Continuum of Compositionality and Formal Description

Éric Laporte
Fabrice Beltran

Outline

Syntactic variation

Formal description

Measurements

Results

Syntactic variation of idioms

One of the main reasons to consider there is a continuum of compositionality is syntactic variation

*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"

Passive

Pronominalization

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

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

Syntactic operations

Two operations

lit. strain the note "overdo it"

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

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

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

lit. lance an abscess "clear the air"

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

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

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

Some syntactic operations don't apply to some idioms

Outline

Fuzzy models

Source: Nash Vail - https://medium.freecodecamp.org

Source: Matlab documentation

In linguistics

Psycholinguistics: fuzzy data, e.g. preference

Corpus linguistics: rare/frequent

Continuum of compositionality

In natural language processing

Statistical computation

Compositionality: statistical affinity, e.g. *back from*

Syntactic variation: syntactic flexibility

Discrete models
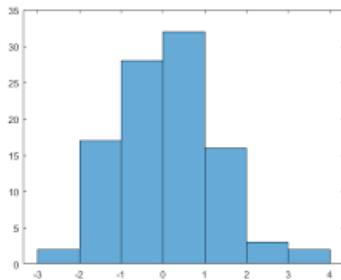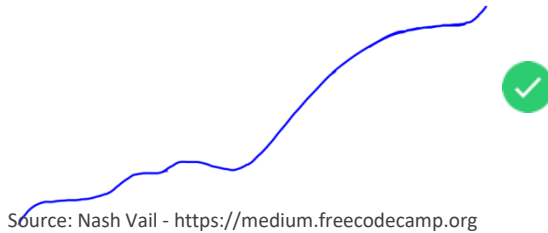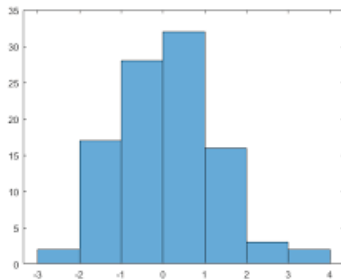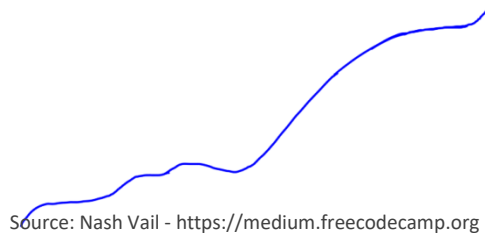
LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

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

Source: Nash Vail - https://medium.freecodecamp.org

Source: Matlab documentation

Categories with definite boundaries

In natural language processing

Symbolic computation: labels for features, senses...

Symbolic parsers

Compositionality: if the expression can be processed on the basis of several lexical entries

envisager l'avenir avec optimisme

vs. *enfouir son visage dans ses mains*

In linguistics

A sentence is accepted or rejected (cf. structural linguistics)
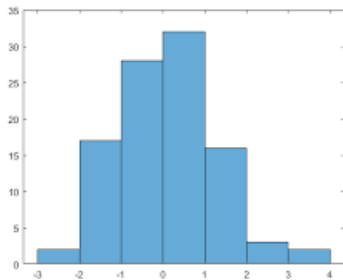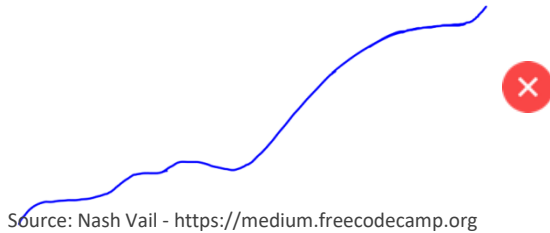
Entries

Formal description

Source: Nash Vail - https://medium.freecodecamp.org

Source: Matlab documentation

Symbolic parsers require formal description with a discrete model

This is our framework in this study

Dimensions of syntactic flexibility

Dimensions of
syntactic flexibility
- passive
- omissions
- pronominalizations
etc.

Is syntactic flexibility a one-dimensional or a
multidimensional feature?

In idioms with more frozen parts, most operations seem less
applicable: *se demander si c'est du lard ou du cochon*

lit. wonder if it's bacon or pork
"wonder where one stands"

If an operation easily applies to an idiom, do others easily
apply to it?

Is this a basis for a notion of a one-dimensional continuum of
compositionality?

fully flexible   fully non-flexible

Dimensions of
syntactic flexibility
- passive
- omissions
- pronominalizations
etc.

How much are the dimensions correlated together?

Fraser (1970): no particular claim

Gross (1984): no particular claim

Nunberg *et al.* (1994): loosely   sporadic evidence

Sag *et al.* (2002): absolutely   no evidence

Baldwin & Kim (2010): absolutely   no evidence

Fake science?

Outline

Measuring correlation

A sample of lexical entries

A sample of syntactic operations

Statistics on applicability of operations

Manual marking of features of lexical entries

Lexicon-Grammar

Source: French verbal idioms (Gross, 1984)

Gross' final, unpublished version, freely downloadable from the LIGM web site

The lexical entries

c1i
c1d
c1p2
c1g
c1gpn

"look to the future
optimistically"

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

5 classes of French idioms

Discard non-lexicalized collocations

envisager l'avenir avec optimisme   non-lexicalized
   (compositional)

enfouir son visage dans ses mains   lexicalized (meaning
   inconsistent with both meanings of *enfouir*)

Work with Fabrice Beltran and Takuya Nakamura, ongoing

Resulting sample

2 129 entries

The syntactic operations

The features marked in the existing data
Example: passive in all 5 classes

The implicit features implied by class membership
Example: omission of all complements doesn't apply to any
 idiom in class c1gpn; not marked in the data

*Je **tiens mes distances avec** les fans de foot*

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

*≠ Je **tiens***

Documentation of the features
Fabrice Beltran (2017)

11 operations
5 types of pronominalization
4 types of omission of complements
passive
a reflexive construction

Morphosyntactic prerequisites for syntactic operations

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

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

lit. advance a pawn on the chessboard
"make a move to advance one's interest"

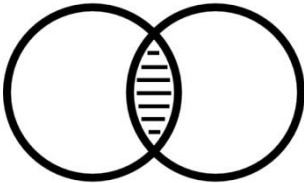Most syntactic operations require morphosyntactic prerequisites

Pronominalization of a free prepositional complement presupposes there is a free prepositional complement

But if there isn't, other operations may apply, e.g. omissions:
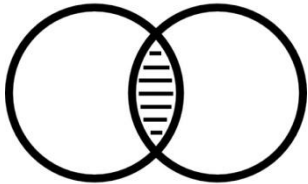
*Ils ont **avancé un pion sur l'échiquier***

*Ils ont **avancé un pion***

To compute the actual correlation between two operations, we take into account only the entries where the morphosyntactic prerequisites of both are met

Method

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

For each pair of operations:

Prerequisites

Select the entries with the morphosyntactic prerequisites for **both** syntactic operations

Actual applicability

Compute the correlation (Pearson's coefficient) between the actual applicability of both operations

A positive correlation means that if one of the operations occurs, the other is more likely to occur

A negative correlation means the other is less likely to occur

One operation or two?

Omission of a frozen prepositional complement and of a free prepositional complement

		Correlation	Sample size
Considered distinct operations	passive/omission of frozen PP	0,61%	933
	passive/omission of free PP	-12 %	246
Merged	passive/omission of PP	-0,13%	1179

Outline

Results

LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE

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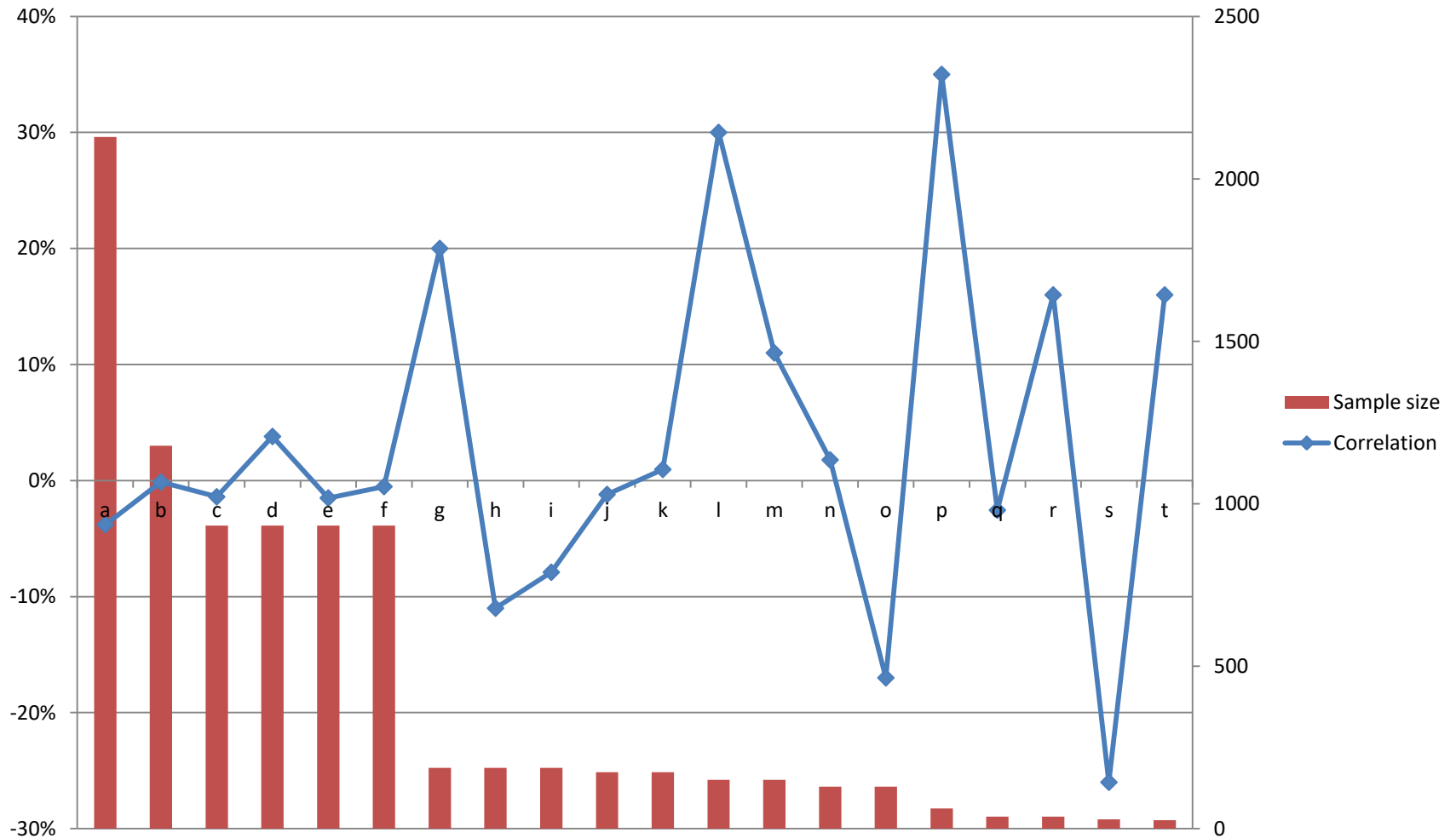ESIEE PARIS
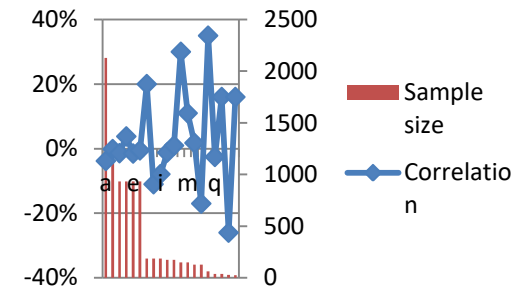UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Results

Sample size

Correlation

25 pairs of operations

5 absolute negative correlations
When one of the operations doesn't apply to any entry in the sample
Example: omission of all complements and dative clitic
    pronominalization of a free complement

9 positive correlations
Strongest: 35%
Dative clitic pronominalizations of adjuncts of direct complement and
    of prepositional complement

11 negative correlations
Strongest: −26%
Clitic pronominalization of prepositional human complement
Adverbial pronominalization of prepositional complement in *sur*

Average
Weighted by sample sizes: −1.4%

Significance

Samples may be too small, not representative
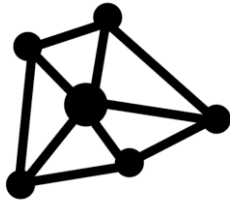
Sample of lexical entries

Syntactic operations marked in available data (explicitly or implicitly)

Values with the largest samples

5 out of the 6 values with the largest samples are weakly negative: –3.8% to –0.13%

Conclusion

The notion of syntactic flexibility as a one-dimensional phenomenon seems shaky

A mix of weakly correlated phenomena

Correlation seems too weak to predict which operations are applicable

**LABORATOIRE D'INFORMATIQUE
GASPARD-MONGE**

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

Thanks

CONTACT

ÉRIC LAPORTE

00 +33 (0)1 60 95 75 52

ERIC.LAPORTE@UNIV-PARIS-EST.FR