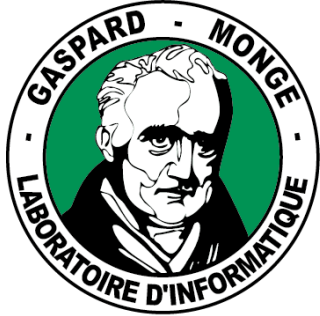




**LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE**

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

20 mars 2019

ATILF

Université de Lorraine

Le continuum de compositionnalité et la description linguistique formelle

Eric Laporte
Fabrice Beltran



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Sommaire

Variabilité syntaxique

Description formelle

Mesures

Résultats



Le continuum de compositionnalité

Merci au projet PARSEME-FR (ANR) pour son soutien

Merci à Takuya Nakamura (Université Paris-Est) pour sa contribution

Il existe un continuum de compositionnalité entre
expressions polylexicales et expressions libres

*Ils ont **avancé un pion** sur l'échiquier*

lit. advance a pawn on the chessboard

"make a move to advance one's interest"

Ils ont avancé de l'argent à un inconnu

Différences dans les constructions syntaxiques

Qu'est-ce qu'ils ont **avancé sur l'échiquier ?*

Combien ont-ils avancé à l'inconnu ?

Où ont-ils ont **avancé un pion ?*

À qui ont-ils ont avancé de l'argent ?



Variabilité syntaxique d'expressions verbales figées

Une des principales raisons de considérer qu'il existe un continuum de compositionnalité est la variabilité syntaxique

*Le joueur a **enlevé une épine du pied de** ses présidents*

lit. remove a thorn from sb's foot

"pull a thorn from sb's side"

Passif

*Tant mieux, **une bonne épine enlevée du pied** !*

Pronominalisation

*Voilà qui nous **enlève une épine du pied** !*



Certaines opérations syntaxiques ne s'appliquent pas à toutes les expressions

Deux opérations

lit. strain the note
"overdo it"

*Il a même **forcé la note** en les offrant à son demi-frère*

Passif : ***La note** est un peu **forcée***

Omission: *On a un peu **forcé** en faisant fonctionner ce sigle*

*On a un peu crié, mais on a **crevé l'abcès***

lit. lance an abscess
"clear the air"

Passif : *Je suis contente ! Comme ça, **l'abcès est crevé***

Omission : *≠ On a un peu crié, mais on a **crevé***

Certaines opérations syntaxiques ne s'appliquent pas à
toutes les expressions



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Sommaire

Variabilité syntaxique

Description formelle

Mesures

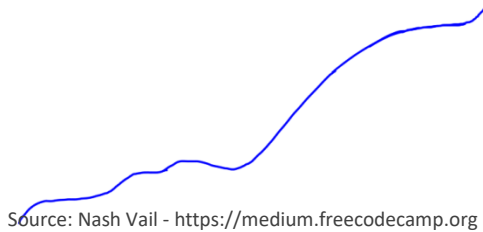
Résultats



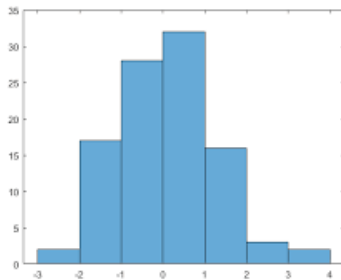
LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Description formelle



Plusieurs façons d'aborder la description formelle
Modèles flous, modèles discrets



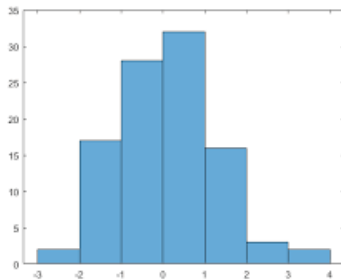
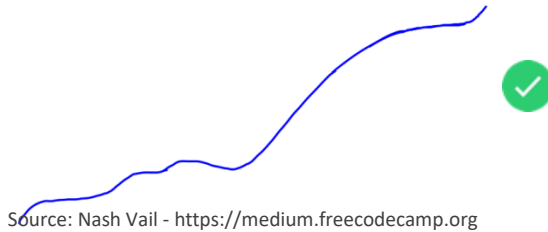
Source: Matlab documentation



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Modèles flous



Source: Matlab documentation

En linguistique

Psycholinguistique : données floues, par exemple préférences

Linguistique de corpus: rare/fréquent

Continuum de compositionnalité

En traitement des langues

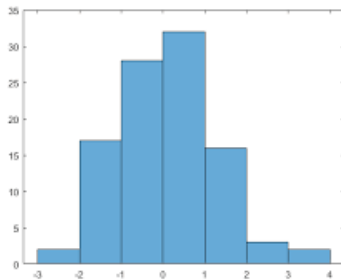
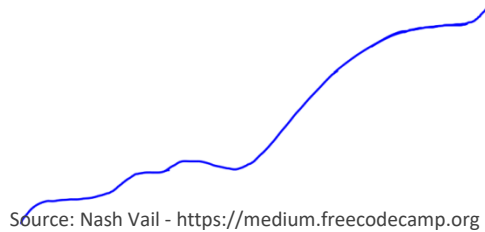
Calculs statistiques

Compositionnalité : affinité statistique, par exemple *nuages et éclaircies*

Variabilité syntaxique : flexibilité syntaxique



Modèles discrets



Source: Matlab documentation



Catégories à frontières nettes

En traitement des langues

Calcul symbolique : étiquettes de propriétés, de sens...

Analyseurs syntaxiques symboliques

Compositionnalité : si l'expression peut être traitée sur la base de plusieurs entrées lexicales

envisager l'avenir avec optimisme

vs. enfouir son visage dans ses mains

En linguistique

Une phrase est acceptée ou rejetée (cf. linguistique structurale)

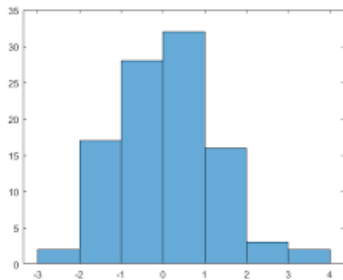
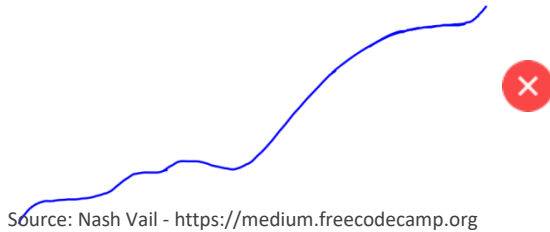
Entrées lexicales



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Description formelle



Source: Matlab documentation

Les analyseurs syntaxiques symboliques
nécessitent une description symbolique dans
un modèle discret
C'est le cadre dans lequel nous nous situons pour
cette étude



Dimensions de la flexibilité syntaxique

Dimensions de la flexibilité syntaxique

- passif
- omissions
- pronominalisations
- etc.

La flexibilité syntaxique est une propriété multidimensionnelle

Peut-on la représenter par une propriété unidimensionnelle ?

Dans les expressions avec beaucoup de parties figées, la plupart des opérations semblent plus difficilement applicables :

se demander si c'est du lard ou du cochon

lit. wonder if it's bacon or pork

"wonder where one stands"

Si une opération s'applique facilement à une expression, les autres s'y appliquent-elles facilement aussi ?

Est-ce une base pour une notion de continuum unidimensionnel de compositionnalité ?



flexibilité totale

aucune flexibilité



Dimensions de la flexibilité syntaxique

Dimensions de la flexibilité syntaxique

- passif
- omissions
- pronominalisations
- etc.

Les dimensions sont-elles corrélées ?

Fraser (1970) : pas d'allégation particulière

Gross (1984) : pas d'allégation particulière

Nunberg *et al.* (1994) : faiblement observations sporadiques

Sag *et al.* (2002) : absolument pas d'observations

Baldwin & Kim (2010) : absolument pas d'observations

Une illusion ?



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Sommaire

Variabilité syntaxique

Description formelle

Mesures

Résultats

Mesure de la corrélation

[illegible]

Un échantillon d'entrées lexicales

Un échantillon d'opérations syntaxiques

Statistiques sur l'applicabilité des opérations

Marquage manuel des propriétés des entrées lexicales

Lexique-grammaire

Source : expressions verbales figées du français (Gross, 1984)

Version finale, inédite, de Gross,
téléchargeable gratuitement sur le site
web du LIGM



Les entrées lexicales

c1i

c1d

c1p2

c1g

c1gpn

5 classes d'expressions françaises

Rejet des collocations non lexicalisées

"look to the future
optimistically"

envisager l'avenir avec optimisme
(compositionnel)

non lexicalisé

"bury one's face in
one's hands"

enfouir son visage dans ses mains
incohérent avec les deux sens de *enfouir*)

lexicalisé (sens

Travail avec Fabrice Beltran et Takuya Nakamura, en cours

Échantillon obtenu

2 129 entrées



Les opérations syntaxiques

Les propriétés marquées dans les données existantes

Exemple: passif dans les 5 classes

Les propriétés implicites déduites de l'appartenance à une classe

Exemple: l'omission de tous les compléments ne s'applique à aucune expression de la classe c1gpn ; non marqué dans les données

Je tiens mes distances avec les fans de foot

lit. hold one's distances with
"keep one's distance from"

≠ *Je tiens*

Documentation des propriétés

Fabrice Beltran (2017)

11 opérations

5 types de pronominalisation

4 types d'omission de compléments

passif

une construction pronominale



Préconditions morphosyntaxiques sur les opérations syntaxiques

*Le joueur a **enlevé une épine du pied de ses présidents***

*Le joueur leur a **enlevé une épine du pied***

La plupart des opérations syntaxiques présupposent des conditions morphosyntaxiques

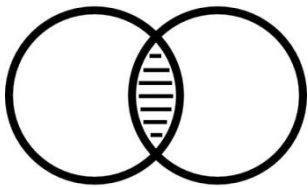
La pronominalisation d'un complément prépositionnel libre présuppose qu'il y a un complément prépositionnel libre

S'il n'y en a pas, d'autres opérations peuvent s'appliquer, par exemple des omissions :

*Ils ont **avancé un pion sur l'échiquier***

lit. advance a pawn on the chessboard

"make a move to advance one's interest"



*Ils ont **avancé un pion***

La corrélation alléguée ne vaut que si les préconditions morphosyntaxiques sur les deux opérations sont réunies



Méthode

Pour chaque paire d'opérations :

Préconditions

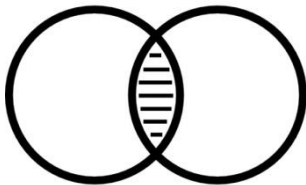
Sélectionner les entrées qui réunissent les préconditions morphosyntaxiques sur les **deux** opérations syntaxiques

Applicabilité effective

Calcul de la corrélation (coefficient de Pearson) entre l'applicabilité effective des deux opérations

Une corrélation positive veut dire que si une des deux opérations est acceptée, l'autre a plus de chances de l'être aussi

Une corrélation négative veut dire que l'autre a moins de chances de l'être aussi



$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$$



Une opération ou deux ?

Omission d'un complément prépositionnel figé ou d'un
complément prépositionnel libre

		Corrélation	Taille de l'échantillon
Considérées comme deux opérations distinctes	passif/omission d'un SP figé	0,61%	933
	passif/omission d'un SP libre	-12 %	246
Fusionnées	passif/omission d'un SP	-0,13%	1179



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Sommaire

Variabilité syntaxique

Description formelle

Mesures

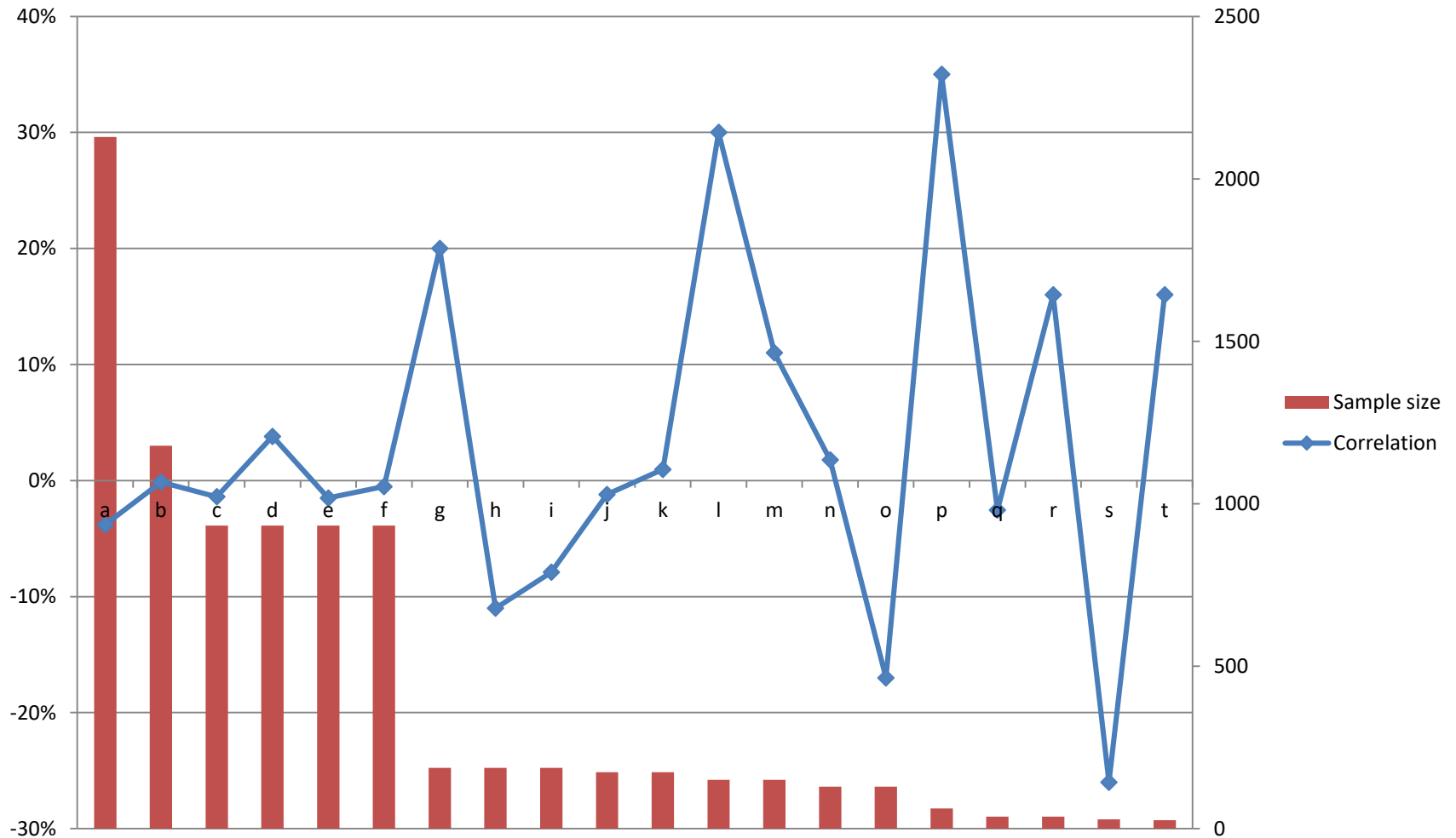
Résultats



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

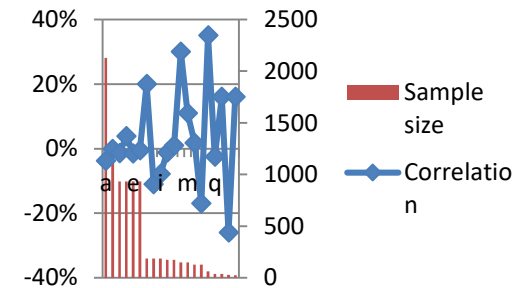
Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Résultats





Résultats



25 paires d'opérations

5 corrélations négatives absolues

Quand une des opérations ne s'applique à aucune entrée dans l'échantillon

Exemple : omission de tous les compléments et pronominalisation dative clitique d'un complément libre

9 corrélations positives

La plus forte : 35 %

Pronominalisation dative clitique de modifieurs du complément direct et du complément prépositionnel

11 corrélations négatives

La plus forte : -26 %

Pronominalisation clitique du complément prépositionnel humain

Pronominalisation adverbiale du complément prépositionnel en *sur*

Moyenne

pondérée par la taille des échantillons : -1.4 %



Significativité statistique

Les échantillons sont peut-être trop petits ou non représentatifs

Échantillons d'entrées lexicales

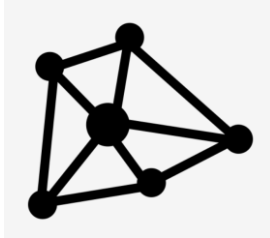
Opérations syntaxiques marquées dans les données disponibles (explicitement ou implicitement)

Valeurs avec les plus grands échantillons

5 des 6 valeurs avec les plus grands échantillons sont faiblement négatives : -3.8% à -0.13%



Conclusion



Un mélange de phénomènes faiblement corrélés

La notion de flexibilité syntaxique en tant que phénomène unidimensionnel ne semble pas cadrer avec un modèle discret

La corrélation semble trop faible pour prédire quelles opérations sont applicables



LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

Sous la co-tutelle de :
CNRS
ÉCOLE DES PONTS PARISTECH
ESIEE PARIS
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Merci

CONTACT

ÉRIC LAPORTE

00 +33 (0)1 60 95 75 52

ERIC.LAPORTE@UNIV-PARIS-EST.FR