁ πρῶτος πρὸς τὸν δεύτερον, τοῦτον ὁ δεύτερος πρὸς τὸν τρίτον ἢ ἄλλος τις πρὸς ἄλλον.

$$A : B \simeq B : C \simeq C : D \simeq D : F$$

Aristotle (*Nic. Eth.*, 1131 a29)

> ἡ γὰρ ἀναλογία ἰσότης ἐστὶ λόγων, καὶ ἐν τέτταρσιν ἐλαχίστοις.
> ἡ μὲν οὖν διῃρημένη ὅτι ἐν τέτταρσιν,δῆλον.
> ἀλλὰ καὶ ἡ συνεχής [...] δὶς οὖν ἡ τοῦ β εἴρηται

Definitions: Continuous and Discrete Analogies (Michel, 1949; Lambert, 1996)

Theon of Smyrna (*Expositio*)

> ἀναλογία δ΄ ἐστὶ πλειόνων λόγων ὁμοιότης ἢ ταυτότης, τουτέστιν ἐν πλείοσιν ὅροις λόγων ὁμοιότης, ὅταν ὃν ἔχει λόγον ὁ πρῶτος πρὸς τὸν δεύτερον, τοῦτον ὁ δεύτερος πρὸς τὸν τρίτον ἢ ἄλλος τις πρὸς ἄλλον.

$$A : B = B : C = C : D = D : F$$

Aristotle (*Nic. Eth.*, 1131 a29)

> ἡ γὰρ ἀναλογία ἰσότης ἐστὶ λόγων, καὶ ἐν τέτταρσιν ἐλαχίστοις.
>
> ἡ μὲν οὖν διῃρημένη ὅτι ἐν τέτταρσιν,δῆλον.
>
> ἀλλὰ καὶ ἡ συνεχής [...] δὶς οὖν ἡ τοῦ β εἴρηται

Definitions: Continuous and Discrete Analogies (Michel, 1949; Lambert, 1996)

Theon of Smyrna (*Expositio*)

> ἀναλογία δ΄ ἐστὶ πλειόνων λόγων ὁμοιότης ἢ ταυτότης, τουτέστιν ἐν πλείοσιν ὅροις λόγων ὁμοιότης, ὅταν ὃν ἔχει λόγον ὁ πρῶτος πρὸς τὸν δεύτερον, τοῦτον ὁ δεύτερος πρὸς τὸν τρίτον ἢ ἄλλος τις πρὸς ἄλλον.

$$A : B = B : C = C : D = D : F$$

Aristotle (*Nic. Eth.*, 1131 a29)

> ἡ γὰρ ἀναλογία ἰσότης ἐστὶ λόγων, καὶ ἐν τετταρσιν ἐλαχίστοις.
> ἡ μὲν οὖν διῃρημένη ὅτι ἐν τέτταρσιν,δῆλον.
> ἀλλὰ καὶ ἡ συνεχής [...] δὶς οὖν ἡ τοῦ β εἴρηται

$$A : B = C : D$$

Definitions: Continuous and Discrete Analogies (Michel, 1949; Lambert, 1996)

Theon of Smyrna (*Expositio*)

> ἀναλογία δ΄ ἐστὶ πλειόνων λόγων ὁμοιότης ἢ ταυτότης, τουτέστιν ἐν πλείοσιν ὅροις λόγων ὁμοιότης, ὅταν ὃν ἔχει λόγον ὁ πρῶτος πρὸς τὸν δεύτερον, τοῦτον ὁ δεύτερος πρὸς τὸν τρίτον ἢ ἄλλος τις πρὸς ἄλλον.

$$A : B = B : C = C : D = D : F$$

Aristotle (*Nic. Eth.*, 1131 a29)

> ἡ γὰρ ἀναλογία ἰσότης ἐστὶ λόγων, καὶ ἐν τετταρσιν ἐλαχίστοις.
> ἡ μὲν οὖν διῃρημένη ὅτι ἐν τέτταρσιν,δῆλον.
> ἀλλὰ καὶ ἡ συνεχής [...] δὶς οὖν ἡ τοῦ β εἴρηται

$$A : B = B : D$$

Definitions: Continuous and Discrete Analogies (Michel, 1949; Lambert, 1996)

Theon of Smyrna (*Expositio*)

> ἀναλογία δ΄ ἐστὶ πλειόνων λόγων ὁμοιότης ἢ ταυτότης, τουτέστιν ἐν πλείοσιν ὅροις λόγων ὁμοιότης, ὅταν ὃν ἔχει λόγον ὁ πρῶτος πρὸς τὸν δεύτερον, τοῦτον ὁ δεύτερος πρὸς τὸν τρίτον ἢ ἄλλος τις πρὸς ἄλλον.

$$A : B = B : C = C : D = D : F$$

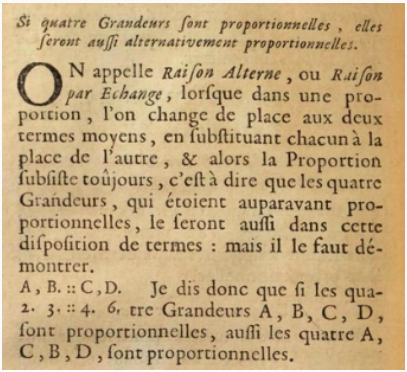
Aristotle (*Nic. Eth.*, 1131 a29)

> ἡ γὰρ ἀναλογία ἰσότης ἐστὶ λόγων, καὶ ἐν τέτταρσιν ἐλαχίστοις.
> ἡ μὲν οὖν διῃρημένη ὅτι ἐν τέτταρσιν,δῆλον.
> ἀλλὰ καὶ ἡ συνεχής [...] δὶς οὖν ἡ τοῦ β εἴρηται

$$A : B = C : D$$

The Exchange the Means

Elements d'Euclide Expliquez, tirés du cours de M. Ozanam, 1711.

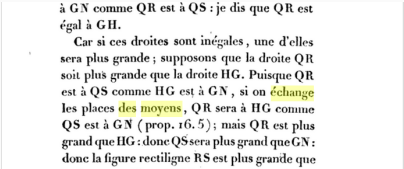


Si quatre Grandeurs ſont proportionnelles, elles ſeront auſſi alternativement proportionnelles.

ON appelle *Raiſon Alterne*, ou *Raiſon par Echange*, lorſque dans une proportion, l'on change de place aux deux termes moyens, en ſubſtituant chacun à la place de l'autre, & alors la Proportion ſubſiſte toûjours, c'eſt à dire que les quatre Grandeurs, qui étoient auparavant proportionnelles, le feront auſſi dans cette diſpoſition de termes : mais il le faut démontrer.

A, B. :: C, D. Je dis donc que ſi les qua-
2. 3. :: 4. 6. tre Grandeurs A, B, C, D, ſont proportionnelles, auſſi les quatre A, C, B, D, ſont proportionnelles.

Euclid (*Elements*), trad. Peyrard, 1804.

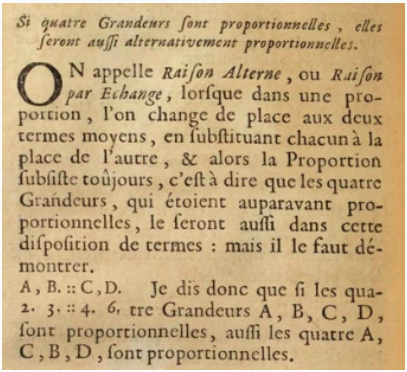


à GN comme QR est à QS : je dis que QR est égal à GH.

Car si ces droites sont inégales, une d'elles sera plus grande ; supposons que la droite QR soit plus grande que la droite HG. Puisque QR est à QS comme HG est à GN, si on échange les places des moyens, QR sera à HG comme QS est à GN (prop. 16. 5) ; mais QR est plus grand que HG : donc QS sera plus grand que GN : donc la figure rectiligne RS est plus grande que

The Exchange the Means

Elements d'Euclide Expliquez, tirés du cours de M. Ozanam, 1711.

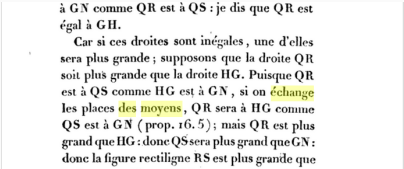


Si quatre Grandeurs ſont proportionnelles, elles ſeront auſſi alternativement proportionnelles.

ON appelle *Raiſon Alterne*, ou *Raiſon par Echange*, lorſque dans une proportion, l'on change de place aux deux termes moyens, en ſubſtituant chacun à la place de l'autre, & alors la Proportion ſubſiſte toûjours, c'eſt à dire que les quatre Grandeurs, qui étoient auparavant proportionnelles, le feront auſſi dans cette diſpoſition de termes : mais il le faut démontrer.

A, B. :: C, D. Je dis donc que ſi les qua-
2. 3. :: 4. 6. tre Grandeurs A, B, C, D, ſont proportionnelles, auſſi les quatre A, C, B, D, ſont proportionnelles.

Euclid (*Elements*), trad. Peyrard, 1804.



à GN comme QR est à QS : je dis que QR est égal à GH.

Car si ces droites sont inégales, une d'elles sera plus grande ; supposons que la droite QR soit plus grande que la droite HG. Puisque QR est à QS comme HG est à GN, si on échange les places des moyens, QR sera à HG comme QS est à GN (prop. 16. 5) ; mais QR est plus grand que HG : donc QS sera plus grand que GN : donc la figure rectiligne RS est plus grande que

$$A : B :: C : D \Leftrightarrow A : C :: B : D$$

From Numbers

$$8 : 2 :: 12 : 3$$

From Numbers

Geometric analogy: $8 \div 2 = 12 \div 3$

From Numbers

$$8 : 5 :: 12 : 9$$

From Numbers

Arithmetic analogy: $8 - 5 = 12 - 9$

From Numbers to Vectors (Turney and Littman, 2005; Turney, 2006, 2008a)

$$\begin{pmatrix} -0.26 \\ -0.77 \\ \vdots \\ 0.39 \end{pmatrix} \quad \begin{pmatrix} 0.34 \\ -0.23 \\ \vdots \\ 0.20 \end{pmatrix} \quad \begin{pmatrix} 0.40 \\ 0.99 \\ \vdots \\ 0.27 \end{pmatrix} \quad \begin{pmatrix} -0.20 \\ -0.35 \\ \vdots \\ 0.44 \end{pmatrix}$$

(fictitious numbers)

From Numbers to Vectors (Turney and Littman, 2005; Turney, 2006, 2008a)

$$\begin{pmatrix} -0.26 \\ -0.77 \\ \vdots \\ 0.39 \end{pmatrix} - \begin{pmatrix} 0.34 \\ -0.23 \\ \vdots \\ 0.20 \end{pmatrix} \stackrel{?}{=} \begin{pmatrix} 0.40 \\ 0.99 \\ \vdots \\ 0.27 \end{pmatrix} - \begin{pmatrix} -0.20 \\ -0.35 \\ \vdots \\ 0.44 \end{pmatrix}$$

(fictitious numbers)

From Numbers to Vector Representations of Words (Turney, 2008b; Turney and Pantel, 2010; Mikolov et al., 2013a,b; Levy and Goldberg, 2014)

$$\overrightarrow{male} - \overrightarrow{female} \simeq \overrightarrow{king} - \overrightarrow{queen}$$

$$\begin{pmatrix} -0.26 \\ -0.77 \\ \vdots \\ 0.39 \end{pmatrix} - \begin{pmatrix} 0.34 \\ -0.23 \\ \vdots \\ 0.20 \end{pmatrix} \simeq \begin{pmatrix} 0.40 \\ 0.99 \\ \vdots \\ 0.27 \end{pmatrix} - \begin{pmatrix} -0.20 \\ -0.35 \\ \vdots \\ 0.44 \end{pmatrix}$$

(fictitious numbers)

From Numbers To words

male : *female* :: *king* : *queen*

Analogy on Numbers and Words

Analogy on Words

Analogy on Word Forms (Douay, 1985; Douay and Pinto, 1991)

Aristarchus of Samothrace on *Iliad* XXIV, 8

πειρῶν/πείρων?

χείρω : χείρων :: πείρω : x $\Rightarrow$ $x =$

Quintilianus, *Ars orationis*, 6, § 10 et 11.

> *[. . .] qui reprehenderant quod hoc verbo usus essem, pepigi; [. . .] Nos, praeter auctoritatem oratorum atque historicorum analogia quoque dictum tuebamur. Nam eum in xii tabulis legeremus, ni ita pagunt inveniebamus simile huic, cadunt; inde prima positio, etiamsi vetustate exoleverat, apparebat; pago ut cado; unde non erat dubium sic pepigi nos dicere, ut cecidi.*

cado : *pago* :: *cecidi* : x $\Rightarrow$ $x =$

Analogy on Word Forms (Douay, 1985; Douay and Pinto, 1991)

Aristarchus of Samothrace on *Iliad* XXIV, 8

πειρῶν/πείρων?

κείρω : κείρων :: πείρω : x $\Rightarrow$ $x =$ πείρων

Quintilianus, *Ars orationis*, 6, § 10 et 11.

> *[. . .] qui reprehenderant quod hoc verbo usus essem, pepigi; [. . .] Nos, praeter auctoritatem oratorum atque historicorum analogia quoque dictum tuebamur. Nam eum in xii tabulis legeremus, ni ita pagunt inveniebamus simile huic, cadunt; inde prima positio, etiamsi vetustate exoleverat, apparebat; pago ut cado; unde non erat dubium sic pepigi nos dicere, ut cecidi.*

cado : *pago* :: *cecidi* : x $\Rightarrow$ $x =$

Analogy on Word Forms (Douay, 1985; Douay and Pinto, 1991)

Aristarchus of Samothrace on *Iliad* XXIV, 8

πειρῶν/πείρων?

κείρω : κείρων :: πείρω : x $\Rightarrow$ $x =$ πείρων

Quintilianus, *Ars orationis*, 6, § 10 et 11.

> *[. . .] qui reprehenderant quod hoc verbo usus essem, pepigi; [. . .] Nos, praeter auctoritatem oratorum atque historicorum analogia quoque dictum tuebamur. Nam eum in xii tabulis legeremus, ni ita pagunt inveniebamus simile huic, cadunt; inde prima positio, etiamsi vetustate exoleverat, apparebat; pago ut cado; unde non erat dubium sic pepigi nos dicere, ut cecidi.*

cado : *pago* :: *cecidi* : x $\Rightarrow$ $x =$ *pepigi*

Analogy across Two States of the Same Language

Mounin, *Histoire de la linguistique – Des origines au* XX^{e} *siècle*, 1967, p. 127.

> *On voit poindre aussi [à la Renaissance] les premières règles de correspondance phonétique (*x *latin =* ss *italien,* i *latin de* litera = e *fermé de* lettera*, etc.) ; p. ex., chez Tolomei, ou Castelvetro en Italie, qui commencent aussi à se servir de l'analogie comparative : si* habeo $>$ haggio*, en italien,* creggio *doit s'expliquer par l'existence d'un* credeo *[forme classique:* credo*]*

Analogy across Two States of the Same Language

Mounin, *Histoire de la linguistique – Des origines au* XX^{e} *siècle*, 1967, p. 127.

> *On voit poindre aussi [à la Renaissance] les premières règles de correspondance phonétique (*x *latin =* ss *italien,* i *latin de* litera *=* e *fermé de* lettera*, etc.) ; p. ex., chez Tolomei, ou Castelvetro en Italie, qui commencent aussi à se servir de l'analogie comparative : si* habeo $>$ haggio*, en italien,* creggio *doit s'expliquer par l'existence d'un* credeo *[forme classique:* credo*]*

$$haggio : habeo :: creggio : x \quad \Rightarrow \quad x =$$

Analogy across Two States of the Same Language

Mounin, *Histoire de la linguistique – Des origines au* XX^{e} *siècle*, 1967, p. 127.

> *On voit poindre aussi [à la Renaissance] les premières règles de correspondance phonétique (*x *latin =* ss *italien,* i *latin de* litera *=* e *fermé de* lettera*, etc.) ; p. ex., chez Tolomei, ou Castelvetro en Italie, qui commencent aussi à se servir de l'analogie comparative : si* habeo $>$ haggio*, en italien,* creggio *doit s'expliquer par l'existence d'un* credeo *[forme classique:* credo*]*

$$haggio : habeo :: creggio : x \quad \Rightarrow \quad x = {}^{*}credeo$$

Analogies Across Several Languages: Comparative Linguistics

Franz Bopp, *Vergleichende Grammatik der indogermanischen Sprachen,* 1833–1852.

Sanscrit.	Grec.	Latin.	Gothique.
Pâda-s	*πούς*	*pes*	*fôtus*
páńćan	*πέμπε*	*quinque*	*fimf*
pûrṇá	*πλέος*	*plenus*	*fulls*
pitár	*πατήρ*	*pater*	*fadar*
upári	*ὑπέρ*	*super*	*ufar*
bhrâtar	*Φράτωρ*	*frater*	*bróthar*
bar	*Φέρω*	*fero*	*baira*
twam	*τύ*	*tu*	*thu*
tam (accusatif)	*τόν*	*is-tum*	*thana*
tráya-s	*τρεῖς*	*tres*	*threis*
dwâu	*δύο*	*duo*	*twai*
dákśiṇâ	*δεξίτα*	*dextra*	*taihswô*
śvan pour *kvan*	*κύων*	*canis*	*hunths*
paśú pour *pakú*		*pecus*	*faihu*
śvaśura pour *svákura*	*ἑκυρός*	*socer*	*svaihra*
dáśan pour *dákan*	*δέκα*	*decem*	*taihun*
dáru pour *dákru*	*δάκρυ*	*lacrima*	*tagr*
haṅsá pour *ghansá*	*χήν*	*(h)anser*	*gans*
hyas pour *ghyas*	*χθές*	*heri*	*gistra*
lih pour *ligh*	*λείχω*	*lingo*	*laigô*
gńâ pour *gnâ*	*γιγνώσκω*	*gnosco*	*kan*
gâti pour *gâti*	*γένος*	*genus*	*kuni*
gânu pour *gânu*	*γόνυ*	*genu*	*kniu.*

Analogies Across Several Languages: Comparative Linguistics

Franz Bopp, *Vergleichende Grammatik der indogermanischen Sprachen,* 1833–1852.

Sanscrit.	Grec.	Latin.	Gothique.
Pâda-s	*πούς*	*pes*	*fôtus*
páńćan	*πέμπε*	*quinque*	*fimf*
pûrṇá	*πλέος*	*plenus*	*fulls*
pitár	*πατήρ*	*pater*	*fadar*
upári	*ὑπέρ*	*super*	*ufar*
bhrâtar	*φράτωρ*	*frater*	*bróthar*
bar	*φέρω*	*fero*	*baira*
twam	*τύ*	*tu*	*thu*
tam (accusatif)	*τόν*	*is-tum*	*thana*
tráya-s	*τρεῖς*	*tres*	*threis*
dwâu	*δύο*	*duo*	*twai*
dákshiṇâ	*δεξία*	*dextra*	*taihswô*
śvan pour *kvan*	*κύων*	*canis*	*hunths*
paśú pour *pakú*		*pecus*	*faihu*
śváśura pour *svákura*	*ἑκυρός*	*socer*	*swaihra*
dáśan pour *dákan*	*δέκα*	*decem*	*taihun*
dáru pour *dákru*	*δάκρυ*	*lacrima*	*tagr*
haṅsá pour *ghansá*	*χήν*	*(h)anser*	*gans*
hyas pour *ghyas*	*χθές*	*heri*	*gistra*
liṅ pour *liṅ*	*λείχω*	*lingo*	*laigô*
ǵnâ pour *gnâ*	*γιγνώσκω*	*gnosco*	*kan*
ǵâti pour *gâti*	*γένος*	*genus*	*kuni*
ǵánu pour *gánu*	*γόνυ*	*genu*	*kniu.*

Latin	Greek	‘meaning’
semi	*hemi*	‘half’
sextem	*hexa*	‘six’
septem	*hepta*	‘seven’
serpens	*herpes*	‘snake’
similis	*homolos*	‘similar’

German	Dutch	‘meaning’
Haus	*huis*	‘house’
Schaum	*schuim*	‘foam’
braun	*bruin*	‘brown’
ausbreiten	*uitbreiden*	‘to spread’
Weltraum	*wereldruim*	‘space’

Analogies in One Same Language: Modern Linguistics (Welcomme, 2010)

Hermann Paul, *Prinzipien der Sprachgeschichte,* 1880, chap. V, p. 106–107.

Doch gibt es auch stoffliche Gruppen, die lediglich auf die Bedeutung und nicht auf den Laut basiert sind, vgl.

Ochse (Stier) – Kuh,
Mann – Weib,
Knabe – Mädchen,
Vater – Mutter,
Sohn – Tochter,
Bruder – Schwester,
Mönch – Nonne,

Analogy of meaning

Hermann Paul, *Prinzipien der Sprachgeschichte,* 1880, chap. V, p. 107.

Die Basis für die Gleichung ist dabei die Übereinstimmung in der Bedeutung des stofflichen Elements nach der einen und des formalen Elements nach der andern Richtung, weshalb wir diese Art als stofflich-formale Proportionengruppen bezeichnen wollen. Es kann dazu auch eine lautliche Übereinstimmung nach beiden Richtungen treten, vgl.

Tag : Tages : Tage =
Arm : Armes : Arme =
Fisch : Fisches : Fische

führen : Führer : Führung =
erziehen : Erzieher : Erziehrung
etc.

Analogy of form

Sound Changes vs. Analogy – Modern Linguistics

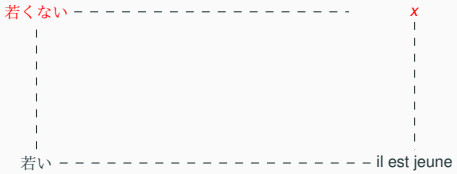
Saussure, *Cours de linguistique générale,* 1916, p. 221–222.

> *[. . .] un procédé [...] et que nous ramenons dès maintenant au calcul de la quatrième proportionnelle:*
>
> $$\text{ōrātōrem} : \text{ōrātor} :: \text{honōrem} : x \quad \Rightarrow \quad x = \text{honor}$$
>
> *On voit donc que, pour contrebalancer l'action diversifiante du changement phonétique (honōs : honōrem), l'analogie a de nouveau unifié les formes et rétabli la régularité (honor : honōrem).*

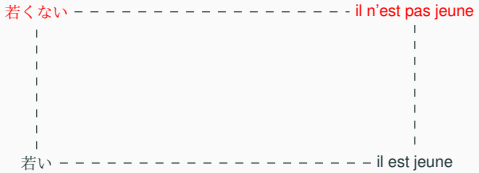
VII. BC.	II BC.	Classical Latin
actor, actorem *orator, oratorem* *honos, honosem*	*actor, actorem* *orator, oratorem* *honos, honorem*	*actor, actorem* *orator, oratorem* *honor, honorem*

Analogy in NLP

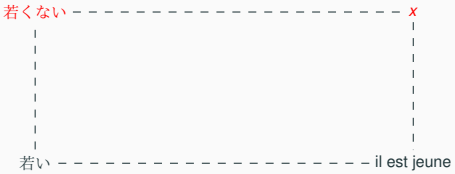
Analogies Across Two Domains (Nagao, 1984)



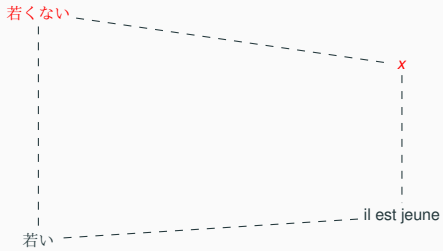
Analogies Across Two Domains (Nagao, 1984)



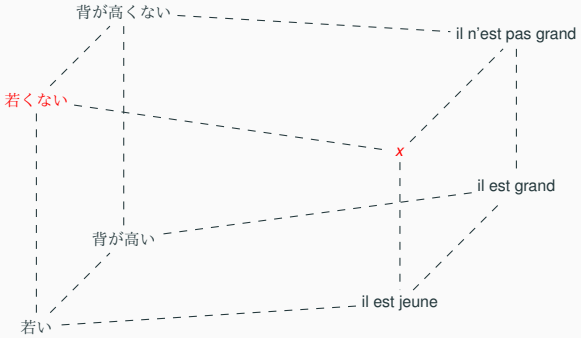
From One Analogy across Two Domains (Nagao, 1984)



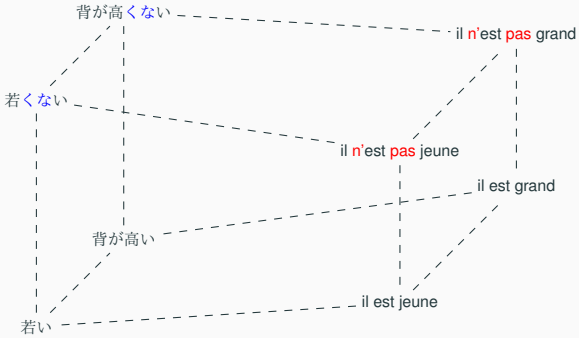
I.e., One Bilingual Analogy



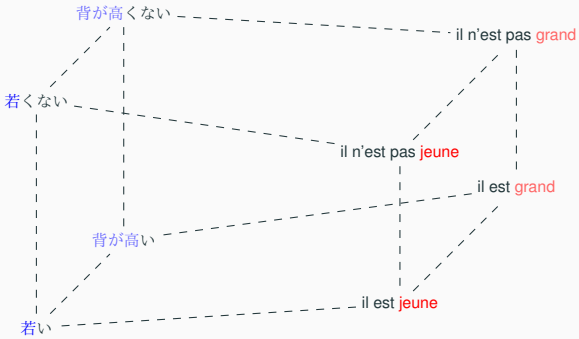
To Two Monolingual Analogies That Correspond



I.e., Two Analogies – Each in One Domain (Lepage and Denoual, 2005)



I.e., Two Analogies – Each in One Domain (Lepage and Denoual, 2005)



One vs. Two Domain(s)

Two domains		
	similarity only	
contiguity only	bird	fish
	wings	fins
	lungs	gills
	air	water
	to fly	to swim
	⋮	⋮

One vs. Two Domain(s)

Two domains		
	similarity only	
contiguity only	bird	fish
	wings	fins
	lungs	gills
	air	water
	to fly	to swim
	⋮	⋮

One domain	
contiguity and similarity	
contiguity and similarity	aslama / muslimun
	arsala / mursilun

Analogy on strings

Analogy on strings

The Vector Part

Parikh Vectors

a string: $s = abacab$

Parikh Vectors

$$\text{a string: } s = abacab \quad \rightarrow \quad \overrightarrow{abacab} = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Parikh Vectors

$$\text{a string: } s = abacab \quad \rightarrow \quad \overrightarrow{abacab} = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{matrix} |s|_a \\ |s|_b \\ |s|_c \\ |s|_d \\ \vdots \\ |s|_z \end{matrix}$$

Parikh Vectors

$$\text{a string: } s = abacab \quad \rightarrow \quad \overrightarrow{abacab} = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{matrix} |s|_a \\ |s|_b \\ |s|_c \\ |s|_d \\ \vdots \\ |s|_z \end{matrix}$$

Such an interpretation excludes reduplication and repetition:

~~*cado* : *pago* :: *cecidi* : *pepigi*~~

~~*guru* : *pelajar* :: *guru guru* : *pelajar pelajar*~~

Parikh Vectors are not Enough for Strings

- handling the number of characters (strings $\simeq$ vectors)

$$ab : abab :: ababab : abababab$$

Parikh Vectors are not Enough for Strings

- handling the number of characters (strings $\simeq$ vectors)

$$ab : abab :: ababab : abababab$$

- no handling of negative numbers (strings $\simeq$ multi-sets)

$$ababab : ab :: ab : x \quad \Rightarrow \quad x = -1\times a \ \ and \ \ -1\times b$$

Parikh Vectors are not Enough for Strings

- handling the number of characters (strings $\simeq$ vectors)

$$ab : abab :: ababab : abababab$$

- no handling of negative numbers (strings $\simeq$ multi-sets)

$$ababab : ab :: ab : x \quad \Rightarrow \quad x = -1\times a \ \ and \ \ -1\times b$$

- no handling of order of characters (strings $=$ what?)

$$ab : abab :: ababab : x \quad \Rightarrow \quad x = aaaabbbb$$

Axioms

Axioms

The Usually Admitted Axioms

Equivalent Forms

PROPORTION: *(Mathémat.) comme on compare deux grandeurs d'où résulte un rapport ou une raison (voyez* RAISON, RAPPORT*) ; aussi l'on peut comparer deux rapports d'où résulte une proportion, lorsque les rapports comparés, ou ce qui est la même chose, leurs exposans se trouvent égaux.*

La proportion fondamentale étant a . b :: c . d, il y a sept manieres d'en déplacer les termes, sous la condition prescrite ; mais de ces sept manieres, il n'y en a que deux qui aient mérité l'attention des anciens géometres, & auxquelles il leur ait plû de donner des noms particuliers.

Ils nomment alternando ou permutando celle-ci, a . c :: b . d, où l'on ne fait que transposer entr'eux les deux moyens.

Ils nomment invertendo cette autre, b . a :: d . c, où l'on ne fait que renverser chacun des deux rapports primitifs, mettant le conséquent à la place de l'antécédent, & réciproquement.

. . .

[Article de Rallier des Ourmes]

L'Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers

Dihedral Group D_8 (Lepage, 2018a)

$$A : B :: C : D \Leftrightarrow \begin{matrix} A & B \\ C & D \end{matrix}$$

. . . .	$\begin{matrix} A & B \\ C & D \end{matrix}$	$\begin{matrix} B & D \\ A & C \end{matrix}$	$\begin{matrix} D & C \\ B & A \end{matrix}$	$\begin{matrix} C & A \\ D & B \end{matrix}$	$\begin{matrix} A & C \\ B & D \end{matrix}$	$\begin{matrix} B & A \\ D & C \end{matrix}$	$\begin{matrix} D & B \\ C & A \end{matrix}$	$\begin{matrix} C & D \\ A & B \end{matrix}$
$\begin{matrix} A & B \\ C & D \end{matrix}$	$\begin{matrix} A & B \\ C & D \end{matrix}$	$\begin{matrix} B & D \\ A & C \end{matrix}$	$\begin{matrix} D & C \\ B & A \end{matrix}$	$\begin{matrix} C & A \\ D & B \end{matrix}$	$\begin{matrix} A & C \\ B & D \end{matrix}$	$\begin{matrix} B & A \\ D & C \end{matrix}$	$\begin{matrix} D & B \\ C & A \end{matrix}$	$\begin{matrix} C & D \\ A & B \end{matrix}$
$\begin{matrix} B & D \\ A & C \end{matrix}$	$\begin{matrix} B & D \\ A & C \end{matrix}$	$\begin{matrix} D & C \\ B & A \end{matrix}$	$\begin{matrix} C & A \\ D & B \end{matrix}$	$\begin{matrix} A & B \\ C & D \end{matrix}$	$\begin{matrix} B & A \\ D & C \end{matrix}$	$\begin{matrix} D & B \\ C & A \end{matrix}$	$\begin{matrix} C & D \\ A & B \end{matrix}$	$\begin{matrix} A & C \\ B & D \end{matrix}$
$\begin{matrix} D & C \\ B & A \end{matrix}$	$\begin{matrix} D & C \\ B & A \end{matrix}$	$\begin{matrix} C & A \\ D & B \end{matrix}$	$\begin{matrix} A & B \\ C & D \end{matrix}$	$\begin{matrix} B & D \\ A & C \end{matrix}$	$\begin{matrix} D & B \\ C & A \end{matrix}$	$\begin{matrix} C & D \\ A & B \end{matrix}$	$\begin{matrix} A & C \\ B & D \end{matrix}$	$\begin{matrix} B & A \\ D & C \end{matrix}$
$\begin{matrix} C & A \\ D & B \end{matrix}$	$\begin{matrix} C & A \\ D & B \end{matrix}$	$\begin{matrix} A & B \\ C & D \end{matrix}$	$\begin{matrix} B & D \\ A & C \end{matrix}$	$\begin{matrix} D & C \\ B & A \end{matrix}$	$\begin{matrix} C & D \\ A & B \end{matrix}$	$\begin{matrix} A & C \\ B & D \end{matrix}$	$\begin{matrix} B & A \\ D & C \end{matrix}$	$\begin{matrix} D & B \\ C & A \end{matrix}$
$\begin{matrix} A & C \\ B & D \end{matrix}$	$\begin{matrix} A & C \\ B & D \end{matrix}$	$\begin{matrix} C & D \\ A & B \end{matrix}$	$\begin{matrix} D & B \\ C & A \end{matrix}$	$\begin{matrix} B & A \\ D & C \end{matrix}$	$\begin{matrix} A & B \\ C & D \end{matrix}$	$\begin{matrix} C & A \\ D & B \end{matrix}$	$\begin{matrix} D & C \\ B & A \end{matrix}$	$\begin{matrix} B & D \\ A & C \end{matrix}$
$\begin{matrix} B & A \\ D & C \end{matrix}$	$\begin{matrix} B & A \\ D & C \end{matrix}$	$\begin{matrix} A & C \\ B & D \end{matrix}$	$\begin{matrix} C & D \\ A & B \end{matrix}$	$\begin{matrix} D & B \\ C & A \end{matrix}$	$\begin{matrix} B & D \\ A & C \end{matrix}$	$\begin{matrix} A & B \\ C & D \end{matrix}$	$\begin{matrix} C & A \\ D & B \end{matrix}$	$\begin{matrix} D & C \\ B & A \end{matrix}$
$\begin{matrix} D & B \\ C & A \end{matrix}$	$\begin{matrix} D & B \\ C & A \end{matrix}$	$\begin{matrix} B & A \\ D & C \end{matrix}$	$\begin{matrix} A & C \\ B & D \end{matrix}$	$\begin{matrix} C & D \\ A & B \end{matrix}$	$\begin{matrix} D & C \\ B & A \end{matrix}$	$\begin{matrix} B & D \\ A & C \end{matrix}$	$\begin{matrix} A & B \\ C & D \end{matrix}$	$\begin{matrix} C & A \\ D & B \end{matrix}$
$\begin{matrix} C & D \\ A & B \end{matrix}$	$\begin{matrix} C & D \\ A & B \end{matrix}$	$\begin{matrix} D & B \\ C & A \end{matrix}$	$\begin{matrix} B & A \\ D & C \end{matrix}$	$\begin{matrix} A & C \\ B & D \end{matrix}$	$\begin{matrix} C & A \\ D & B \end{matrix}$	$\begin{matrix} D & C \\ B & A \end{matrix}$	$\begin{matrix} B & D \\ A & C \end{matrix}$	$\begin{matrix} A & B \\ C & D \end{matrix}$

Dihedral Group D_8 (Lepage, 2018a)

$$A : B :: C : D \Leftrightarrow \begin{matrix} A & B \\ C & D \end{matrix}$$

D_8	e	a	a^2	a^3	x	ax	a^2x	a^3x
e	e	a	a^2	a^3	x	a^x	a^2x	a^3x
a	a	a^2	a^3	e	ax	a^2x	a^3x	x
a^2	a^2	a^3	e	a	a^2x	a^3x	x	ax
a^3	a^3	e	a	a^2	a^3x	x	ax	a^2x
x	x	a^3x	a^2x	ax	e	a^3	a^2	a
ax	ax	x	a^3x	a^2x	a	e	a^3	a^2
a^2x	a^2x	ax	x	a^3x	a^2	a	e	a^3
a^3x	a^3x	a^2x	ax	x	a^3	a^2	a	e

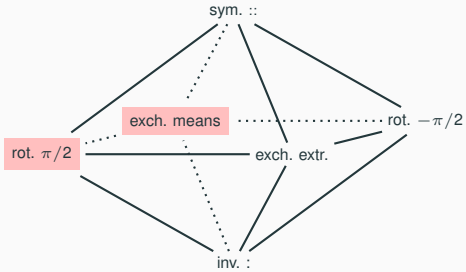
Dihedral Group D_8 (Lepage, 2018a)

$$A : B :: C : D \Leftrightarrow \begin{matrix} A & B \\ C & D \end{matrix}$$

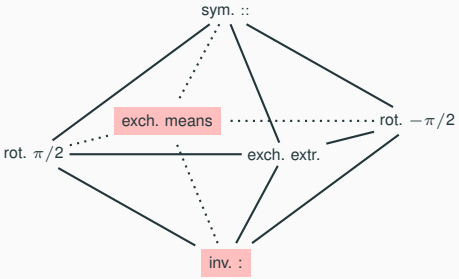
D_8	e	a	a^2	a^3	x	ax	a^2x	a^3x
e	e	a	a^2	a^3	x	a^x	a^2x	a^3x
a	a	a^2	a^3	e	ax	a^2x	a^3x	x
a^2	a^2	a^3	e	a	a^2x	a^3x	x	ax
a^3	a^3	e	a	a^2	a^3x	x	ax	a^2x
x	x	a^3x	a^2x	ax	e	a^3	a^2	a
ax	ax	x	a^3x	a^2x	a	e	a^3	a^2
a^2x	a^2x	ax	x	a^3x	a^2	a	e	a^3
a^3x	a^3x	a^2x	ax	x	a^3	a^2	a	e

D_8	corner transformation	square	analogy transformation	equivalent form
e	identity	$\begin{matrix} A & B \\ C & D \end{matrix}$	identity	$A : B :: C : D$
a	rotation by $\pi/2$	$\begin{matrix} B & D \\ A & C \end{matrix}$		$B : D :: A : C$
a^2	rotation by π	$\begin{matrix} D & C \\ B & A \end{matrix}$	inverse of reading	$D : C :: B : A$
a^3	rotation by $3\pi/2 = -\pi/2$	$\begin{matrix} C & A \\ D & B \end{matrix}$		$C : A :: D : B$
x	reflection about first diagonal	$\begin{matrix} A & C \\ B & D \end{matrix}$	exchange of the means	$A : C :: B : D$
ax	reflection about vertical axis	$\begin{matrix} B & A \\ D & C \end{matrix}$	inverse of ratios	$B : A :: D : C$
a^2x	reflection about second diagonal	$\begin{matrix} D & B \\ C & A \end{matrix}$	exchange of the extremes	$D : B :: C : A$
a^3x	reflection about horizontal axis	$\begin{matrix} C & D \\ A & B \end{matrix}$	symmetry of conformity	$C : D :: A : B$

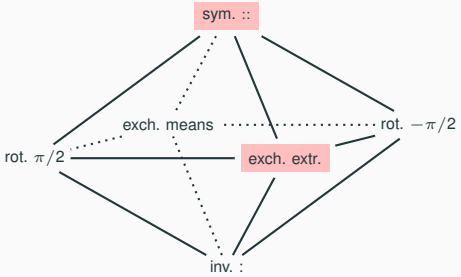
Basic Axioms: Octahedron of Transformations



Basic Axioms: Octahedron of Transformations



Basic Axioms: Octahedron of Transformations



Axioms

Need for Additional Axioms or Postulates

Natural Interpretation Analogy on Sets

A	B	C	$D = (B \Leftrightarrow \neg A) \Leftrightarrow C$
F	F	F	F
F	F	T	T
F	T	F	T
F	T	T	F
T	F	F	T
T	F	T	F
T	T	F	F
T	T	T	T

- Logical interpretation of sets: against intuition

$$\{a, b\} : \{b\} :: \{c\} : \{a, c\}$$

- "Container" intuition of sets: no solution to $\{a, b\} : \{b\} :: \{c\} : x$ because a appears neither in $B = \{b\}$ nor in $C = \{c\}$.

Natural Interpretation Analogy on Sets

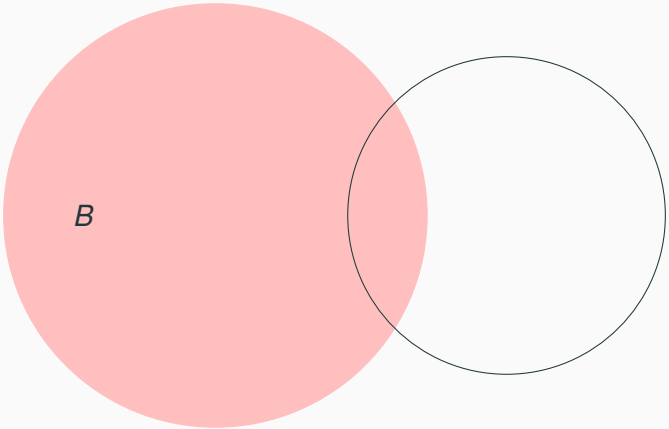
A	B	C	$D = (a \in B \Leftrightarrow a \notin A) \Leftrightarrow a \in C$
F	F	F	F
F	F	T	T
F	T	F	T
F	T	T	F
T	F	F	T
T	F	T	F
T	T	F	F
T	T	T	T

- Logical interpretation of sets: against intuition

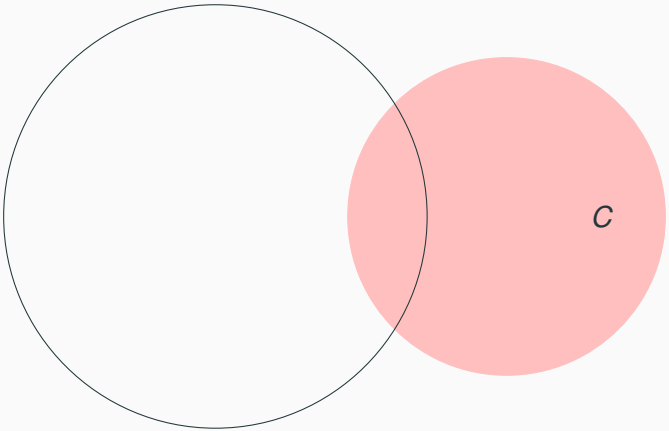
$$\{a, b\} : \{b\} :: \{c\} : \{a, c\}$$

- "Container" intuition of sets: no solution to $\{a, b\} : \{b\} :: \{c\} : x$ because a appears neither in $B = \{b\}$ nor in $C = \{c\}$.

Additional Constraints for Analogy on Sets

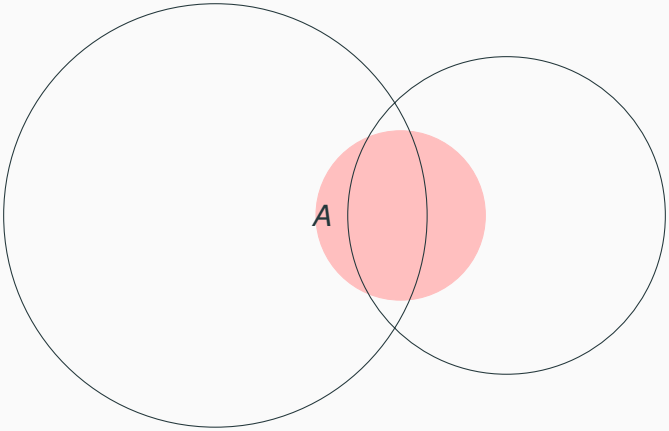


Additional Constraints for Analogy on Sets



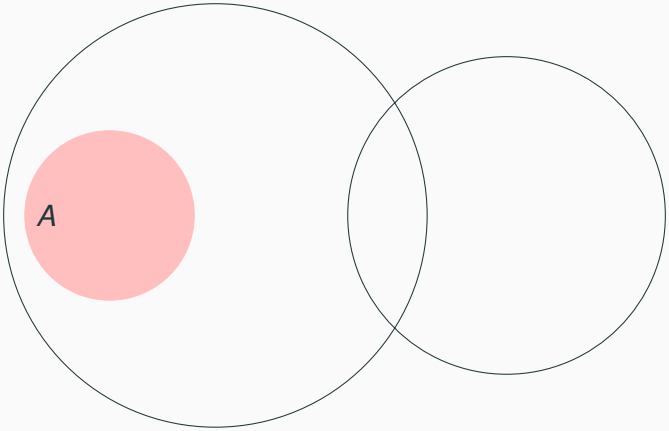
Additional Constraints for Analogy on Sets

$A \subseteq B \cup C$



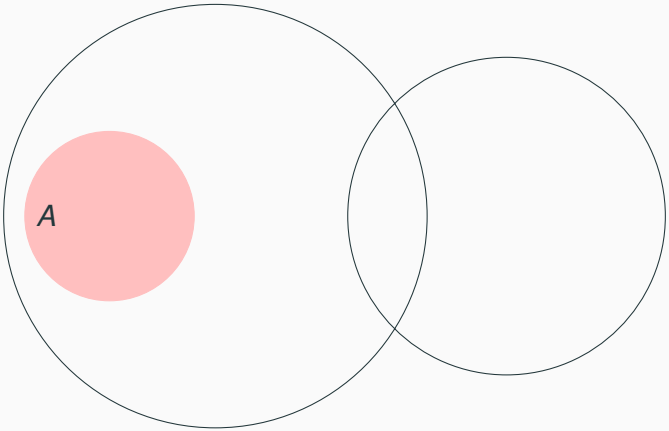
Additional Constraints for Analogy on Sets

$A \subseteq B \cup C$



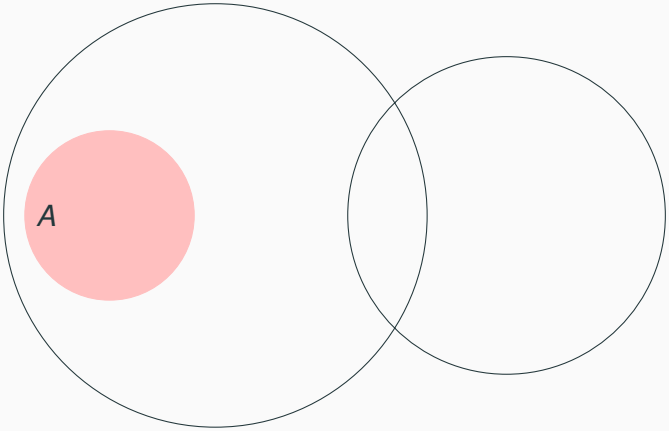
Additional Constraints for Analogy on Sets

$$A \subseteq B \cup C \land \mathcal{U} \setminus A \subseteq \mathcal{U} \setminus B \cup \mathcal{U} \setminus C$$



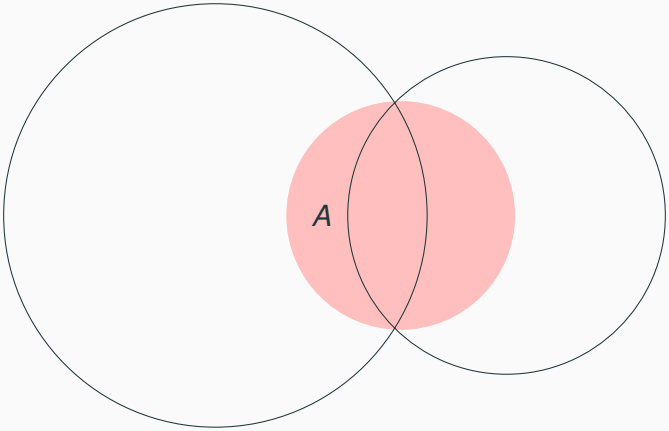
Additional Constraints for Analogy on Sets

$A \subseteq B \cup C \ \wedge \ A \supseteq B \cap C$



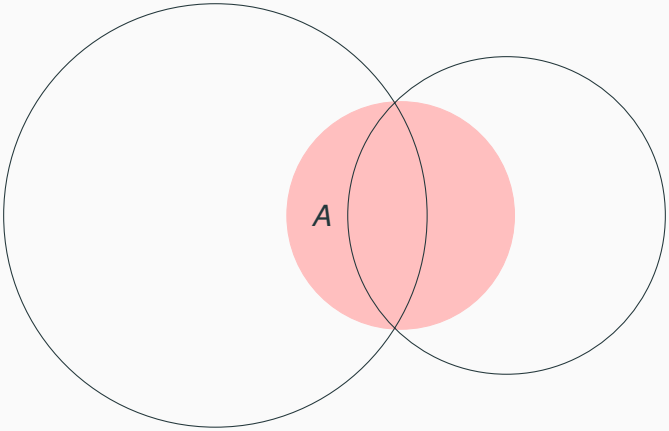
Additional Constraints for Analogy on Sets

$A \subseteq B \cup C \ \wedge \ A \supseteq B \cap C$



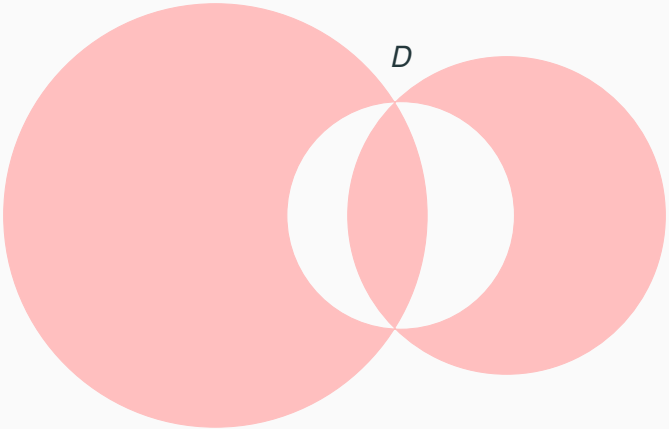
Additional Constraints for Analogy on Sets

$$A \subseteq B \cup C \ \wedge \ A \supseteq B \cap C \quad \Rightarrow \quad A : B :: C :$$



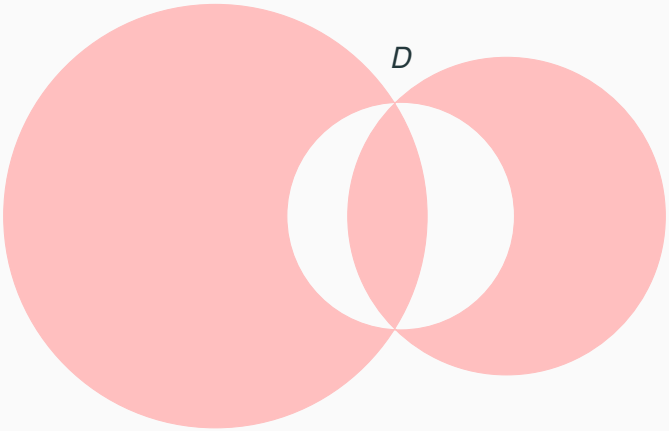
Additional Constraints for Analogy on Sets

$$A \subseteq B \cup C \ \wedge \ A \supseteq B \cap C \quad \Rightarrow \quad A : B :: C : D$$



Additional Constraints for Analogy on Sets

$$A \subseteq B \cup C \ \wedge \ A \supseteq B \cap C \quad \Rightarrow \quad A : B :: C : ((B \cup C) \setminus A) \cup (B \cap C)$$



Additional Postulates (Lepage, 2003)

- Similarity: Distribution of features of A onto B or C
- Contiguity: Preservation of analogy by inversion of objects (inversion of objects is an extreme case of contiguity)

Axioms

Analogy Induced by Algebraic Structures

Algebraic Structures (Stroppa, 2005; Lepage, 2018c)

	internal op. $\star$	associativity	zero elt.	inverse elt for all elt
magma	yes			
semi-group	yes	yes		
monoid	yes	yes	yes	
group	yes	yes	yes	yes

Algebraic Structures (Stroppa, 2005; Lepage, 2018c)

	internal op. $\star$	associativity	zero elt.	inverse elt for all elt
magma	yes			
semi-group	yes	yes		
monoid	yes	yes	yes	
group	yes	yes	yes	yes

- Commutativity is a different notion

Algebraic Structures (Stroppa, 2005; Lepage, 2018c)

	internal op. $\star$	associativity	zero elt.	inverse elt for all elt
magma	yes			
semi-group	yes	yes		
monoid	yes	yes	yes	
group	yes	yes	yes	yes

- Commutativity is a different notion
- $A : B :: C : D \overset{\Delta?}{\Longleftrightarrow} A \star D = C \star D$

Algebraic Structures (Stroppa, 2005; Lepage, 2018c)

	internal op. $\star$	associativity	zero elt.	inverse elt for all elt
magma	yes			
semi-group	yes	yes		
monoid	yes	yes	yes	
group	yes	yes	yes	yes

- Commutativity is a different notion
- $A : B :: C : D \overset{\Delta?}{\iff} A \star D = C \star D$ or $D \star C$?

Algebraic Structures (Stroppa, 2005; Lepage, 2018c)

	internal op. $\star$	associativity	zero elt.	inverse elt for all elt
magma	yes			
semi-group	yes	yes		
monoid	yes	yes	yes	
group	yes	yes	yes	yes

- Commutativity is a different notion
- $A : B :: C : D \overset{\Delta?}{\iff} A \star D = C \star D$ or $D \star C$? $\overset{\Delta?}{\iff} A \star B^{-1} = C \star D^{-1}$

Analogy Induced on Monoids and Groups (Lepage, 2018b)

The analogy induced by the structure of the commutative groups
$(\mathcal{P}(\mathcal{E}), \triangle)$ or $(\mathcal{P}(\mathcal{E}), \doteq)$
is the same as
the analogy induced by the two monoids $(\mathcal{P}(\mathcal{E}), \cap)$ and $(\mathcal{P}(\mathcal{E}), \cup)$
holding at the same time
under the condition that $A \subset B \cup C$ and $A \supset B \cap C$.

$$A : B \overset{\triangle}{::} C : D \Leftrightarrow$$

$$\begin{aligned} A : B \overset{\doteq}{::} C : D &\Leftrightarrow A \triangle B = C \triangle D \\ &\Leftrightarrow (A \setminus B) \cup (B \setminus A) = (C \setminus D) \cup (C \setminus D) \\ &\Leftrightarrow (A \cap D) = (C \cap B) \wedge (A \cup D) = (C \cup B) \\ &\Leftrightarrow A : B \overset{\cap}{::} C : D \wedge A : B \overset{\cup}{::} C : D \end{aligned}$$

Same thing for $\mathbb{B}$ with $\wedge$, $\vee$, $\Leftrightarrow$, etc.

Analogy on Algebraic Structures – The Case for Strings

set	**internal operation**	associativity	zero elt.	**monoid?**	inverse elt for a	**group?**	commutativity
$(\mathbb{N},$	$+)$	yes	0	yes			yes
$(\mathbb{N} \setminus \{0\},$	$\times)$	yes	1	yes			yes
$(\mathbb{Z},$	$+)$	yes	0	yes	$-a$	yes	yes
$(\mathbb{Q} \setminus \{0\},$	$\times)$	yes	1	yes	$\frac{1}{a}$	yes	yes
$(\mathbb{R},$	$+)$	yes	0	yes	$-a$	yes	yes
$(\mathbb{R} \setminus \{0\},$	$\times)$	yes	1	yes	$\frac{1}{a}$	yes	yes
$(\mathbb{C},$	$+)$	yes	0	yes	$-a$	yes	yes
$(\mathbb{C} \setminus \{0\},$	$\times)$	yes	1	yes	$\frac{1}{a}$	yes	yes

Analogy on Algebraic Structures – The Case for Strings

set	**internal operation**	associativity	zero elt.	**monoid?**	inverse elt for a	**group?**	commutativity
$(\mathbb{B},$	$\wedge)$	yes	V	yes			yes
$(\mathbb{B},$	$\vee)$	yes	F	yes			yes
$(\mathbb{B},$	$\oplus)$	yes	F	yes	a	yes	yes
$(\mathbb{B},$	$\Leftrightarrow)$	yes	V	yes	a	yes	yes
$(\mathcal{P}(\mathcal{E}),$	$\cap)$	yes	$\mathcal{E}$	yes			yes
$(\mathcal{P}(\mathcal{E}),$	$\cup)$	yes	$\emptyset$	yes			yes
$(\mathcal{P}(\mathcal{E}),$	$\triangle)$	yes	$\emptyset$	yes	a	yes	yes
$(\mathcal{P}(\mathcal{E}),$	$\doteq)$	yes	$\mathcal{E}$	yes	a	yes	yes
$(\mathcal{V}^*,$	$.)$	yes	ε	yes			

Axioms and Postulates for Analogy on Strings (Lepage, 2003, 2014b)

Number	Axiom name		Property
(iv)	Reflexivity of conformity	$A : B :: A : B$	is always true
(i)	Symmetry of conformity	$A : B :: C : D \Rightarrow$	$C : D :: A : B$
(ii)	Inversion of ratios		$B : A :: D : C$
(iii)	Inversion of objects (= centrality)		$A^{-1} : B^{-1} :: C^{-1} : B^{-1}$
(iv)	Distribution over objects (= similarity)		any feature in A is present in B, in C, or in both
(v)	Exchange of the means	$A : B :: C : D \Rightarrow$	$A : C :: B : D$

Axioms and Postulates for Analogy on Strings (Lepage, 2003, 2014b)

Number	Axiom name	Property
(0)	Reflexivity of *conformity*	$A : B :: A : B$ is always true.
(I)	Symmetry of *conformity*	$A : B :: C : D \Rightarrow C : D :: A : B$
(II)	Inversion of *ratios*	$B : A :: D : C$
(III)	Inversion of *objects* (= contiguity)	$A^{-1} : B^{-1} :: C^{-1} : D^{-1}$
(IV)	Distribution over *objects* (= similarity)	$\exists x$: feature in A is present in B, in C, or in both
(V)	Exchange of the *means*	$A : C :: B : D$

Axioms and Postulates for Analogy on Strings (Lepage, 2003, 2014b)

Number	Axiom name		Property
(0)	Reflexivity of *conformity*	$A : B :: A : B$	is always true
(I)	Inversion of *conformity*	$A : B :: C : D \Rightarrow$	$C : D :: A : B$
(II)	Inversion of *ratios*	"	$B : A :: D : C$
(III)	Inversion of *objects*	"	$A^{-1} : B^{-1} :: C^{-1} : D^{-1}$
(V)	Distribution over *Objects* (= antisymmetry)	"	any feature in A is present in B, in C, or in both
(X)	Exchange of the means	"	$A \quad C \quad B \quad D$

Axioms and Postulates for Analogy on Strings (Lepage, 2003, 2014b)

Number	Axiom name				Property
(0)	Reflexivity of conformity		$A : B :: A : B$	is always true	
(I)	Inversion of conformity		$A : B :: C : D$	$\Rightarrow$	$C : D :: A : B$
(II)	Inversion of ratios		$A : B :: C : D$	$\Rightarrow$	$B : A :: D : C$
(III)	Inversion of *objects*	(contiguity)	$A : B :: C : D$	$\Rightarrow$	$A^{-1} : B^{-1} :: C^{-1} : D^{-1}$
(IV)	Distribution over *objects*	(similarity)	"		any feature in A is present in B, in C, or in both.
(0)	Exchange of the means		$A : C :: B : D$	$\Rightarrow$	$A : B :: C : D$

Axioms and Postulates for Analogy on Strings (Lepage, 2003, 2014b)

Number	Axiom name	Property
(O)	Reflexivity of *conformity*	$A : B :: A : B$ is always true.
(I)	Inversion of *conformity*	$A : B :: C : D \Rightarrow C : D :: A : B$
(II)	Inversion of *ratios*	" $B : A :: D : C$
(III)	Inversion of *objects* (contiguity)	" $A^{-1} : B^{-1} :: C^{-1} : D^{-1}$
(IV)	Distribution over *objects* (similarity)	" any feature in A is present in B, in C, or in both.
(V)	Exchange of the means	" $A : C :: B : D$

Axioms and Postulates for Analogy on Strings (Lepage, 2003, 2014b)

Number	Axiom name		Property
(0)	Reflexivity of *conformity*	$A : B :: A : B$	is always true.
(I)	Inversion of *conformity*	$A : B :: C : D \Rightarrow$	$C : D :: A : B$ (redundant with(II))
(II)	Inversion of *ratios*	"	$B : A :: D : C$ (redundant with(I))
(III)	Inversion of *objects* (contiguity)	"	$A^{-1} : B^{-1} :: C^{-1} : D^{-1}$
(IV)	Distribution over *objects* (similarity)	"	any feature in A is present in B, in C, or in both.
(V)	Exchange of the means	"	$A : C :: B : D$

Analogies on Strings: A Proposal

Analogies on Strings: A Proposal

A Tentative Definition of Analogy on Strings

Distribution and Edit Distances (Lepage, 2001, 2004a)

Definition by distances:

- Definition of ratio as difference of Parikh vectors and LCS distance
- The LCS distance is induced by the axiom of distribution:
  The characters in A should appear in either B or C, in the same order
  $|A| \leq \mathrm{s}(A,B) + \mathrm{s}(A,C)$ and $\mathrm{d}(A,B) = |A| + |B| - 2 \times \mathrm{s}(A,B)$

$$A : B :: C : D \quad \Rightarrow \quad \begin{cases} \mid A \mid_a - \mid B \mid_a = \mid C \mid_a - \mid D \mid_a, \forall a \\ \mathrm{d}(A,B) = \mathrm{d}(C,D) \\ \mathrm{d}(A,C) = \mathrm{d}(B,D) \end{cases}$$

Distribution and Edit Distances (Lepage, 2001, 2004a)

Definition by distances:

- Definition of ratio as difference of Parikh vectors and LCS distance
- The LCS distance is induced by the axiom of distribution:
  The characters in A should appear in either B or C, in the same order
  $|A| \leq \mathrm{s}(A,B) + \mathrm{s}(A,C)$ and $\mathrm{d}(A,B) = |A| + |B| - 2 \times \mathrm{s}(A,B)$

$$A : B :: C : D \quad \Rightarrow \quad \begin{cases} \mid A \mid_a - \mid B \mid_a = \mid C \mid_a - \mid D \mid_a, \forall a \\ \mathrm{d}(A,B) = \mathrm{d}(C,D) \\ \mathrm{d}(A,C) = \mathrm{d}(B,D) \end{cases}$$

Definition by shuffle (Yvon, 2003):

$$A : B :: C : D \quad \Leftrightarrow \quad A \bullet D \cap B \bullet C \neq \emptyset$$

Definition by algorithmic complexity (Murena et al., 2020):

$$A : B :: C : D \quad \Leftarrow \quad D = \left(p = \min_{\mathrm{prog}/B=\mathrm{prog}(A)}\right)(C)$$

String Transformations Preserving Analogy (Lepage, 2018d)

	Theorems	Verif.
Max-length casting	␣*putra* : *putera* :: ␣*putri* : *puteri* *putra*␣ : *putera* :: *putri*␣ : *puteri*	100 % 100 %
Mirroring	*artup* : *aretup* :: *irtup* : *iretup*	100 %
Systematic reduplication	*ppuuttrraa* : *ppuutteerraa* :: *ppuuttrrii* : *ppuutteerrii*	100 %
Concaternation with opposite term's mirror	*putrairetup* : *puterairtup* :: *putriaretup* : *puteriartup*	100 %
	Conjecture	
Systematic insertion	*p␣u␣t␣r␣a␣* : *p␣u␣t␣e␣r␣a␣* :: *p␣u␣t␣r␣i␣* : *p␣u␣t␣e␣r␣i␣*	100 %
	Refutations	
Concatenation with own mirror	*putraartup* : *puteraaretup* :: *putriirtup* : *puteriiretup*	98 %
Repetition	*putraputra* : *puteraputera* :: *putriputri* : *puteriputeri*	90 %

Analogies on Strings: A Proposal

Implementation

Analogical Clusters and Analogical Grids (Lepage, 2012, 2014a; Fam and Lepage, 2017)

Analogical clusters

$$\begin{array}{l} A_1 : B_1 \\ A_2 : B_2 \\ \quad\vdots \\ A_n : B_n \end{array} \overset{\Delta}{\Longleftrightarrow} \forall (i,j) \in \{1,\ldots,n\}^2,\ \ A_i : B_i :: A_j : B_j$$

Analogical grids

$$\begin{array}{ccccc} P_1^1 & : & P_1^2 & : \cdots : & P_1^m \\ P_2^1 & : & P_2^2 & : \cdots : & P_2^m \\ \vdots & & \vdots & & \vdots \\ P_n^1 & : & P_n^2 & : \cdots : & P_n^m \end{array} \overset{\Delta}{\Longleftrightarrow} \begin{array}{l} \forall (i,j) \in \{1,\ldots,n\}^2, \\ \quad \forall (k,l) \in \{1,\ldots,m\}^2, \\ \quad\quad P_i^k : P_i^l :: P_j^k : P_j^l \end{array}$$

The Nlg Package (Fam and Lepage, 2018)

- Python module for analogy, analogical clusters and analogical grids

Function Name	Functionality	Input	Output
Words2Clusters	Extraction	words	analogical clusters
Words2Grids	Construction	words	analogical grids
Words2Vectors	Building	words	words + their vectors
Vectors2Clusters	Extraction	words + their vectors	analogical clusters
Vectors2Grids	Construction	words + their vectors	analogical grids
Clusters2Grids	Construction	analogical clusters	analogical grids

- Versions:

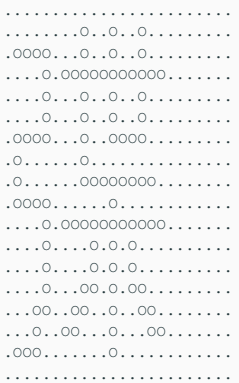
Version Number	Date	Comment
v1.0.0	2018-05-07	first version in Python2.7
v3.0.0	2020-04-14	Python3 version
v3.1.0	2021-04-30	scripts for vector input
v3.2.0	2021-09-08	better clarity on clustering function
v3.2.1	2021-12-20	NumPy for faster matrix manipulation and multiprocessing for faster computation

- Creative Commons Attribution-NonCommercial 4.0 International License.

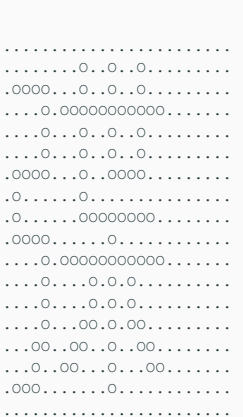
Analogies on Strings: A Proposal

Resources

Applications on Bitmaps (Lepage, 2012, 2014a)

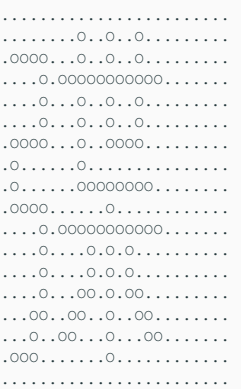


Applications on Bitmaps (Lepage, 2012, 2014a)



$\rightarrow \quad \overrightarrow{\text{碟}} = \begin{pmatrix} 0 \\ 3 \\ 7 \\ 12 \\ 4 \\ 4 \\ 9 \\ 2 \\ \vdots \\ 3 \\ 0 \end{pmatrix}$

Applications on Bitmaps (Lepage, 2012, 2014a)



$\rightarrow \quad \overrightarrow{\text{蝶}} = \begin{pmatrix} 0 \\ 3 \\ 7 \\ 12 \\ 4 \\ 4 \\ 9 \\ 2 \\ \vdots \\ 3 \\ 0 \end{pmatrix}$

of ○ on 1st row

of ○ on 2nd row

of ○ on 3rd row

of ○ on 4th row

of ○ on 5th row

of ○ on 6th row

of ○ on 7th row

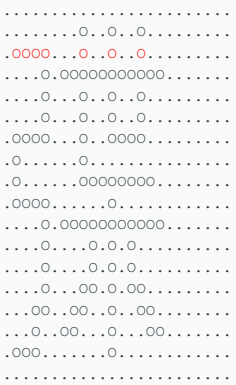
of ○ on 8th row

$\vdots$

of ○ on 23rd col.

of ○ on 24th col.

Applications on Bitmaps (Lepage, 2012, 2014a)



$\rightarrow \quad \overrightarrow{\text{碟}} = \begin{pmatrix} 0 \\ 3 \\ 7 \\ 12 \\ 4 \\ 4 \\ 9 \\ 2 \\ \vdots \\ 3 \\ 0 \end{pmatrix}$

of ○ on 1st row

of ○ on 2nd row

of ○ on 3rd row

of ○ on 4th row

of ○ on 5th row

of ○ on 6th row

of ○ on 7th row

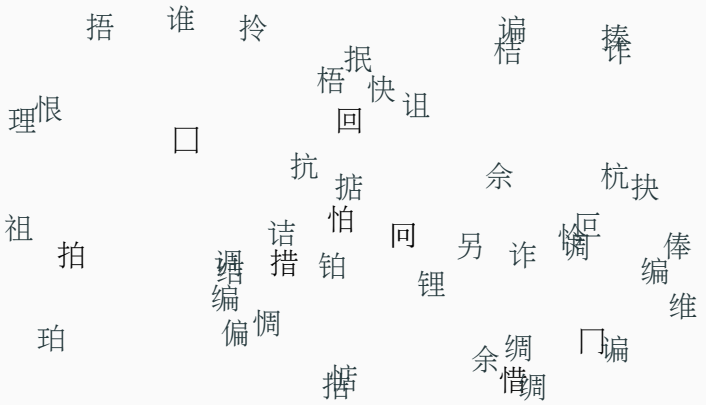
of ○ on 8th row

$\vdots$

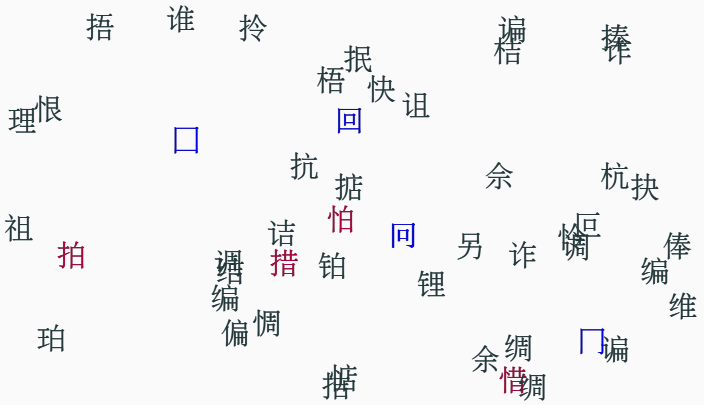
of ○ on 23rd col.

of ○ on 24th col.

Applications on Bitmaps (Lepage, 2012, 2014a)



Applications on Bitmaps (Lepage, 2012, 2014a)



Applications on Bitmaps (Lepage, 2012, 2014a)

怕:拍
惜:措
恨:抿
快:抉
怜:拎
惦:掂
俸:捧

诂:结
调:绸
谝:编
谁:维

捂:梧
抗:杭
拮:桔

偏:惆
谝:调
编:绸

诅:祖
诈:祚

铂:珀
锂:理

冂:问
口:回
匚:匝

:另
佘:余

Overview of Some Resources Produced

Units	Original Resource	Languages	Size	Type	Year
Character bitmaps	font sets	Chinese, Japanese		clusters	2014
Word forms	Gospel of St Matthew, St Luke	12 world languages, 4 continents (so, sw, xh, acu, nah, qu, zh, id, te, en, fr, el)		grids	2016
Word forms	SIGMORPHON	10 world languages, 3 continents (ar, de, es, fi, hu, ka, mt, nv, ru, tr)	1.2 million	analogies	2017
Word forms	BPPT	Indonesian		grids	2019
N-grams (1 to 6)	Europarl	11 European languages (da, de, el, en, es, it, fi, fr, nl, pt, sv)		clusters, grids	2018
Chunks, Sentences	BTEC	Japanese, English		analogies	2011
Sentences	BTEC	Chinese, Japanese, English	1.5∼2 million	analogies	2004
Sentences (char, word, BPE, etc.)	Tatoeba, Multi30k, CommonCrawl	5 languages		analogical density	2021
Sentences	Tatoeba	2 languages (en, fr)	1.2 million	formal aligned analogies	2020
Sentences	”	”	5,600	semantico-formal analogies	2021

Word Forms (Lepage, 2017)

Test set of 9.2 million analogies between word forms in 10 different languages extracted from the SIGMORPHON 2016 Shared Task: Morphological Reinflection.

Finnish	*jdekoodaamaisillaan* : *kakomaisillaan* :: *dekoodata* : *katko*
German	*abgehärtet* : *abgünstig* :: *gehärtet* : *günstig*
Hungarian	*lógtál* : *abbahagytál* :: *lóg* : *abbahagy*
Maltese	*hóóshkeed* : *jijoozba'* :: *dahóóshkeed* : *dajijoozba'*
Navajo	*niłdlaad* : *wołdlaad* :: *niłjooł* : *woljooł*
Spanish	*mutaban* : *muten* :: *apatrullaba* : *apatrulle*
Turkish	*yutmayacaklar* : *yutmak* :: *doymayacaklar* : *doymak*

Comparison of Proposals for Analogy on Strings (Murena et al., 2020)

Language	# analogies	distance	shuffle	complexity
Arabic	165,113	93.33 %	81.91 %	87.18 %
Finnish	313,011	92.76 %	78.75 %	93.69 %
Georgian	3,066,273	97.54 %	88.42 %	99.35 %
German	730,427	96.21 %	95.42 %	98.84 %
Hungarian	2,912,310	92.61 %	86.02 %	95.71 %
Maltese	28,365	84.72 %	91.84 %	96.38 %
Navajo	321,473	86.87 %	78.95 %	81.21 %
Russian	552,423	97.26 %	95.46 %	96.41 %
Spanish	845,996	96.13 %	94.42 %	96.73 %
Turkish	245,721	69.97 %	70.06 %	89.45 %
Total	9,181,112	94.34 %	87.93 %	96.41 %

Chunks (Lepage et al., 2007a,b; Takeya and Lepage, 2011b,a)

my insurance company : *your insurance company* :: *my passport and ticket* : *your passport and ticket*

a dollar bill : *five thousand yen* :: *a hundred dollar bill* : *five hundred thousand yen*

a snack bar : *the snack bar* :: *a pina colada* : *the pina colada*

any free rooms : *some free time* :: *any other rooms* : *some other time*

any concert tonight and tomorrow night : *any concert**s** tonight and tomorrow night* $\neq$ *this Mr. Ono* : *this Mr**s**. Ono*

N-grams (Fam and Lepage, 2018)

needs of the	:	*directives on the*
the needs of	:	*the directives on*
of this need	:	*on this directives*
aware of this	:	*to extend this*
aware of the	:	*to extend the*
is aware of	:	*is to extend*

by the commission ,	:	*by both the commission*
the member states ,	:	*both the member states*
1993 and 2003 ,	:	*both 1993 and 2003*
, in line with	:	*in line with both*
, to remind us	:	*to remind us both*
in favour of ,	:	*both in favour of*
legally and illegally ,	:	*both legally and illegally*

Sentences (Lepage, 2004b; Yang et al., 2014)

i 'd like to	*cash this traveler 's check . :*	*can i cash this traveler 's check here ?*
i 'd like to	*make a hotel reservation . :*	*can i make a hotel reservation here ?*
i 'd like to	*make a reservation . :*	*can i make a reservation here ?*
i 'd like to	*check my baggage . :*	*can i check my baggage here ?*
i 'd like to	*leave my baggage . :*	*can i leave my baggage here ?*
i 'd like to	*leave my luggage . :*	*can i leave my luggage here ?*
i 'd like to	*reserve a room . :*	*can i reserve a room here ?*
i 'd like to	*have dinner . :*	*can i have dinner here ?*
i 'd like to	*check in . :*	*can i check in here ?*
i 'd like to	*swim . :*	*can i swim here ?*

is there a subway station	*around here ? :*	*is there a youth hostel around here ?*
how can i get to the subway	*station ? :*	*how can i get to the youth hostel ?*
is there a subway station	*near here ? :*	*is there a youth hostel near here ?*
is there a subway station	*nearby ? :*	*is there a youth hostel nearby ?*

could you keep this baggage	*? :*	*could you draw me a map ?*
keep this baggage , please	*. :*	*draw me a map , please .*
will you keep this baggage	*? :*	*will you draw me a map ?*
please keep this baggage	*. :*	*please draw me a map .*

Native speaker's judgement: more than 96% of the analogies are semantically valid.

Sentences – Form and Meaning (Lepage, 2019; Wang and Lepage, 2020)

There's hardly any coffee left in the pot . : *There's almost no coffee left in the pot .* :: *There's hardly any water in the bucket .* : x $\Rightarrow$ $x =$ *There's almost no water in the bucket .*

I do not know what to say about that . : *I do not know what to do now .* :: *I do not know about that .* : x $\Rightarrow$ $x =$ *I do not think so .*

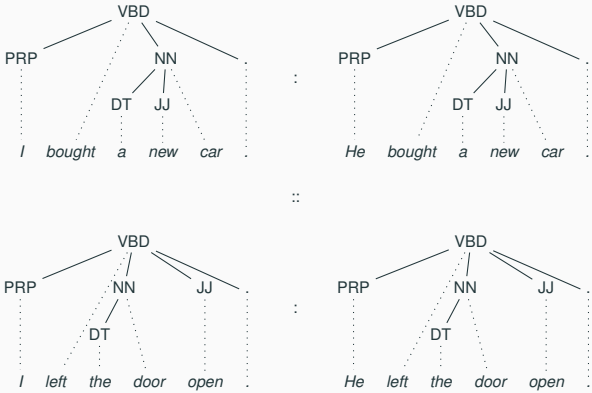
You 're not from around here , are you ? : *You 're not staying here , are you ?* :: *You 're confused again , are n't you ?* : x $\Rightarrow$ $x =$ *You 're disappointed , are n't you ?*

Je veux ce que vous voulez . : *Je veux ce que tu veux .* :: *Je veux vraiment que vous compreniez .* : x $\Rightarrow$ $x =$ *Je veux vraiment que tu comprennes .*

Je ne suis pas si veinard . : *Je ne suis pas si chanceuse .* :: *Je ne suis pas ton ami .* : x $\Rightarrow$ $x =$ *Je ne suis pas ton amie .*

Je veux ce que vous voulez . : *Je veux ce que tu veux .* :: *Je sais ce que vous faisiez .* : x $\Rightarrow$ $x =$ *Je sais ce que tu faisais .*

Syntactic Trees (Lepage and Auclerc, 2000; Stroppa, 2005; Zhou et al., 2022)



Conclusion

Conclusion and Summary

- Survey of the history of the concept in grammar and linguistics
  $\Rightarrow$ one domain, objects of the same type (strings)
  But, no inverse, no commutativity of concatenation.

Conclusion and Summary

- Survey of the history of the concept in grammar and linguistics
  $\Rightarrow$ one domain, objects of the same type (strings)
  But, no inverse, no commutativity of concatenation.
- Production of analogy sets
  at various granularity, in various languages, aligned or not
  $\Rightarrow$ Applications in morphology, parsing, example-based or statistical machine translation

Conclusion and Summary

- Survey of the history of the concept in grammar and linguistics
  $\Rightarrow$ one domain, objects of the same type (strings)
  But, no inverse, no commutativity of concatenation.
- Production of analogy sets
  at various granularity, in various languages, aligned or not
  $\Rightarrow$ Applications in morphology, parsing, example-based or statistical machine translation
- Come back of continuous analogies between sentences:
  computation of the middle of two sentences
  $\Rightarrow$ Application to text morphing, data augmentation for machine translation

Acknowledgments

Work supported by JSPS Grant 21K12038 (Kakenhi C)
“Theoretically funded algorithms for the automatic production
of analogy test sets in NLP”.

Thank you.

Publications / The Nlg Package

References

Douay, F. (1985). La contre-analogie – Réflexion sur la récusation de certaines analogies pourtant bien formées cognitivement. *texto*, 10:155–170.

Douay, F. and Pinto, J.-J. (1991). Analogie / anomalie – Reflet de nos querelles dans un miroir antique. *Communications*, 53:7–16.

Fam, R. and Lepage, Y. (2017). A study of the saturation of analogical grids agnostically extracted from texts. In *Proceedings of the Computational Analogy Workshop at the 24th International Conference on Case-Based Reasoning (ICCBR-17)*, pages 7–16, Trondheim, Norway.

Fam, R. and Lepage, Y. (2018). Tools for the production of analogical grids and a resource of n-gram analogical grids in 11 languages. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, pages 1060–1066, Miyazaki, Japan. ELRA.

Lambert, R. (1996). L'observation des ressemblances, d'après Aristote. *Laval théologique et philosophique*, 22(2):169–185.

Lepage, Y. (2001). Analogy and formal languages. In *Proceedings of FG/MOL 2001*, pages 1–12, Helsinki.

Lepage, Y. (2003). *De l'analogie rendant compte de la commutation en linguistique (French) (Of that kind of analogies capturing linguistic commutations)*. Habilitation thesis, Joseph Fourier Grenoble University.

Lepage, Y. (2004a). Analogy and formal languages. *Electronic notes in theoretical computer science*, 53:180–191.

Lepage, Y. (2004b). Lower and higher estimates of the number of "true analogies" between sentences contained in a large multilingual corpus. In *Proceedings of COLING-2004*, volume 1, pages 736–742, Genève.

Lepage, Y. (2012). (Re)-discovering the graphical structure of Chinese characters. In Richard, G., editor, *Working papers of the 1st international workshop on similarity and analogy-based methods in AI (SAMAI 2012)*, IRIT/RR–2012-20—FR, pages 57–64, Montpellier. IRIT.

Bibliography II

Lepage, Y. (2014a). Analogy between binary images: application to Chinese characters. In Prade, H. and Richard, G., editors, *Computational Approaches to Analogical Reasoning: Current Trends*, Studies in Computational Intelligence 548, pages 25–57. Springer Verlag, Berlin Heidelberg.

Lepage, Y. (2014b). Proportional analogy in written language data. In Gala, N., Rapp, R., and Bel-Enguix, G., editors, *Language, Production, Cognition and the Lexicon*, volume 48 of *Text, Speech and Language Technology*, pages 151–173. Springer International Publishing Switzerland.

Lepage, Y. (2017). Character–position arithmetic for analogy questions between word forms. In *Proceedings of the Computational Analogy Workshop at the 24th International Conference on Case-Based Reasoning (ICCBR-17)*, pages 17–26, Trondheim, Norway.

Lepage, Y. (2018a). Cahiers d'analogie (1). Master's degree lectures at IPS, Waseda University.

Lepage, Y. (2018b). Cahiers d'analogie (2). Master's degree lectures at IPS, Waseda University.

Lepage, Y. (2018c). Cahiers d'analogie ($2 - \epsilon$). Master's degree lectures at IPS, Waseda University.

Lepage, Y. (2018d). String transformations preserving analogies. In *Proceedings of the 2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS 2018)*, pages 189–194, Yogyakarta, Indonesia.

Lepage, Y. (2019). Semantico-formal resolution of analogies between sentences. In Vetulani, Z. and Paroubek, P., editors, *Proceedings of the 9th Language & Technology Conference (LTC 2019) – Human Language Technologies as a Challenge for Computer Science and Linguistics*, pages 57–61.

Lepage, Y. and Auclerc, N. (2000). A tool to build a tree bank for conversational Chinese. In *Proceedings of ICSLP 2000*, volume III, pages 985–988, Beijing.

Lepage, Y. and Denoual, E. (2005). Purest ever example-based machine translation: detailed presentation and assessment. *Machine Translation*, 19:251–282.

Lepage, Y., Migeot, J., and Guillerm, E. (2007a). Analogies of form between chunks in Japanese are massive and far from being misleading. In *Proceedings of the 3rd Language and Technology Conference (LTC 2007)*, pages 503–507, Poznań, Poland.

Bibliography III

Lepage, Y., Migeot, J., and Guillerm, E. (2007b). A corpus study on the number of true proportional analogies between chunks in two typologically different languages. In *Proceedings of the seventh international Symposium on Natural Language Processing (SNLP 2007)*, pages 117–122, Pattaya, Thailand. Kasetsart University, ISBN 978-974-623-062-9.

Levy, O. and Goldberg, Y. (2014). Linguistic regularities in sparse and explicit word representations. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*, pages 171–180, Ann Arbor, Michigan. Association for Computational Linguistics.

Michel, P.-H. (1949). Les médiétés. *Revue d'histoire des sciences et de leurs applications*, 2(2):139–178.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.

Mikolov, T., Yih, W.-T., and Zweig, G. (2013b). Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2013)*, pages 746–751, Atlanta, Georgia. Association for Computational Linguistics.

Murena, P.-A., Al-Ghossein, M., Dessalles, J.-L., and Cornuéjols, A. (2020). Solving analogies on words based on minimal complexity transformation. In Bessière, C., editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 1848–1854. International Joint Conferences on Artificial Intelligence Organization. Main track.

Nagao, M. (1984). A framework of a mechanical translation between Japanese and English by analogy principle. In Elithorn, A. and Banerji, R., editors, *Proceedings of the international NATO symposium on Artificial and human intelligence*, pages 173–180. Elsevier Science Publishers, NATO.

Schironi, F. (2007). *Bilinguisme et terminologie grammaticale gréco-latine*, volume 27, chapter Analogia (Greek), analogia, proportio, ratio: loanwords, calques, and reinterpretation of a Greek technical term, pages 321–338. Peeters, Leuven – Paris – Dudley, MA.

Silva de Carvalho, A. C. (2013). O conceito de analogia sob a ótico de Marco Terêncio Varrão. *Estudos linguísticos*, 42(3):1244–1253.

Stroppa, N. (2005). *Définitions et caractérisation de modèles à base d'analogies pour l'apprentissage automatique des langues naturelles*. Thèse de doctorat, École nationale supérieure des télécommunications.

Bibliography IV

Takeya, K. and Lepage, Y. (2011a). Fully-automatic marker-based chunking in 11 European languages and counts of the number of analogies between chunks. In *Proceedings of the 25th Pacific Asia Conference on Language Information and Computing (PACLIC 25)*, pages 567–576, Singapore.

Takeya, K. and Lepage, Y. (2011b). A study of the number of proportional analogies between marker-based chunks in 11 European languages. In Vetulani, Z., editor, *Proceedings of the 5th Language & Technology Conference (LTC'11)*, pages 284–288, Poznań. Fundacja uniwersytetu im. Adama Mickiewicza.

Turney, P. (2008a). A uniform approach to analogies, synonyms, antonyms, and associations. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 905–912, Manchester, UK. Coling 2008 Organizing Committee.

Turney, P. and Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37:141–188.

Turney, P. D. (2006). Similarity of semantic relations. *Computational Linguistics*, 32(2):379–416.

Turney, P. D. (2008b). The latent relation mapping engine: Algorithm and experiments. *J. Artif. Int. Res.*, 33(1):615–655.

Turney, P. D. and Littman, M. L. (2005). Corpus-based learning of analogies and semantic relations. *Machine Learning*, 60(1–3):251–278.

Vitrac, B. (1996). La définition V. 8 des Eléments d'Euclide. *Centaurus*, 38(2–3):97–121.

Wang, L. and Lepage, Y. (2020). Vector-to-sequence models for sentence analogies. In IEEE, editor, *Proceedings of the 2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS 2020)*, pages 441–446.

Welcomme, A. (2010). Hermann Paul et le concept d'analogie. *CÍRCULO de Lingüística Aplicada a la Comunicación (clac)*, 43:49–122.

Yang, W., Wang, H., and Lepage, Y. (2014). Deduction of translation relations between new short sentences in Chinese and Japanese using analogical associations. *International Journal of Advanced Intelligence*, 6(1):13–34.

Yvon, F. (2003). Analogy-based NLP: implementation issues. Technical report, ENST.

Zhou, Y., Fam, R., and Lepage, Y. (2022). Extraction of analogies between sentences on the level of syntax using parse trees. In *Proceedings of the Workshop on Analogies: from Theories to Applications at ICCBR 2022*.