

# A probabilistic multifactorial approach to the nature of contact-induced change

A case-study on English substitutive complex prepositions

23<sup>rd</sup> ISTAL

31 Mar. – 2 Apr. 2017

Thessaloniki

Christophe Béchet (ULg)

# Introduction

- Complex prepositions (henceforth CPs) of the type *Prep (Det) N Prep*
- Grammatical(izing) elements (Hoffmann 2005, Fagard 2009)
- Constructions in CxG terms (Vranjes 2012, Béchet 2014)
- Contact phenomena: attested in many EU languages (Hüning 2014)

# Introduction

- Interesting fact: many English CPs have a Romance nominal + cognate cxsc(Lebenstedt 2015: 2), e.g.
  - *in virtue of* (Fr. *en vertu de*)
  - *in respect of*
  - *in favour of* (Fr. *en faveur de*)
  - *in spite of* < *despite of* (Fr. *en dépit de* < *en despit de*)
  - *in place of* (Fr. *à la place de* < *en (la) place de*)
  - *in lieu of* (Fr. *au lieu de* < *en/el/es lieu(s) de*)  
el = en+le  
es = en+les

# Introduction

- **Contact hypothesis:** such a parallelism with regard to CPs cannot be ascribed to chance and to language-specific developments alone
- Nature of the change?
  - Loan translation or calque: adoption of structure + meaning (nuances) from donor language (Lebenstedt 2015)
  - Analogy (Hoffmann 2005: 4): abrupt establishment of the cx facilitated by its formal parallelism to a previously grammaticalized item.

# Introduction

- Aim: determining how (dis)similar English has been to French with regard to the CPs *in place of* and *in lieu of*, answering the following questions:
  - How much does the behavior of the English cxs resemble that of the French ?
  - What are the areas where the behavior of the English cxs exhibits the largest difference from that of French?
- Steps:
  1. Establishing a taxonomy of co-occurring semantic and morpho-syntactic features which are cross-linguistically valid
  2. Data extraction + manual annotation
  3. Exploring the data for a first language-specific and cross-linguistic characterization
  4. Establishing a multifactorial probabilistic model (see Gries 2013) that can predict the outcome cx in French
  5. Applying the French model to the English data and characterizing the (dis)similarities (MuPDAR approach, see Gries & Deshors 2014)

# Step 1: variable selection

- **DET** (determiner): ‘yes’ or ‘no’
- **COMPSYNT** (syntax of the complement): ‘NP’, ‘NFClau’, ‘ProForm’, ...
- **POSITION** (position within the clause): ‘Init’, ‘Med’, ‘Fin’
- **COMPTYPE** (type of complement): ‘Hum’, ‘Anim’, ‘Plant’, ‘Conc’, Abst\_1, Abst\_2
- **SUBSTYPE** (type of substitutive meaning): ‘Repl’, ‘Comp’, ‘Contr’, ('Conc')
- **VARIANT**: ‘1’ (*au lieu de* | *in lieu of* | *instead of*), ‘2’ (*à la place de* | *in place of*)
- **LANGUAGE**: ‘French’ or ‘English’
- **YEAR**: from 1510s to 1710s (EModE)

# Step 2: data extraction

- Historical corpora:
  - English: EEBO, Lampeter (pamphlets), PCEEC (correspondence)
  - French: FRANTEXT (all possible genres)
- Exhaustive extraction of the substitutive CPs in all their morpho-syntactic and orthographical realizations
- Random selection of 500 observations for each French variant ( $\approx$  1000 observations) = ‘training’ data
- Random selection of 400 observations for each English variant ( $\approx$  1200 observations) = ‘test’ data

## Step 2: annotation

E.g.

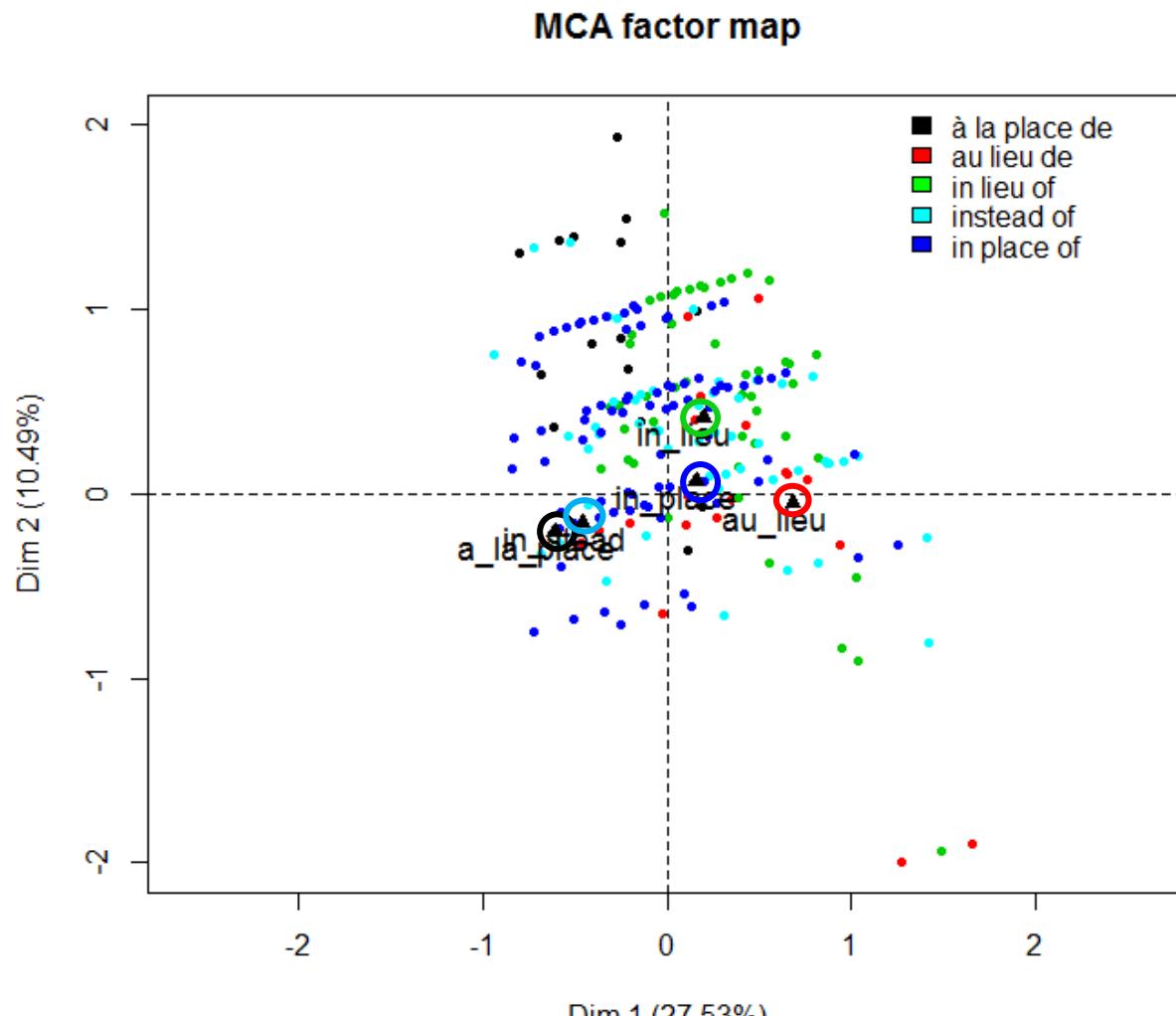
[...] sometime it is shingle , wherewith houses are covered *in the stead of* tyle. (EEBO)

- **DET**: yes
- **COMPSYNT**: NP
- **POSITION**: Fin
- **COMPTYPE**: Conc
- **SUBSTYPE**: Repl
- **VARIANT**: 1
- **LANGUAGE**: English
- **YEAR**: 1542

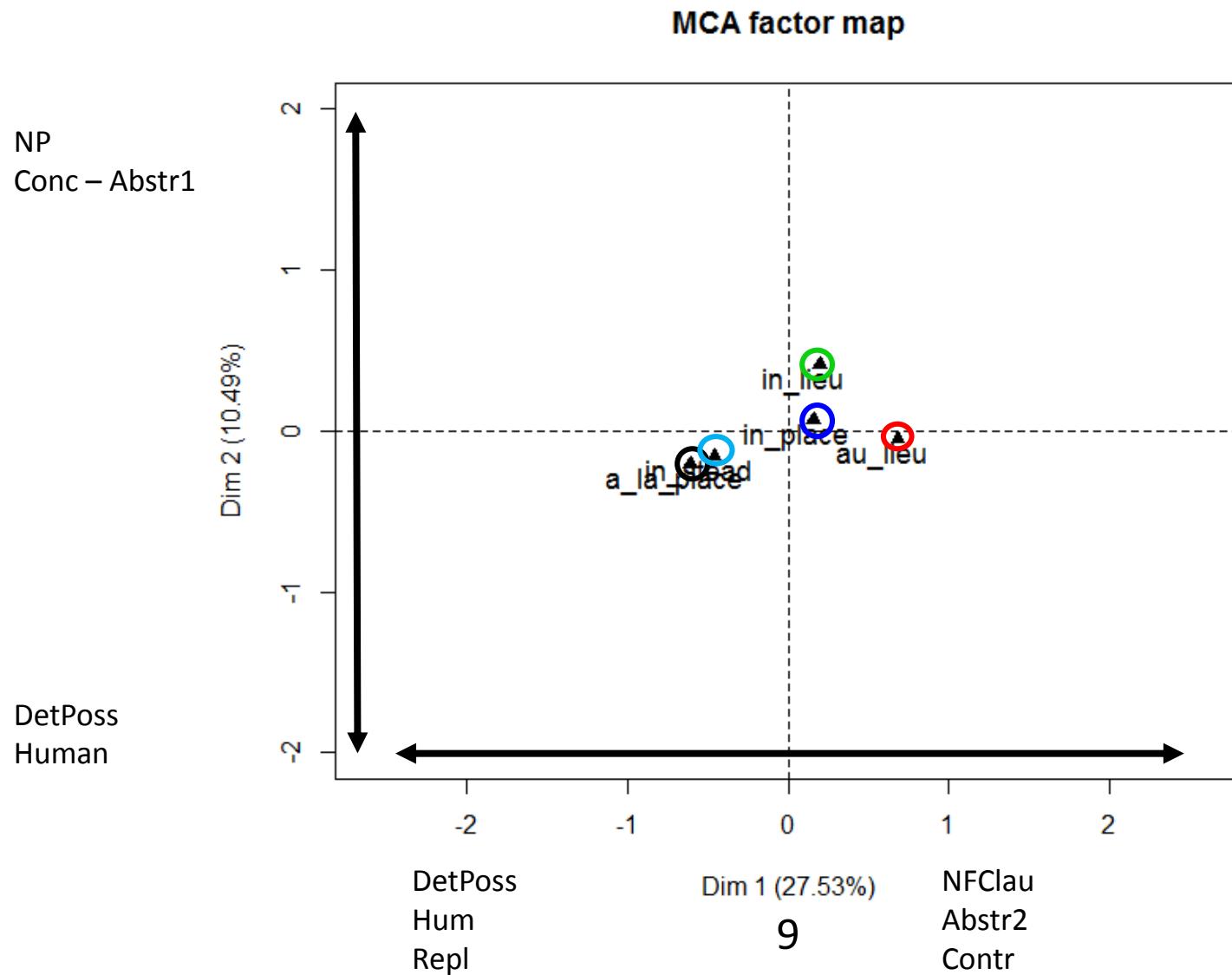
# Step 3: exploration

- Due to data sparseness (e.g. complete separation: no value exists for a given combination of variable values), explanatory variables were better conflated into fewer dimensions
- Technique: Multiple Correspondence analysis (MCA)
- Purpose: transforming row profiles (i.e. combinations of variable values) into coordinates + representation on a bidimensional plot
- Advantages:
  - Gives a visual summary of the data instead of mere counting
  - Conforms to the gradient view of language change

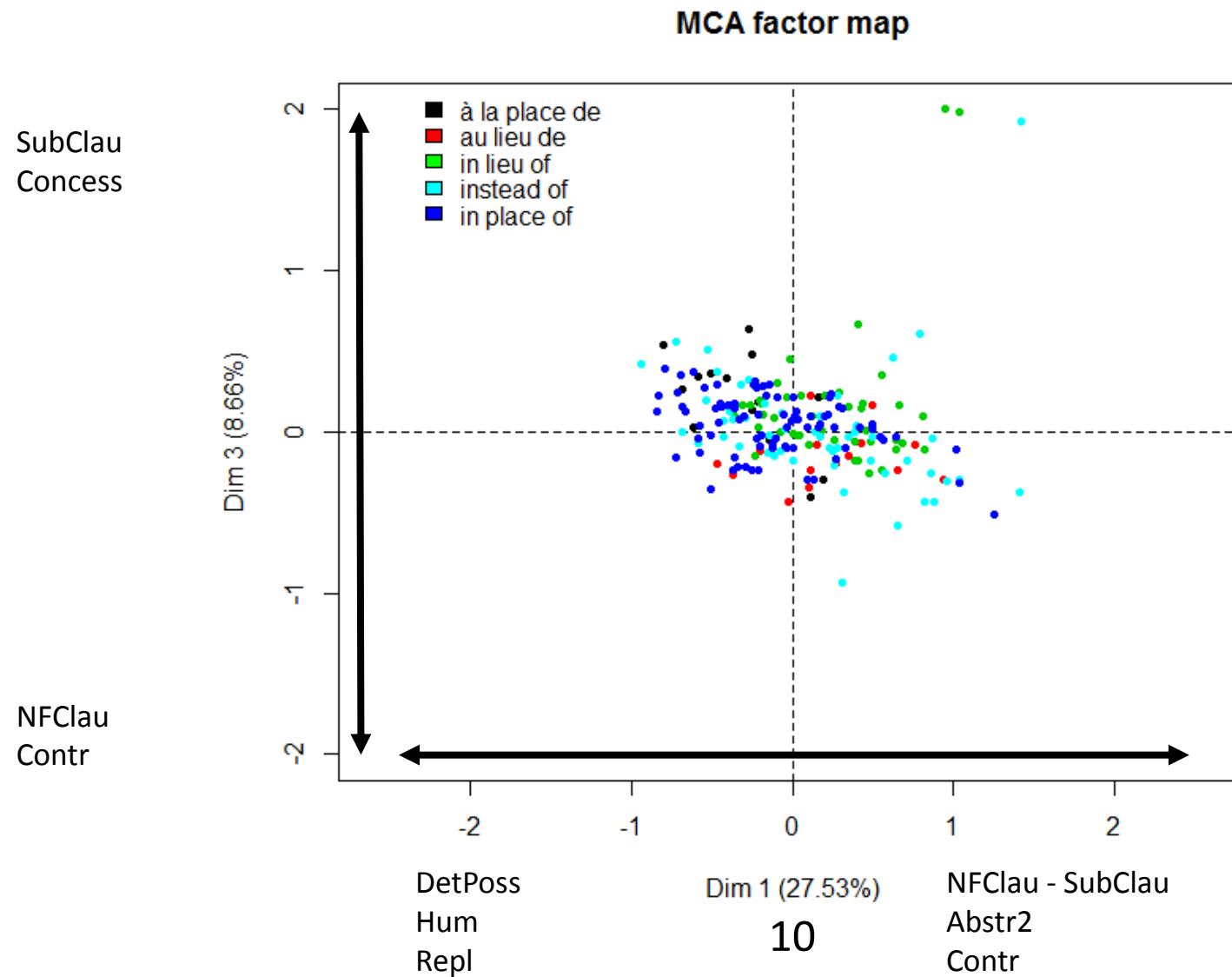
# Step 3: exploration



# Step 3: exploration



# Step 3: exploration



- Multifactorial Prediction and Deviation Analysis with Regressions (Gries & Deshors 2014):
  - Fit a regression model  $R1$  on the French data to predict the CP choices
  - Apply  $R1$  on the English data (“Which variant would a French writer use, given the context of occurrence of the English observation?”)
  - Compute the deviations from  $R1$  for each observation
    - if English choice = French choice, Dev = 0
    - if English choice ≠ French choice, Dev =  $p - 0.5$  (probability of variant 2 - threshold)
  - Fit a regression model  $R2$  with Dev as outcome ~ parameters of  $R1$
  - Analyse and interpret the deviations

# Step 4: $R1$

Université  
de Liège

	Estimate	std. Error	z value	p	Signif.
(Intercept)	-44.977223	5.262759	-8.546	<2e-16 ( $\simeq 0$ )	***
Year	0.026472	0.003176	8.334	<2e-16	***
Dim 1	-6.646023	0.584180	-11.377	<2e-16	***
Dim 2	-0.390299	0.459233	-0.850	0.3954	
Dim 3	-59.315111	23.173587	-2.560	0.0105	*
Year:Dim 3	0.034954	0.013845	2.525	0.0116	*

- Highly significant model: likelihood ratio = 1033.44,  $df = 5$ ,  $p_{model} = 3.48e-221$
- Outstanding classificatory power:  $C = 0.981$ ,  $R^2 = 0.857$

# Step 4: $R1$

	<b>Prediction <i>au lieu de</i></b>	<b>Prediction <i>à la place de</i></b>	<b>TOTALS</b>
Choice <i>au lieu de</i>	461	43	504
Choice <i>à la place de</i>	32	468	500
<b>TOTALS</b>	493	511	1004

- Classification accuracy: 92.53% (against a chance accuracy of 50.2%)

# Step 5: application

## Variant 1: *in lieu of*

	Prediction <i>au lieu de</i>	Prediction <i>à la place de</i>	TOTALS
Choice <i>in lieu of</i>	298	103	401
Choice <i>in place of</i>	97	294	391
TOTALS	395	397	792

- Classification accuracy: 74.74% (against a chance accuracy of 50.63%)

# Step 5: application

## Variant 1: *instead of*

Université  
de Liège

	Prediction <i>au lieu de</i>	Prediction <i>à la place de</i>	TOTALS
Choice <i>instead of</i>	264	136	400
Choice <i>in place of</i>	97	294	391
TOTALS	361	430	791

- Classification accuracy: 70.54% (against a chance accuracy of 50.57%)

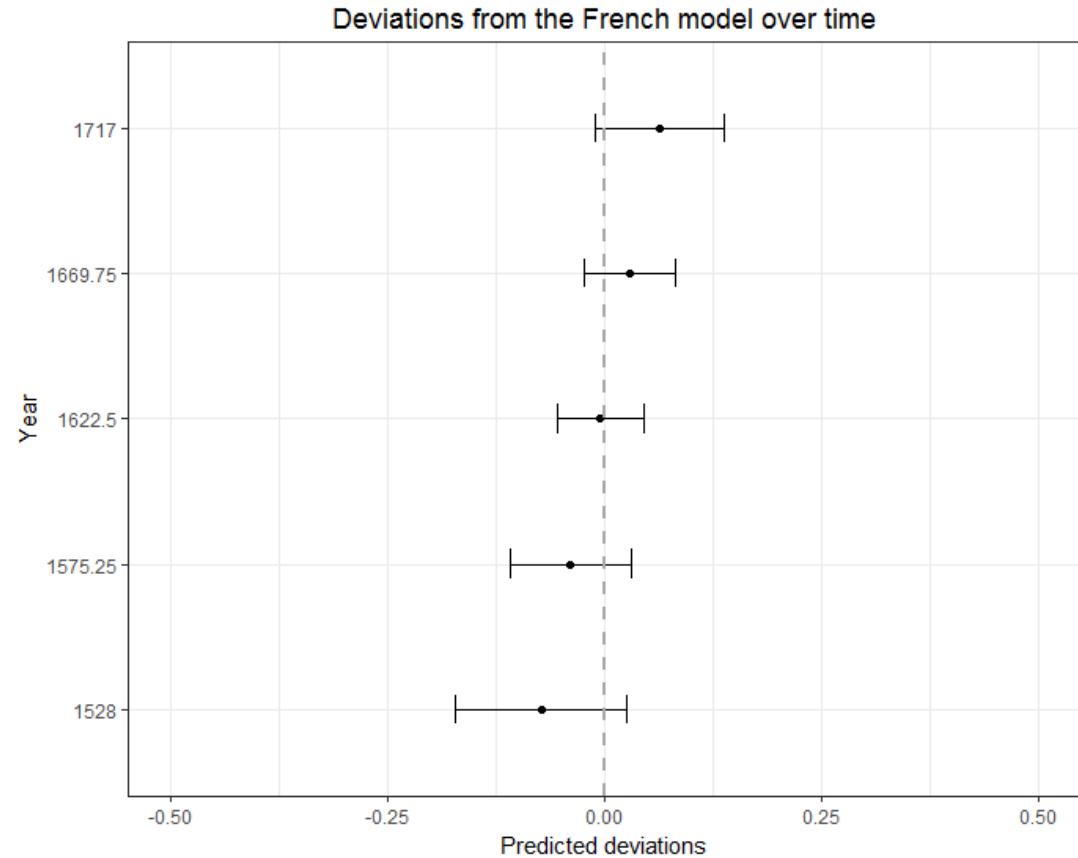
# Step 5: $R^2$

Université  
de Liège

	Estimate	std. Error	t value	p	Signif.
(Intercept)	-1.9406543	0.2417325	-8.028	3.60e-15	***
Year	0.0011705	0.0001470	7.961	5.98e-15	***
Dim 1	-0.0640535	0.0102365	-6.257	6.43e-10	***
Dim 2	0.0006293	0.0101659	0.062	0.951	
Dim 3	-0.7005699	0.5700446	-1.229	0.219	
Year:Dim 3	0.0004479	0.0003489	1.284	0.200	

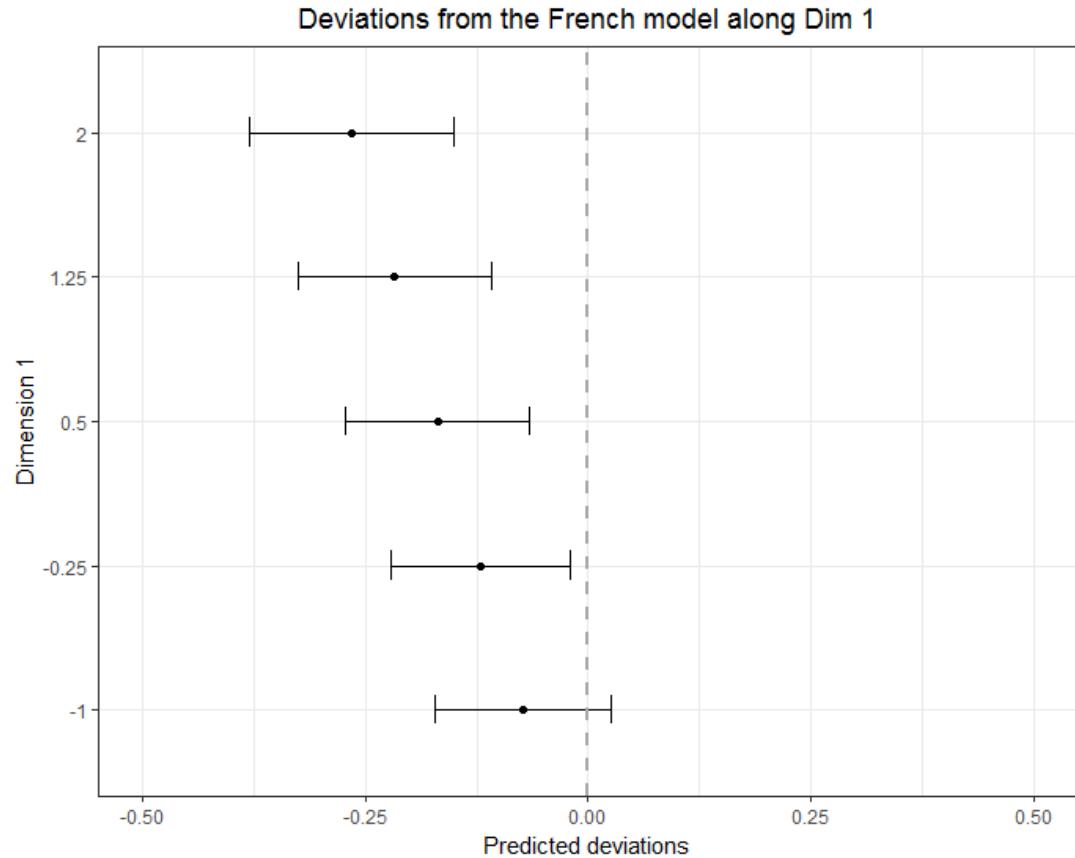
- Highly significant model:  $F = 24.14$ ,  $df = 24.14$  on 5 and 786,  $p_{\text{model}} = p_{\text{model}} < 2.2e-16$  (almost 0)
- BUT weak correlation between the deviations and the dimensions and time: adj.  $R^2 = 0.13$

# Step 5: $R^2$



- As time goes on, the deviation of English from French when French would use *au lieu de*, diminishes (i.e. English starts to use *in lieu of* more often)
- However, when French would use *à la place de*, English deviates by using *in lieu of* more often.

# Step 5: $R^2$



- For every unit increase along the first dimension, it appears that English is less conservative and tends to use *in place of* when *in lieu of* would be expected (if truly equivalent to *au lieu de*).

# Interpretation

- The panchronic exploration of the data suggested that the two cxs *in place of* and *in lieu of*, even though formally related to their French counterparts, did not develop fully according to the French model. This has been (partially) confirmed by the decrease in classification accuracy when applying the model to the English data.
- The asymmetry between the English and French constructions since the earliest times of their use shows that the substitutive CPs in English have developed their own niches (e.g. absolute use of *instead*, compensating meaning of *in lieu of* in legal dealing, restricted use of *in place of*) (cf. Schwenter & Traugott 1995)

# Interpretation

- *in place of* and *in lieu of*:
  - didn't grammaticalize analogically to *instead of*
  - do not seem to be fully translated from their French cognates
- HOWEVER
  - the period chosen for this pilot study may be too short to witness contact-induced language change for the CPs under scrutiny
  - there might be socio-historical reasons for their borrowing (e.g. prestige, stylistic variation)
  - some explanatory features might be lost (hypothesis of an early development in Anglo-Norman)

# Conclusion

- Pluses:
  - Contact-induced language change can be explored in a multivariate design
  - The MuPDAR approach is promising in many respects and found its first application in a pilot study of a contact phenomenon
- Minuses:
  - The methodology proposed needs a finer-grained operationalization of the data (6 explanatory variables conflated into 4 dimensions, too few variability explained by these dimensions to explain deviations)
  - The size of the dataset ( $\simeq 2200$  observations) is still too small to get a full picture

# Conclusion

- Suggestions for further research:
  - Study other periods than EModE (e.g. ME to trace the development of *instead of*)
  - Working on larger data sets when available
  - Taking the categorical nature of the data into account (reducing their dimensions can lead to loss of important information)
  - Including other variables (e.g. frames of ‘substitution’, register, domain, ...)
  - Taking within-author and within-text variation into consideration (by e.g. using Mixed-Effects Logistic Regression)

Thank you very much for your  
attention!

# References

- Béchet, C. (2014). *The Influence of French on Grammatical Constructions in English and Dutch: A cognitive approach in a contrastive-historical study* (Unpublished master's thesis). University of Liège, Liège, Belgium.
- Fagard, B. (2009). Prépositions et locutions prépositionnelles: Un sémantisme comparable? *Langages*, 173 : 95-113.
- Gries, S. T. (2013). *Statistics for Linguistics with R: A Practical Introduction* (2<sup>nd</sup> ed.). Berlin/New York: De Gruyter Mouton.
- Gries, S. T. and S. C. Deshors (2014). Using regressions to explore deviations between corpus data and standard/target: two suggestions. *Corpora*, 9(1): 109-136.
- Hoffmann, S. (2005). *Grammaticalization and English Complex Prepositions: A corpusbased study*. London: Routledge.
- Hüning, M. (2014). Over complexe preposities en convergentie. In F. Van de Velde, H. Smessaert et al. (eds.), *Patroon en argument: Een dubbelfeestbundel bij het emeritaat van William Van Belle en Joop van der Horst*. Leuven: Universitaire Pers Leuven, 433-445.
- Lebenstedt, A.-L. (2015). The Influence of French on the Formation of English Complex Prepositions: Loan Translation and Grammaticalization by Analogy, Unpublished bachelor thesis, University of Trier, Trier, Germany.
- Schwenter, S. A. and E. C. Traugott (1995). The Semantic and Pragmatic Development of Substitutive Complex Prepositions in English. In A. H. Jucker (ed.), *Historical Pragmatics: Pragmatic developments in the history of English*. Amsterdam: John Benjamins Publishing Company.
- Vranjes, J. (2012). *Voorzetseluitdrukkingen in het Nederlands sinds de 16de eeuw: een diachroon corpusonderzoek* (Unpublished master thesis). University of leuven, Leuven, Belgium.