References

Capturing historical causality via game-theoretic interactions

Alexandra.Simonenko@UGent.be

Correlations

- Typologically, subject features (person, number) tend to be encoded *either* as verbal agreement affixes *or* as pronouns (Berdičevskis et al. 2020);
- Diachronically, in a number of documented languages verbal **agreement syncretisation** was accompanied by the **rise of subject pronouns**.

	Old	Modern French
1st	aim	j'aime
3rd	aimet	il aime

Table 1: Illustration of agreement syncretisation and overt subject emergence in French

- Using the treebank of Kroch and Santorini (2021), Simonenko et al. (2019) quantified the emergence of syncretic (=ambiguous) endings in French.

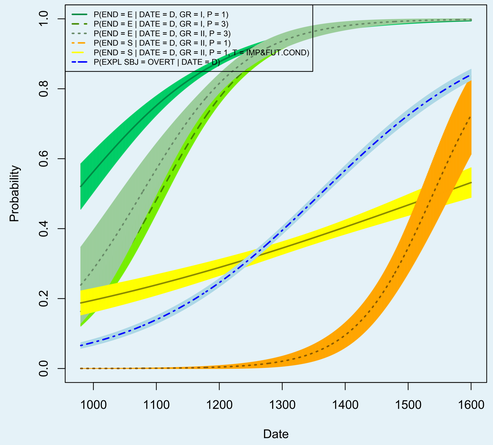

Figure 1: Log. regressions fitted to variables "old/new ending" & "null/overt subject"

- Causality *syncretisation* → *overt subjects* has been hypothesized.
- **Can we build a (quantitatively, pragmatically, grammatically) convincing model of causality?**

Models

- **A**. A speaker makes a stochastic **choice between two encodings** ("null pronoun" vs. "overt pronoun"). The probability distribution depends on whether a given ending is ambiguous: "null pronoun" is predicted to be less likely with ambiguous ones.
- **B**. A speaker makes a stochastic **choice between two abstract grammars** ("null pronoun" vs. "overt pronoun"). The probability distribution depends on how successful a given grammar has been. The latter depends on the degree of ending ambiguity in the data. [in progress]

Model A

Rational Speech Act model (Frank and Goodman 2012):

- Speaker = probability distribution over utterances.
- Listener = probability distribution over information states.

Speaker, when assigning probabilities to utterances u_1 ... u_n to convey a meaning m_i, takes into account how **likely** the Listener is to get m_i upon hearing a given u_j, and how costly u_j is.

(1) Speaker's utility function
$U_S(u_j; s_i) = \mathbf{\log L(s_i \mid u_j)} - C(u_j)$ Scontras et al. (2017)

(2) $P_S(u \mid s) \propto \exp(\alpha(\log L_0(s \mid u) - C(u)))$

Implementation

ENDING	MEANING
"V-zero"	{[1st person]}
"je-V-zero"	{[1st person]}
"V-s"	{[2nd person]}
"tu-V-s"	{[2nd person]}
"V-t"	{[3rd person]}
"il-V-t"	{[3rd person]}
"V-e"	{[1st person], [3rd person]}
"il-V-e"	{[3rd person]}

Table 2: Simplified utterance inventory

	INPUT	OUTPUT
LISTENER: "V-zero"		
	a) meaning of "V-zero" **b**) priors	
	P([1st person]) = 0.33	P([1st person]) = 1
	P([2nd person]) = 0.33	P([2nd person]) = 0
	P([3rd person]) = 0.33	P([3rd person]) = 0
LISTENER: "V-e"		
	a) meaning of "V-e" **b**) priors	
	P([1st person]) = 0.33	P([1st person]) = **0.5**
	P([2nd person]) = 0.33	P([2nd person]) = 0
	P([3rd person]) = 0.33	P([3rd person]) = 0.5
LISTENER: "je-V-e"		
	a) meaning of "je-V-e" **b**) priors	
	P([1st person]) = 0.33	P([1st person]) = **1**
	P([2nd person]) = 0.33	P([2nd person]) = 0
	P([3rd person]) = 0.33	P([3rd person]) = 0

Prediction

SPEAKER: [1st person]

a) model of the (Literal) Listener
b) utterance C(osts)

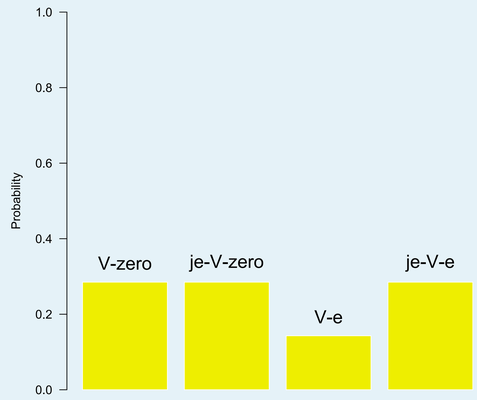

Figure 2: Predicted probabilities of subject encoding

Historical French data

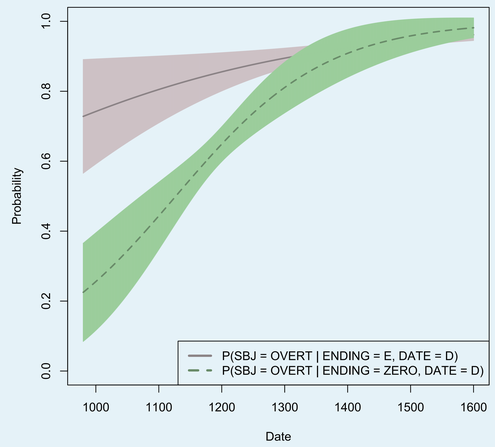

Figure 3: Log. regressions fitted to variable "null/overt subject"

Problems

- The model tolerates null subjects, Modern French does not.
- Null subjects disappear with unambiguous endings (e.g. "zero") as well.

Model B [in progress]

- Reinforcement learning agents with two competing abstract grammars.
- Success = the message is interpreted as intended.

GRAMMAR	OUTPUT	SUCCESS/FAILURE	REINFORCEMENT
SBJ-NULL	V-zero	1/0	always
SBJ-NULL	V-e	p/q	?
SBJ-OVERT	il-V-zero	1/0	always
SBJ-OVERT	il-V-e	1/0	always

- To build in: degree of ambiguity (≈ Shannon's entropy) for new endings growing over time (∼ success probability for SBJ-NULL going down)

References

Berdičevskis, Aleksandrs, Karsten Schmidtke-Bode, and Ilja Seržant. 2020. Subjects tend to be coded only once: Corpus-based and grammar-based evidence for an efficiency-driven trade-off. In *Proceedings of the 19th International Workshop on Treebanks and Linguistic Theories*, 79–92.

Frank, Michael, and Noah Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336:998–998.

Kroch, Anthony, and Beatrice Santorini. 2021. *Penn-BFM Parsed Corpus of Historical French (PPCHF)*. URL `https://github.com/beatrice57/mcvf-plus-ppchf/`.

Scontras, Gregory, Michael Henry Tessler, and Michael Franke. 2017. Probabilistic language understanding: An introduction to the Rational Speech Act framework. Technical report, https://michael-franke.github.io/probLang.

Simonenko, Alexandra, Benoit Crabbé, and Sophie Prévost. 2019. Null subject loss and subject agreement syncretisation: Quantificational models for Medieval French. *Language Variation and Change* 31:275–301.