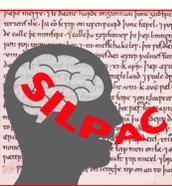


Learnability in diachrony: applying the Tolerance Principle to historical data

Collaborative work with Mercator Fellow Charles Yang, UPenn



1. Loss of verb-particle constructions in French (H1, Piccione)

Phenomenon: Verb-particle constructions (VPCs) were used in Medieval French before disappearing by the 16th century

Research question: How are VPCs learnt throughout the history of French?

Method: simulation of learning trajectories based on the *Tolerance/Sufficiency Principle* (TSP)

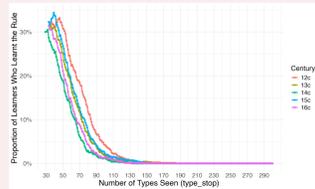
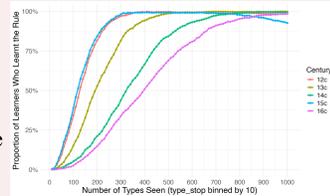
- to account for variation, **simulation of 1000 learners through input sampling** (adapted from Trips, Kodner, & Rainsford, 2025): the input samples are created by collecting one verb at a time drawn randomly from a list containing *verb* and *verb+particle* based on their frequency of occurrence in the corpus.

| Lemma | Frequency w/o | Frequency w/ particles |
|---------------------|---------------|------------------------|
| aller ('to go') | 3 | 2 |
| courir ('to run') | 2 | 1 |
| parler ('to speak') | 1 | 1 |

Input aller, courir+particle, aller+particle, parler, courir, aller, aller, parler+particle, courir, aller+particle

Results:

- Motion** is a necessary condition for a productive rule for VPCs
- However, it is not sufficient, suggesting to look for a productive rule within subclasses of motion verbs



- A productive rule in the subclasses of motion (e.g., **caused motion**) does not hold for all the learners → VPCs not learnt productively, but memorised

Piccione, Mariapaola (in preparation). *A (child) Learner-based Approach to Changes in Motion Lexicalization in French and Italian* [PhD Dissertation], U Stuttgart.

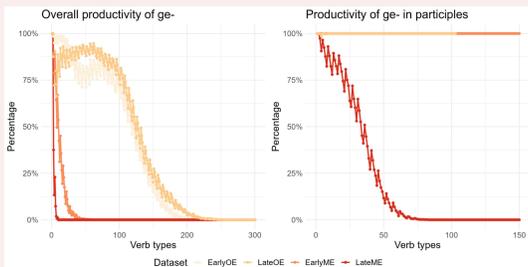
2. (Re)-assessing the productivity of early English a- and ge-: what the Tolerance Principle can tell us

Aim: Test the hypotheses from the literature that the prefix *ge-* shifts from derivational to inflectional prefix in early English, and that the *ge-* and *a-* prefix can be used interchangeably.

Research Question: Is prefixation of both *ge-* and *a-* productive? Does *ge-* shift from a productive derivational affix to an inflectional affix only productive on participles?

Methodology: Simulation of 1000 learners through random input sampling, using an exposure threshold of 10 tokens before a type is assumed to be acquired as an alternative to frequency trimming to avoid predetermining the set of acquired types at a certain N. A 1 token = 1 type approach leads to unreliable results w.r.t. optional prefixation in a full lexicon.

Results: While prefixation is productive in OE and becomes unproductive towards ME, *ge-* is productive for only a subset of speakers at small vocabulary sizes. Participles, however, show a fully productive system until eME, suggesting that *ge-* is already in the transition from a derivational to an inflectional prefix in OE. *A-* is never productive as a catch-all prefix, nor as a marker of participles. *Ge-* and *a-* have a different status in the grammar and



are not freely interchangeable variants. The TP is thus able to statistically confirm and nuance the trajectory of change: the prefix system was hard to acquire already in OE, which eventually resulted in its loss.

The Tolerance Principle

A productive rule applicable to N items in the learning data cannot have more than

$$\theta_N = \frac{N}{\ln N} \text{ exceptions.}$$

| N | θ_N | % |
|------|------------|------|
| 10 | 4 | 40 % |
| 20 | 6 | 30 % |
| 50 | 12 | 24 % |
| 100 | 21 | 21 % |
| 500 | 80 | 16 % |
| 1000 | 144 | 14 % |

- Parameter free
- Adequate rather than optimal learning
- Recursive applications for sub-regularities
- Full computational implementation
- A formal and general mechanism that holds across domains
- Robust evidence from artificial language experiments and many corpus, behavioral and computational studies of phonological and morphological learning, and syntactic acquisition (argument structure, recursion, islands).

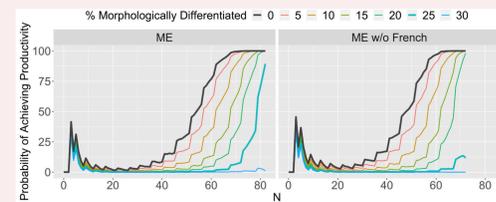
3. Diachronic (learn)ability: Modelling acquisition of labile verbs in Middle English (H1 Rainsford, H3 Trips, Jordan Kodner)

Phenomenon: Modelling the acquisition of labile verbs in Middle English

Method: frequency-weighted sampling: Sample N change-of-state verbs many times and evaluate each sample with the Sufficiency Principle

RQ: How did children learn the class of labile verbs in Middle English (ME) under the contact hypothesis and applying the TSP?

- including French verbs it is acquired slightly earlier
- loss of morphological differentiation strengthens the lability rule
- influence from French can be seen as having an acceleratory effect in conjunction with morphological leveling.



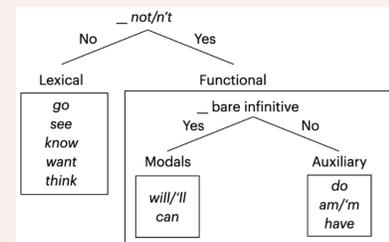
Talk by Trips, Rainsford, Kodner at DIGS in Oxford 2025 on the data basis of Trips & Stein 2024.

4. Distributional learning and grammaticalization: Modals in the history of English (H3, Yang & Trips, submitted)

- TP as a model of distributional categorization: A class of N words is deemed *sufficiently* homogeneous if no more than θ_N items are distributionally distinctive. Otherwise *subdivision* is necessary.
- Infant's sensitivity to **distributional information** including (linear) word order with reference to high frequency items and segmentation and representation of morphological inflection.

Results

Two **distributional signatures** concerning **negation** and **bare infinitive** complementation are sufficient to identify the main verbal categories:



Syntactic categories are the formal consequences of distributional learning in child language acquisition, a **discovery procedure** in the original conception of generative grammar: No additional principles or parameters are necessary.

Yang, Charles, Trips, Carola. Distributional learning as grammaticalization: Modals in the history of English. Submitted to *Journal of Linguistics*.

4. The development of anticausatives in French (H2, Cassarà)

Phenomenon: Anticausative verbs: sub-class of unaccusatives that alternate between two valencies (transitive vs. anticausative), e.g. to *break*: i) the wind broke the vase. vs. ii) the vase broke.

Modern French (MF):

- UAC = Unmarked anticausatives (labile verbs, no change in form when anticausatives), select *avoir*, e.g. *augmenter* (increase)
- RAC = Verbs that take the reflexive morpheme *se* when anticausatives, select *être*. E.g. *se casser* (break) Modern French: 1700 RAC and only 311 UAC. RAC are the default way to form anticausativity.

Changes in the diagnostics of this class in the history of French:

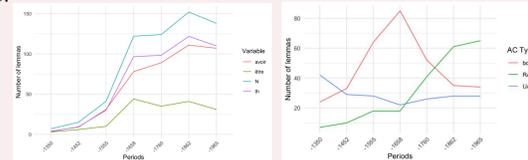
- RAC expanded (e.g. OF: *briser* 'break' => MF: *se briser* 'break').
- Auxiliary selection changed for UAC (OF: *être* => MF *avoir*).

Research question:

How do unaccusatives diagnostics changed in the history of French?

Main claim:

Due to the expansion of RAC, UAC became too scarce to form a class on their own. Consequently, RAC absorbed auxiliary *être* and *avoir* was generalised to the remaining UAC.



Applying the TP:

The TP calculation in the table considers RAC as the rule and UAC as exceptions. N is the sum of RAC+UAC. It excludes the cases showing variation between RAC and UAC. The threshold is θ . The column **RAC_productive** shows at which point RAC became a productive rule, 'yes' in the table means productive.

Discussion:

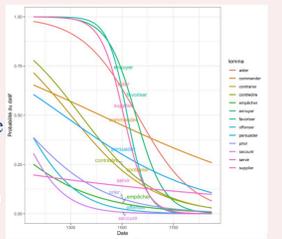
- The chronology of the changes observed seems to indicate that the class of RAC has expanded before the change of *être* to *avoir*.
- Auxiliary change is due to RAC becoming the rule and UAC the exception.

Cassarà, Alessia. Paper presented at the Diachronic Generative Syntax Conference (DiGS25), Mannheim

6. Dative-taking monotransitive (DTM) verbs in French (H2, Stein, Troberg)

Phenomenon: A stable group of DTM have survived, but a subgroup have replaced the dative with an accusative object (e.g. *aider* 'help'; Troberg 2013, see Fig.).

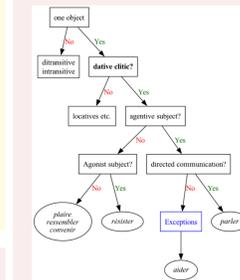
→ There is a rule (a) proper to the *aider* type, (b) relevant for acquisition, (c) cutting across semantic classes:



Agonist subjects cooccur with datives.

TP analysis: Case variation was resolved as soon as this rule was acquired (around 15th-16th c.)

Verbs like *obéir* 'obey' selected stable dative. Verbs like *aider* had no Agonist subject, selected accusative and were memorized as exceptions to that rule. (Fig. recursive application of the TP).



Stein, Achim & Troberg, Michelle. Presented at Diachro XI, Madrid, May '24.

7. The rise of raising in to-infinitives and of -ing-complements in Middle and Modern English (H4, Björnsdóttir, Gotthard & Walkden, under review; Meiszahl & Walkden 2025)

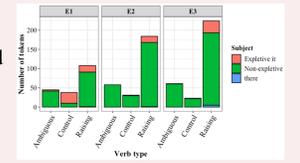
Phenomenon: Raising with *to*-infinitives spreads in Middle English (ME) and Early Modern English (EModE). *Verbal -ing-complements* spread in EModE and Late Modern English (LModE). Can the TP shed light on these developments?

Research questions:

- Does the spread of (i) raising with *to*-infinitives and (ii) *-ing*-complements proceed via **lexical diffusion**, from verb to verb?
- If so, is this lexical diffusion governed by a) formal properties such as co-occurrence with **expletive subjects** and/or b) **lexical semantic classes** of the verbs in question?

Method: Extraction of *to*-infinitives from PPCME2 and PPCME1; lemmatization using BASICS toolkit and coding for formal properties; verb classes coded based on Levin (1993); TP analysis using frequency-trimming following Kodner (2019) and Trips & Rainsford (2022).

Results: 1. Lexically specific patterns are found for both developments. TP predicts 2a) productive association between expletive *it* and raising (ME and EModE), but 2b) no clear and consistent association with any non-trivial lexical semantic class, either for raising or for *to*-infinitives.



Discussion: According to the TP, both raising and *-ing*-complementation should be unproductive and lexically specific rather than generalised to verb classes.

| Class | -ing | non-ing | θ | TP | SP |
|---------------------|------|---------|----------|-------|-------|
| Remove Verbs | 1 | 2 | TRUE | FALSE | |
| Verbs Of Fulfilling | 1 | 1 | TRUE | TRUE | |
| Hold Verbs | 1 | 0 | Inf | TRUE | TRUE |
| Keep Verbs | 1 | 1 | 2 | TRUE | TRUE |
| See Verbs | 2 | 1 | 2 | TRUE | TRUE |
| Leave Verbs | 1 | 0 | Inf | TRUE | TRUE |
| Avoid Verbs | 1 | 0 | Inf | TRUE | TRUE |
| Characterize Verbs | 3 | 9 | 4 | FALSE | FALSE |
| Fit Verbs | 2 | 4 | 3 | FALSE | FALSE |

This seems to be borne out by comparison to present-day child language acquisition data. Formal properties were likely more influential in both cases – for raising, the emergence of expletives was crucial.

Björnsdóttir, Sigríður, Lisa Gotthard, & George Walkden. Under review. Lexical learning and language change. Ms. Meiszahl, Marc, & George Walkden. 2025. Paper presented at the Diachronic Generative Syntax Conference (DiGS26), Oxford.