

# Towards the Degrammaticalization Database

L. S. Fayzeeva  
Supervisor: K. V. Filatov  
HSE University, Moscow 2025

## 01 Introduction

Over the past century, a substantial body of research regarding grammaticalization has emerged. It was first introduced by Meillet [1912: 131] as a functional shift of a linguistic entity from an autonomous to a grammatical meaning. This definition, as well as the following ones, presupposed grammaticalization to be a unidirectional process. Following this assumption, the very idea of degrammaticalization was long considered to be impossible. As the result, cases that appeared to go “upward” grammaticalization cline were considered marginal and statistically insignificant. This view began to shatter with Ramat’s [1992] groundbreaking paper, which challenged the unidirectionality principle by discussing attested cases of degrammaticalization. This has led to an ongoing debate: not only about the existence of degrammaticalization but also about how to define it, and where to draw its boundaries in order to support or challenge the unidirectionality principle. As a result, the conceptual space surrounding degrammaticalization has become increasingly entangled with overlapping and even contradictory definitions. Scholars in the field continue to introduce new terms and refine the old ones, making systematic comparison across cases more and more difficult. The field is thus at risk of circulating definitions restlessly, overlooking the actual data, buried too deeply into the debating labels rather than mechanisms. It becomes increasingly clear that what the scholarly debate around degrammaticalization is missing is a shift in methodology — a way to return to the empirical level without pre-classifying the phenomena we aim to describe, not to define.

## 02 General Architecture of the Database

The aim of my coursepaper is to develop a systematic, empirically grounded database of attested degrammaticalization instance across languages. The database is structured around primitive parameters that capture key aspects of change. This parameters are grouped into four broader clusters, following practices established in World Lexicon of Grammaticalization [Heine, Kuteva 2002; Kuteva et al. 2019]:

- source
    - form
    - type
      - content word, content wordform, function word, clitic or phrasal affix, bound affix-like morpheme, derivational affix, inflectional affix, wordpart, construction
    - meaning
    - language
    - host
- process
  - target
    - (is broken down the same way “source”-cluster is)
  - metadata
    - citation
    - critique
    - primary source
    - comments

-ology	used to form nouns indicating a branch of learning	stem	English	bound affix-like morphemes	generalization of an affix into a hypernym driven by mass borrowing	
-onomy	used to form nouns indicating a branch of learning	stem	English	bound affix-like morphemes	generalization of an affix into a hypernym driven by mass borrowing	
pro-	used to form nouns indicating someone who supports or agrees with some ideas	nominal stem	English	bound affix-like morphemes	generalization of an affix into a hypernym driven by mass borrowing	
-achtig	used to form adjectives with an approximate, diminutive or similative meaning	noun stem or adjective stem	Dutch	derivational affix	increase in syntagmatic autonomy	
-ish	used to form adjectives with an approximate, diminutive or similative meaning	noun stem or adjective stem	English	derivational affix	increase in syntagmatic autonomy	
-mos	1PL	verbal stem	Spanish	inflectional affix	increase in syntagmatic autonomy	
-muid	1PLS	verbal stem	Irish	inflectional affix	increase in syntagmatic autonomy	
generalization of an affix into a hypernym driven by mass borrowing	ology	'a branch of learning'		N/A	English	content word
generalization of an affix into a hypernym driven by mass borrowing	onomy	'a branch of learning'		N/A	English	content word
generalization of an affix into a hypernym driven by mass borrowing	pro	'a person who supports a concept or a principle'		N/A	English	content word
increase in syntagmatic autonomy	=achtig	'typical, similar, like; somewhat, rather; approximately'		noun/NP	Dutch	clitic or phrasal affix
increase in syntagmatic autonomy	=ish	'typical, similar, like; somewhat, rather; approximately'		noun/NP	English	clitic or phrasal affix
increase in syntagmatic autonomy	=nos	'we'		N/A	Mexican Spanish	clitic or phrasal affix
increase in syntagmatic autonomy	muid	'we'		N/A	Irish	function word

Figure 1. Rows 39 to 45 of the database

## 03 Classification of Processes and Preliminary Findings

The “process” parameter in the database aims to capture the transformation observed in each case of attested grammaticalization in minimal terms. These are not broad clssificatory labels, but rather short descriptions of an underlying mechanism. The classification was created bottom-up by grouping the examples that seemed to be driven by the same motivation.

- bound entities gaining lexical status
    - generalization of an affix into a hypernym
      - generalization of an affix into a hypernym driven by mass borrowing
    - semantically unpredictable generalization of an affix
    - separation of a wordpart
  - increase in syntagmatic autonomy
  - inflectional marker to derivational marker
    - inflectional marker to grammatical derivational marker
    - inflectional marker to lexical derivational marker
  - derivation from a function word
    - delocutive word-formation
- conversion
    - taboo-driven conversion
  - clipping
    - clipping followed by conversion
  - wordform to lexeme
  - functional expansion
    - functional expansion into a specific context
  - lexical shift
  - modal verb to lexical verb
  - univerbation
  - loss of an inflectional category
    - loss of an inflectional category with traces

While developing the process-based tagging model, several cross-process trends emerged from th data, such as:

1. bound items with clearer meanings are more likely to undergo upgrading
2. lower degrees of autonomy tend to correlate with lower degrees of semantic content (see also [Dahl 2004: 164] for discussion)
3. previously overlooked pragmatic inference and discourse salience play great role in at least some of these processes

## 04 Future Directions

This database offers a framework for comparison and search for recurring patterns, allowing researchers to work with pre-collected data according to their own priorities and definitions of degrammaticalization. It is a flexible tool for empirical comparison, and is designed to accomodate collaborative annotation for gaining more covarage, expanding into underrepresented language families and areals.

### Access the Database

- GitHub: [lililek/the\\_degrammaticalization\\_database](#)
- Contact: [Isfaizeeva@edu.hse.ru](mailto:Isfaizeeva@edu.hse.ru)
- Open for collaboration!

### References

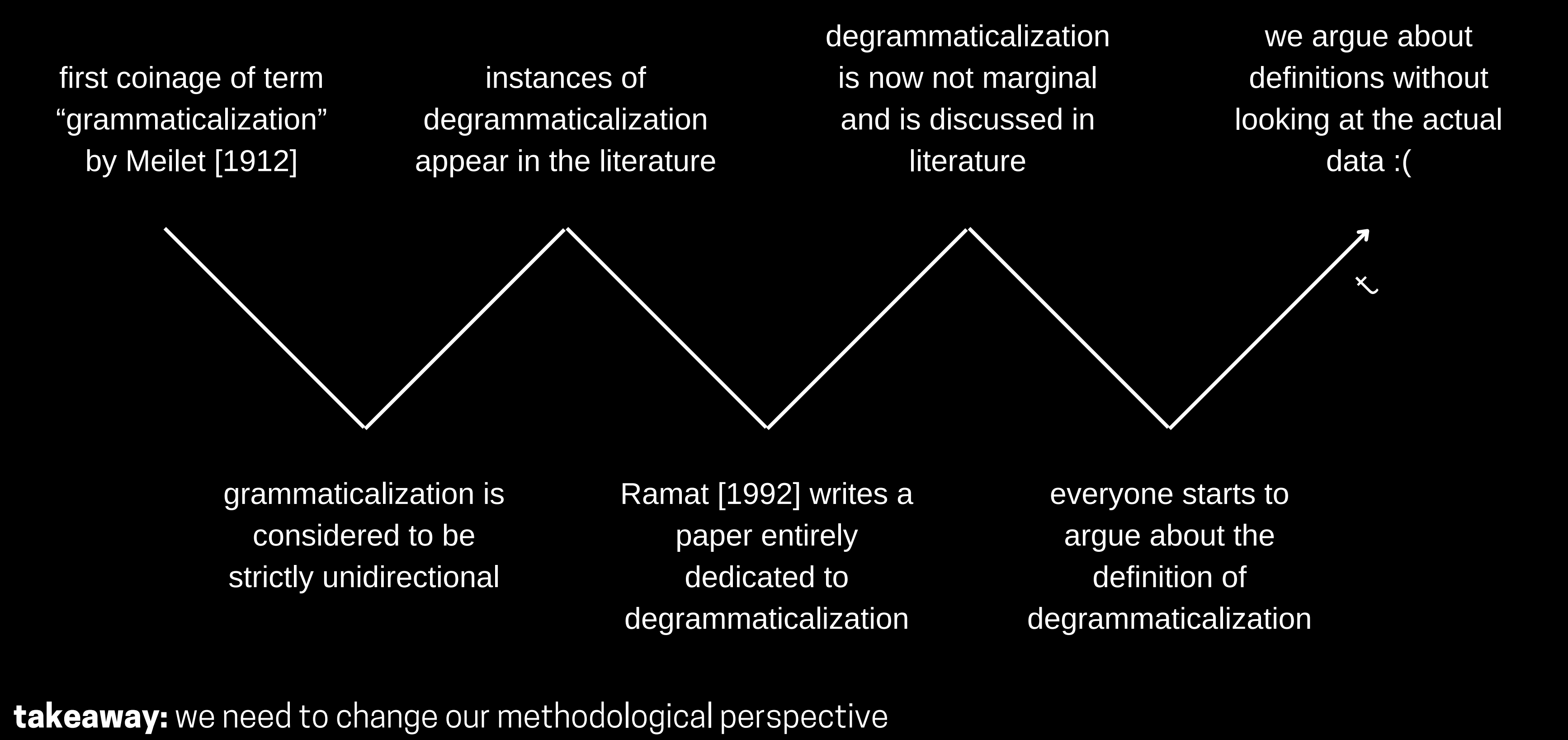
Dahl 2004 — Ö. Dahl. *The growth and maintenance of linguistic complexity*. Amsterdam/Philadelphia: John Benjamins, 2004.  
Heine, Kuteva 2002 — B. Heine, T. Kuteva. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press, 2002.  
Kuteva et al. 2019 — T. Kuteva, B. Heine, B. Hong, H. Long, H. Narrog, S. Rhee. *World lexicon of grammaticalization. 2nd ed*. Cambridge University Press.  
Meillet 1912 — A. Meillet. *Introduction à l'étude des langues indo-européennes*. Paris: Hachette, 1912.  
Ramat 1992 — P. Ramat. Thoughts on degrammaticalization // *Linguistics* 30(2), 1992. P. 547–565.



# Towards the Degrammaticalization Database

L. S. Fayzeeva  
Supervisor: K. V. Filatov  
HSE University, Moscow 2025

## 01 Introduction



## 02 General Architecture of the Database

problem: we need a systematic, empirically grounded database of attested degrammaticalization instances across languages

solution: to create such a dataset by collecting examples from the literature and to annotate it using primitive parameters that capture key aspects of change

- source
    - form
    - type
      - content word, content wordform, function word, clitic or phrasal affix, bound affix-like morpheme, derivational affix, inflectional affix, wordpart, construction
    - meaning
    - language
    - host
- process
  - target
    - (is broken down the same way “source”-cluster is)
  - metadata
    - citation
    - critique
    - primary source
    - comments

-ology	used to form nouns indicating a branch of learning	stem	English	bound affix-like morphemes	generalization of an affix into a hypernym driven by mass borrowing	
-onomy	used to form nouns indicating a branch of learning	stem	English	bound affix-like morphemes	generalization of an affix into a hypernym driven by mass borrowing	
pro-	used to form nouns indicating someone who supports or agrees with some ideas	nominal stem	English	bound affix-like morphemes	generalization of an affix into a hypernym driven by mass borrowing	
-achtig	used to form adjectives with an approximate, diminutive or similative meaning	noun stem or adjective stem	Dutch	derivational affix	increase in syntagmatic autonomy	
-ish	used to form adjectives with an approximate, diminutive or similative meaning	noun stem or adjective stem	English	derivational affix	increase in syntagmatic autonomy	
-mos	1PL	verbal stem	Spanish	inflectional affix	increase in syntagmatic autonomy	
-muid	1PLS	verbal stem	Irish	inflectional affix	increase in syntagmatic autonomy	
generalization of an affix into a hypernym driven by mass borrowing	ology	'a branch of learning'		N/A	English	content word
generalization of an affix into a hypernym driven by mass borrowing	onomy	'a branch of learning'		N/A	English	content word
generalization of an affix into a hypernym driven by mass borrowing	pro	'a person who supports a concept or a principle'		N/A	English	content word
increase in syntagmatic autonomy	=achtig	'typical, similar, like; somewhat, rather; approximately'		noun/NP	Dutch	clitic or phrasal affix
increase in syntagmatic autonomy	=ish	'typical, similar, like; somewhat, rather; approximately'		noun/NP	English	clitic or phrasal affix
increase in syntagmatic autonomy	=nos	'we'		N/A	Mexican Spanish	clitic or phrasal affix
increase in syntagmatic autonomy	muid	'we'		N/A	Irish	function word

Figure 1. Rows 39 to 45 of the database

## 03 Classification of Processes and Preliminary Findings

The “process” parameter in the database aims to capture the transformation observed in each case of attested grammaticalization in minimal terms. These are not broad classificatory labels, but rather short descriptions of an underlying mechanism. The classification was created bottom-up by grouping the examples that seemed to be driven by the same motivation.

- bound entities gaining lexical status
    - generalization of an affix into a hypernym
      - generalization of an affix into a hypernym driven by mass borrowing
    - semantically unpredictable generalization of an affix
    - separation of a wordpart
  - increase in syntagmatic autonomy
  - inflectional marker to derivational marker
    - inflectional marker to grammatical derivational marker
    - inflectional marker to lexical derivational marker
  - derivation from a function word
    - delocutive word-formation
- conversion
    - taboo-driven conversion
  - clipping
    - clipping followed by conversion
  - wordform to lexeme
  - functional expansion
    - functional expansion into a specific context
  - lexical shift
  - modal verb to lexical verb
  - univerbation
  - loss of an inflectional category
    - loss of an inflectional category with traces

While developing the process-based tagging model, several cross-process trends emerged from th data, such as:

- bound items with clearer meanings are more likely to undergo upgrading
- lower degrees of autonomy tend to correlate with lower degrees of semantic content (see also [Dahl 2004: 164] for discussion)
- previously overlooked pragmatic inference and discourse salience play great role in at least some of these processes

## 04 Future Directions

- to use this database to recognise typologically recurring degramamticalization patterns
- expand the coverage by adding degrammaticalization instances from underrepresented language families and areas

### Access the Database

- GitHub: [lililek/the\\_degrammaticalization\\_database](#)
- Contact: [Isfaizeeva@edu.hse.ru](mailto:Isfaizeeva@edu.hse.ru)
- Open for collaboration!

### References

- Dahl 2004 — Ö. Dahl. *The growth and maintenance of linguistic complexity*. Amsterdam/Philadelphia: John Benjamins, 2004.
- Meillet 1912 — A. Meillet. *Introduction à l'étude des langues indo-européennes*. Paris: Hachette, 1912.
- Ramat 1992 — P. Ramat. Thoughts on degrammaticalization // *Linguistics* 30(2), 1992. P. 547–565.