

From qualifiers to quantifiers: semantic shift at the paradigm level

Quentin FELTGEN

Ghent University / Blandijnberg 2 9000 Gent Belgium

quentin.feltgen@gmail.com

Abstract

Language change has often been conceived as a competition between linguistic variants. However, language units may be complex organizations in themselves, e.g. in the case of schematic constructions, featuring a free slot. Such a slot is filled by words forming a set or 'paradigm' and engaging in inter-related dynamics within this constructional environment. To tackle this complexity, a simple computational method is offered to automatically characterize their interactions, and visualize them through networks of cooperation and competition. Applying this method to the French paradigm of quantifiers, I show that this method efficiently captures phenomena regarding the evolving organization of constructional paradigms, in particular the constitution of competing clusters of fillers that promote different semantic strategies overall.

1 Introduction

Language change is often depicted as a competition between an entrenched variant, and an innovative, rising competitor; e.g. the replacement of *of course* by *obviously* (Tagliamonte and Smith, 2021) in Present Day Canadian English, of *werðan* by *becuman* in Middle English (Petré and Cuyckens, 2008), of *moult* by *beaucoup* and *très* in Middle French (Marchello-Nizia, 2000), of *en par dans* as the chief locative preposition in Modern French (Fagard and Combettes, 2013), or of the former syntactic patterns for negation and interrogation by the periphrastic *do* pattern (Kroch, 1989).

Such a competition, however, is difficult to evidence. For instance, what can the *way* construction (Israel, 1996; Perek, 2018) (e.g. 'the Black Prince *plundered his way* eastward to Narbonne and back') possibly replace? Moreover, in the case of a clear replacement, the replacement is seldom total. For instance, the periphrastic *do* did not replace auxiliaries, and some verbs like *need* are still found with the older pattern; *moult* was replaced by two

different words, but they both show uses that *moult* had not, and they do not cover the whole functional range that was carried by *moult*. The French *en* was replaced by *dans* in most locative contexts, but it remains more frequent than its newer counterpart (Eckart and Quasthoff, 2013; Corpus and language statistics for corpora of the Leipzig Corpora Collection, 2021). Similarly, while *be going to* can be seen as a competitor for *will*, both auxiliaries differ semantically; furthermore, it has been argued that they both feature semantic retention pertaining to their respective origins (Nicolle, 1998). The same phenomenon has been observed for discourse markers based on prepositional adverbs in French (Fagard and Charolles, 2018).

In the meanwhile, it has been posited that frequency rise evidenced by lexical items or constructions is a sign of semantic expansion (Feltgen et al., 2017). This hypothesis would provide a convenient account of the phenomena mentioned above: language change is, first and foremost, a semantic shift; if this semantic shift spills over the semantic domain of an existing form, competition arises over this overlap; if this semantic shift leads to meanings and functions that were not formerly expressed in the language, no competition occurs and there is no competitor. In this sense, lexical competitions are a sign of a semantic shift, and help identify ongoing language changes; yet language change may happen without any obvious competition.

In this paper, another perspective on the interplay between semantic change and competition is offered. Indeed, competition may arise *within* a given linguistic form dominion, especially in the case of schematic constructions, that is, constructions that feature an open slot that can be filled by different arguments (e.g. the *way too* + {ADJ/ADV} intensifier construction). These arguments can compete against one another, within the construction. Moreover, this competition needs not be one-to-one: since a large number of arguments are involved,

the competition may unfold between different clusters of arguments. Therefore, we can detect such paradigmatic competitions by looking at the correlations and anti-correlations between the frequency dynamics of the different fillers.

To give a somehow hypothetical example of this semantic shift at the paradigm level, we may consider the paradigm of classifiers (e.g. in Thai or in Korean), which categorizes nouns according to a set of principles. This categorization can, for instance, be driven by considerations of shape, or by considerations of function (Carpenter, 1992). These two broad principles may in principle both co-exist and compete over time, and individual classifiers typically fall into one or the other broad group of function-based classifiers and shape-based classifiers (things being more blurry in practice).

In this regard, we can conceive of three levels of semantic change, that are not mutually exclusive. The first is the constructional level, which corresponds to a significant change in the broad scope of the nouns to which the construction applies, for instance through the recruitment of new classifiers to operate on nouns that were previously beyond the scope of the construction, e.g. the emergence of a new classifier for the class of machines. The second level is that of the individual classifiers; e.g. the reanalysis of Thai /*khan*/ from a shape-based classifier to a function-based one. The third level is what I refer to as the 'paradigmatic level'. This happens when one group takes over the nouns that were classified by the other group. Nouns become then re-classified according to a new set of principles. In this case, the individual classifiers do not have to undergo semantic change; and the broad 'classifiers' construction still applies to the same nouns. Yet, these nouns are now preferentially categorized in a new way (e.g. function over shape); therefore, the properties that are made salient by the choice of a specific classifier are now different. This kind of paradigmatic reorganization is for instance illustrated by the different yet overlapping semantic roles of verb classifiers in the related Nyulnyul and Warrwa languages (McGregor, 2018).

This kind of semantic change only reveals itself at the scale of a system of linguistic units, such as constructional paradigms. As a result, traditional tools, such as word embeddings that rely on collocations (Mikolov et al., 2013), cannot be readily expected to account for it. In this paper, I offer a simple method to detect such a paradigmatic reor-

ganization. I illustrate it on the French quantifier construction *un N de* (e.g. *une profusion de*), which exactly mirrors the English quantifier construction *a N of* (e.g. *a lot of*), whose historical development has already been studied (Traugott and Trousdale, 2013). Besides the entrenchment of the construction in Middle French, I evidence, by looking at the network of correlation between the fillers frequencies, a major paradigmatic shift occurring in Modern French. A qualitative analysis of the competing clusters is also offered.

2 Corpus and frequency profiles

2.1 The French quantifier construction

The French quantifier construction, *un Q(N) de N* ('*a Q(N) of N*'), is a construction in which a nominal quantifier, *Q(N)*, is used to introduce a noun, *N*, by giving an estimate of its overall count. Its structure is closely similar to a more general genitive construction *NP de NP* (similar to the English *NP of NP* from which the English quantifier construction also originates), so that automatically (and even manually) sorting the relevant occurrences from the spurious ones poses a serious challenge. To cope with this difficulty, I decided to be conservative and select only the arguments which are clear quantifiers. One useful test in this regard is that the verb may agree with the quantified noun (plural) instead of agreeing with the quantifier noun (singular), which is incompatible with a genitive reading. For one early example of it (1330): "*Tantost une foule de gent firent cesser leur parlement*" ("A crowd of people interrupted their discussions.")

I have retained 36 quantifier nouns for the *Q(N)* slot of the construction. Partitive constructions, which behave very similarly, have been excluded (e.g. *un morceau de*, 'a bit of'), as well as plural quantifiers, such as *des litres de vin* ('liters of wine'), which don't seem to warrant the same constructional reading.

2.2 Frantext corpus

This construction has been investigated on the Frantext database (ATILF, 1998), restricted to the 1321-2020 period. This is the longest period such that every decade is covered in the corpus with at least one text (with a minimum of 5 for the 1321-1330 decade). The selected corpus encompasses close to 300 M words. In total, I found 50k occurrences of the quantifier construction.

2.3 Frequency profile

For each decade, frequency is based on the number of tokens of the quantifier construction found in the corpus. To obtain such a count, I performed individual queries for all identified fillers, cleaned the results manually when necessary (e.g. removing from *un tas de* occurrences such as '*un tas de sable*', 'a sandpile'), and then aggregated all fillers to get a count for the construction as a whole. Next, this number of tokens is divided by the corpus size - in number of words - associated with the decade. Finally, frequency is smoothed using a moving average over the five previous data points (e.g. the data point for the decade 1801-1810 is actually an average over the whole period 1761-1810).

The frequency profile of the quantifier construction (Figure 1) first features a pattern of latency (low frequency, slowly increasing), followed by an S-curve covering the whole sixteenth century. This pattern is commonly associated with the entrenchment of a linguistic unit (Croft, 2000; Aitchison, 2001; Blythe and Croft, 2012; Feltgen et al., 2017). That the frequency immediately decreases instead of stabilizing is a known phenomenon (Van de Velde and De Smet, 2021), but has not been associated with any semantic account so far. The subsequent and massive frequency rise does not follow a clear pattern, and is mostly due to the individual rise of three leading fillers: *nombre* ('number'), *foule* ('crowd'), and *infinité* ('infinity').

This observation is, in itself, interesting: first, during an entrenchment phase, the frequency rise of the construction is cohesive at the construction level, showing a well-formed S-curve. During this time, the functional scope of the free slot increases, but no individual filler drives the frequency profile of the construction. Next, the frequency profile of the post-S-curve period is dominated by individual fillers which become increasingly dominant over the paradigm. This clearly indicates that frequency of the construction alone, besides the well-established S-curve pattern, is not a reliable indicator of the functional changes undergone by the construction. More likely, after an early period of entrenchment, the frequency profile of the construction as a whole is mostly a by-product of the dynamics of the individual fillers. Therefore, to detect relevant changes at the constructional level, we must turn towards other quantitative measures.

3 Network of interactions

In this section, I present indicators of a major shift within the paradigmatic organization of the quantifier construction, based on the dynamical interactions between the paradigm members.

3.1 Building the network

3.1.1 Correlation matrices

The frequency profiles of the quantifiers are first extracted and computed individually. Next, we can measure the correlations between the different time series. However, it is pointless to compute the correlation over the series as a whole, since the corpus spans too long a time period: the correlation needs to be more local in time. Therefore, I decided to use a time window of 10 decades, which strikes a convenient balance between computing the correlation over a sufficiently long time series for the correlation to be meaningful, and over a sufficiently focused window to efficiently capture change phenomena. It also corresponds to the mean time of entrenchment of a form (Feltgen et al., 2017).

By computing such correlations, we can build matrices $A(t)$ for the time period t (e.g. 1451-1550), whose elements $A_{ij}(t)$ are the Pearson correlation coefficients between the frequencies of forms i and j over the corresponding time period.

Additionally, time series can show spurious correlations if they are driven by a common process, e.g. both individual forms could be driven by the frequency profile of the construction as a whole (Koplenig, 2018). Therefore, we complement this measure with a second correlation matrix, this time between the derivatives of the individual frequencies. For each time period, a matrix $B(t)$ is built for the correlations between the time series of the derivatives in the same way as $A(t)$ had been built. For a different method to compute correlation between the time series of word frequencies, the reader may refer to Koplenig (2017).

3.1.2 Filtering the matrices

At this point, the matrices need filtering, to only capture the interactions that are significant enough. Therefore, a matrix $C(t)$ is introduced, whose elements are 0 everywhere, except when both $A_{ij}(t) > \theta$ and $B_{ij}(t) > \theta$, in which case $C_{ij}(t)$ is set to 1, or when both $A_{ij}(t) < -\theta$ and $B_{ij}(t) < -\theta$, in which case $C_{ij}(t)$ is set to -1. The threshold θ is set to 0.45. The rationale behind this choice is the following. We might want

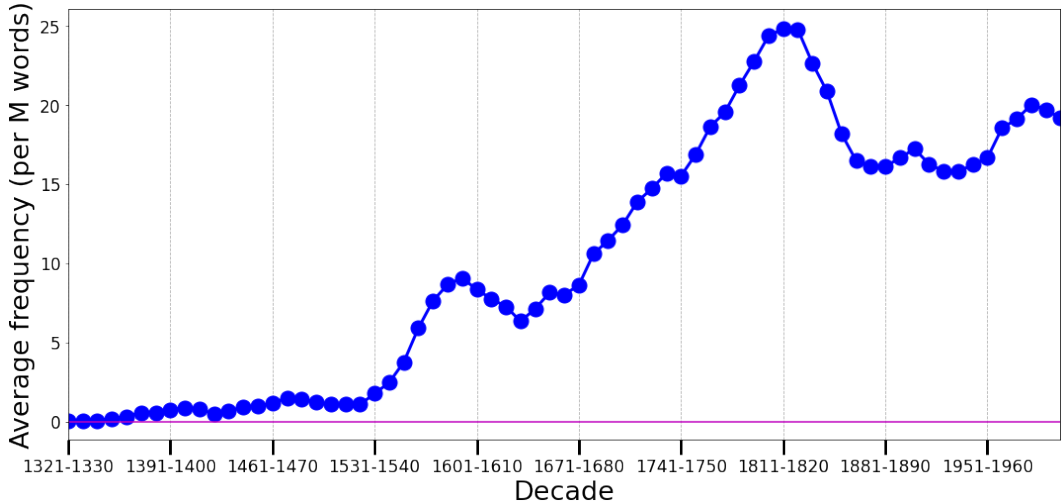


Figure 1: Smoothed frequency, per million words, of the quantifiers construction in the Frantext database.

to choose the threshold θ so that the p -value of observing such a value for the Pearson correlation coefficient is 0.05. However, if A and B are assumed to be independent, then the probability of a false positive for the joint observation of a correlation above the threshold for both quantities is $0.05^2 = 0.0025$, which is a much stricter threshold. Therefore, to set a p -value significance threshold at 0.05 for the joint observation, we must choose a Pearson correlation coefficient threshold over single quantities corresponding to a p -value of $\sqrt{0.05}$. For a Pearson correlation coefficient computed over 10 data points (our moving window covers this many points), this threshold is approximately 0.45.

From the $C(t)$ matrix, the network can be drawn by drawing two kind of links, correlation ones when $C_{ij}(t) = 1$ and anti-correlation ones when $C_{ij}(t) = -1$. In Figure 2, the latter are depicted with a thick extremity pointing towards the form whose derivative is the smaller on average over the time period considered. This way, a network of interactions between the fillers of the construction can be drawn for each time period. Note that the construction as a whole has been included among the inventory of forms, to track which cluster drives its frequency evolution; on the networks of Figure 2, the associated node is labeled 'paradigm'.

3.2 Results

The networks of earlier periods (from 1601-1610 onward) are very sparse and provide little insight into the paradigmatic dynamics. That the frequency increase over the seventeenth and eighteenth centuries is associated with the frequency rise of a few

fillers that behave independently is corroborated by the sparsity of the interactions.

However, from 1791, two competing clusters are clearly emerging. On the one side, the entrenched quantifiers, with *foule* ('crowd'), *multitude* ('multitude') and *infinité* ('infinity'); on the other side, the innovative forms, with specific quantities such as *dizaine* ('dozen'), *millier* ('thousand'), etc. The cluster of entrenched quantifiers still drives the frequency profile of the construction, as evidenced by its correlation with the 'paradigm' node.

For the 1831-1930 and 1851-1950 windows, numerous anti-correlations are found between the members of the two major clusters, hinting at a competition. Since the anti-competition links point toward the members of the entrenched forms cluster, it shows that these forms are in decline relative to the forms belonging to the newly emerging cluster. The absence of anti-correlation links for the 1811-1910 network is however intriguing.

3.3 Further quantitative evidence of the cluster competition

The automated quantitative characterization of the fillers' interactions has evidenced two clusters competing against one another. However, one might argue that the identification of these two clusters only reveals that we have conflated within the same alleged quantifier construction, two separate constructions that are quite disparate in their scope and use. Therefore, we provide additional empirical evidence for the ongoing competition to stress the high level of interaction between the two groups.

Comparing the frequency profiles of the fillers

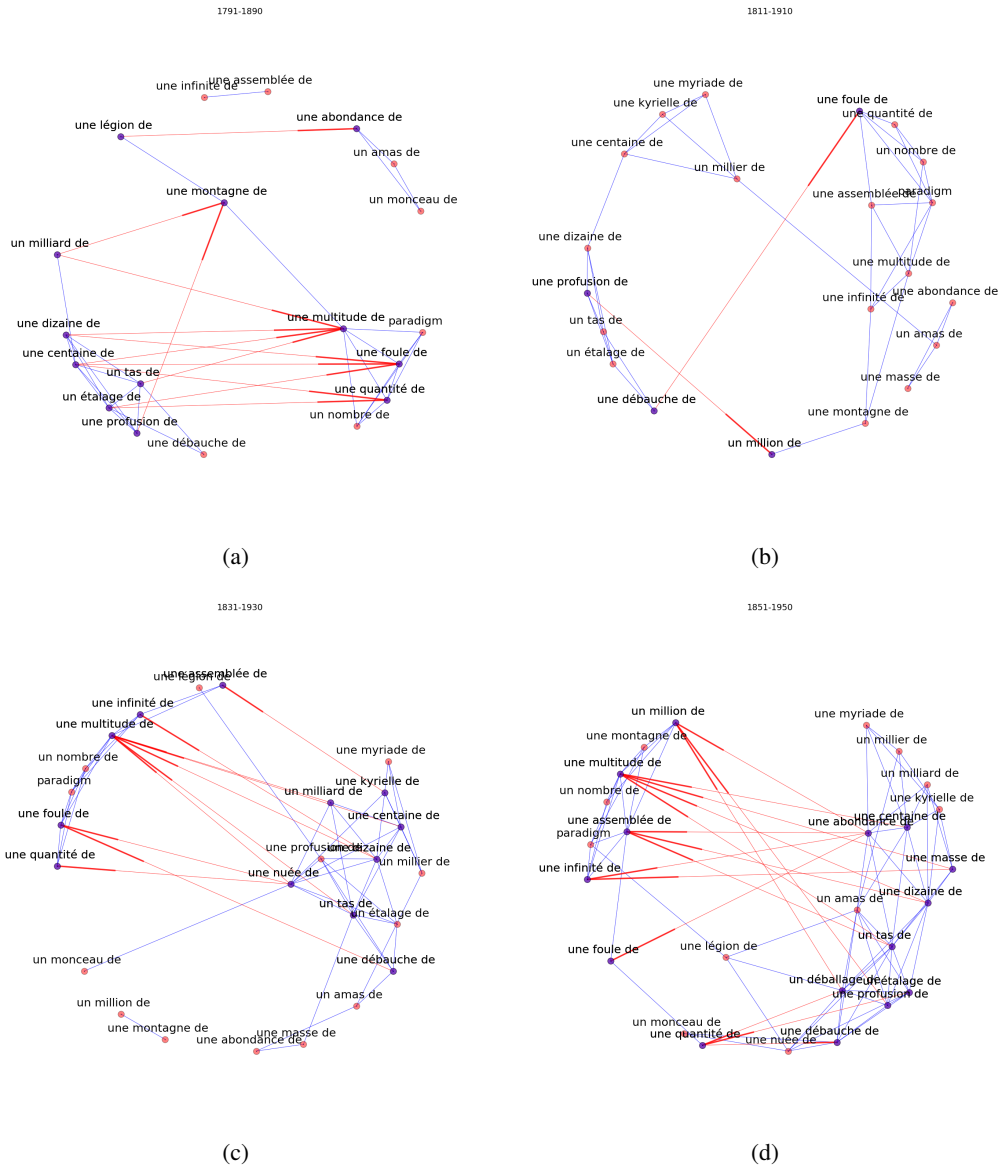


Figure 2: Network of interactions within the quantifiers construction paradigm, for four time periods: (a) 1791-1800; (b) 1811-1910; (c) 1831-1930; (d) 1851-1950. Simple edges (in blue) show correlation between nodes, edges with a wider end (in red) show an anti-correlation, and point towards the declining form.

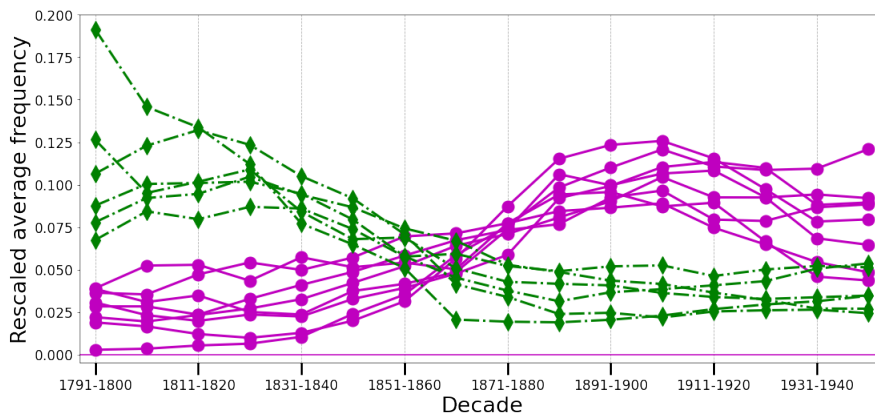


Figure 3: Rescaled frequencies of the fillers belonging to the two clusters identified through the network analysis (diamonds: entrenched members; circles: innovative members).

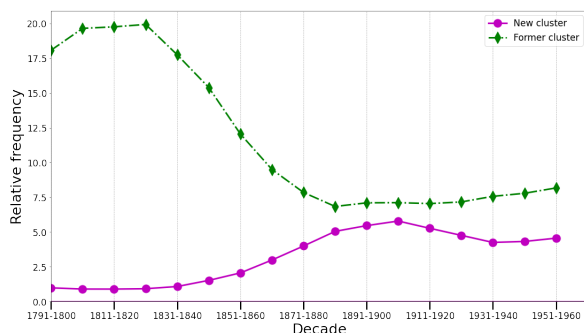


Figure 4: Frequencies, per million words, of the two clusters identified through the network analysis (diamonds: entrenched cluster; circles: innovative cluster).

is not straightforward, because different fillers can show very different magnitudes of frequency, due to the Zipfian structure of a paradigm organization (Ellis and Ogden, 2017). Therefore, we rescale each frequency profile by the mean frequency of the form for the time period under consideration (here 1791-1950, where the competition occurs). The rescaled frequencies shown in Figure 3 are clearly consistent with the competition picture.

If we furthermore plot the aggregated frequencies for each filler (Figure 4), it becomes apparent that the decline of the former cluster and the rise of the newer cluster are concomitant, even though the frequency gain of the innovative cluster does not compensate for the frequency loss of the entrenched one, which is reflected in the frequency decline of the construction as a whole. Interestingly, the outcome of this competition is a coexistence rather than an eviction of the former cluster, which remains dominant in terms of frequency.

4 Linguistic interpretation

Now that we have evidenced an ongoing competition between two different clusters of the quantifiers construction, it is worthwhile to shed light on the extent to which these clusters provide conflicting perspectives on the quantification of things. In order to better understand the functional range of each cluster, we can consider the arguments that are associated with each quantifier - that is to say, the individual paradigms of nouns attached to the single members of the construction.

Quantitative techniques based on word embeddings have been developed to automatically assess how the semantic organization of a constructional paradigm evolves diachronically Hilpert and Perek (2015); Perek (2016, 2018). An even more relevant

quantitative tool to capture the semantic shift of the quantifiers is the cluster characterization based on collexeme analysis (Gries and Stefanowitsch, 2010). Since a range of methods already exist, and their application to this case study would constitute a contribution in itself, yet without any original methodology to offer, I shall remain here at a qualitative level. The more modest goal of this section is therefore to briefly illustrate how the dynamics-based methods outlined above evidence phenomena that make sense from a linguistic point of view. In what follows, I discuss the functional roles of each different fillers by simply looking at their ten most frequent collocates.

4.1 Entrenched cluster semantics

The ten most frequent arguments of the main five quantifiers of the entrenched cluster are listed on Table 1. First, generic arguments, such as *gens* ('people'), *hommes* ('men') and *choses* ('things'), are associated with most quantifiers. Next, *assemblée* ('gathering') and *quantité* ('quantity') immediately stand out from the other three quantifiers. Indeed, the former is almost exclusively used to quantify people, in agreement with its immediate lexical root, while the latter is often associated with uncountable things, like *eau* ('water') and *argent* ('money'), although its representation may be biased by its strong association with scientific texts (*quantité de chaleur* is the French name for heat energy in thermodynamics). In contrast to this, *nombre* is mostly associated with countable arguments, such as *cas* ('cases'), *fois* ('times'), *jours* ('days'), *exemplaires* ('copies'). It is also revealing that *quantité* is associated with *gens* ('people'), while *nombre* is associated with *individuels* ('individuals'): both refers to groups of persons, yet one underlines their indistinction and uncountability, while the other conceives them as separate entities.

The quantifier *foule* is often associated with abstract things, e.g. *idées* ('ideas'), *détails* ('details'), *questions* ('questions'), while *multitude* shows a surprising specialization into generic categories, as with *oiseaux* ('birds'), *insectes* ('insects'), *plantes* ('plants'), *êtres* ('beings'). After *un nombre de*, which remains the most frequent quantifier throughout the studied period, *une foule de* and *une multitude de* are respectively the second and third most frequent quantifiers of the entrenched cluster.

To summarize, the different roles of the quantifiers of this cluster are distinguished by ontological

nombre (number)	multitude	foule (crowd)	quantité (quantity)	assemblée (gathering)
hommes (men)	hommes (men)	choses (things)	chaleur (heat)	hommes (men)
heures (hours)	êtres (beings)	hommes (men)	choses (things)	femmes (women)
années (years)	faits (facts)	gens (people)	eau (water)	notables
exemplaires (copies)	choses (things)	idées (ideas)	gens (people)	gens (people)
fois (times)	idées (ideas)	détails	mots (words)	législateurs (lawmakers)
jours (days)	oiseaux (birds)	questions	lettres (letters)	médecins (doctors)
cas (cases)	objets (objects)	objets (objects)	acide (acid)	créanciers (creditors)
coups (blows)	plantes (plants)	faits (facts)	produits (products)	députés (deputies)
députés (deputies)	gens (people)	cas (cases)	argent (money)	évêques (bishops)
individus (individuals)	insectes (insects)	mots (words)	hommes (men)	poètes (poets)

Table 1: List of the ten most frequent quantified nouns for each of the five most prominent quantifiers of the entrenched cluster. I did not gloss words that are the same as their English counterparts.

considerations; they are sensitive to *what* is quantified. Furthermore, they are all impressively vague regarding the actual quantity of what they quantify: *quantité* ('quantity') and *nombre* ('number') could not be more generic, while *foule* ('crowd') and *multitude* only hints at a 'big' quantity.

4.2 Innovative cluster semantics

The arguments of the innovative cluster of quantifiers are displayed on Table 2. First of all, a lot of quantifiers become available to express the quantity in a pretty precise way: *dizaine*, *douzaine*, *vingtaine*, *centaine*, *millier*, etc. Comparing the arguments for *dizaine* and *centaine* shows that they follow very similar semantic distributions. Their arguments are items that typically need to be counted and quantified. As such, there is a significant overlap with the semantic profile of *un nombre de*.

The quantifier *nuée* ('cloud') draws on metaphoric expansion: *une nuée d'oiseaux* ('a cloud of birds'), *une nuée de flèches* ('a cloud of arrows'), are transparent, while *une nuée de solliciteurs* ('a cloud of petitioners') goes one step further on the metaphorical expansion path, and leans more towards a pure quantifier meaning. Interestingly, the arguments of *nuée* are widely different from those of the other members, while the semantic profiles of the fillers belonging to the entrenched cluster all had a more extensive overlap. This observation also applies in the case of *profusion*, whose arguments, with the exception of *détails*, are not commonly associated with the other quantifiers. Most of the arguments here indicate that *profusion* emphasizes the excess of futile or superfluous items.

Finally, *un tas de* ('a heap of') seems to take on the semantic role of *une foule de*, but its arguments are more disparate. It is remarkable that *un tas de* doubles up on the expression of indetermination;

not only, as a quantifier, it expresses an uncertain quantity, but it also preferentially associates with undetermined arguments such as *choses* and *trucs*. Importantly, *un tas de* is by far the most frequent of the fillers of the innovative cluster, and as an outcome of the competition, it becomes also more frequent than *une foule de*, becoming thus the second most frequent quantifier after *un nombre de*.

The semantic pattern of quantification expressed by this innovative cluster is less clear than that of the entrenched cluster. On the one hand, a large family of quantifiers allows for an accurate assessment of the quantity in which the quantified is found; on the other hand, very specific quantifiers also appear, that no longer highlight which sort of things are quantified, but how the plurality comes to constitute a quantity: a lot of small things that coalesce into a whole (*nuée*), a profligate plethora of frivolities or luxury items (*profusion*), a disorganized collection of assorted stuff (*tas*). In that sense, the pragmatic coloring of these quantifiers is stronger than that of the former cluster.

Crucially, the two clusters offer two very different strategies to delineate the semantic space associated with the quantification: the first cluster focuses on the nature of what is quantified, while the second cluster focuses on how the set of items manifests itself *as a quantity*, leading to more heterogeneous semantic distributions that may span an arbitrary number of ontological categories.

5 Conclusion

Language change, as it unfolds over several, inter-related levels of the linguistic organization, is inherently complex. To understand the diachronic processes that a form or a construction participates in, tracking its frequency soon faces drastic limitations: besides the S-curve pattern of entrenchment, frequency is volatile and extremely variable, for no

profusion (multitude)	tas (heap)	dizaine (dozen)	centaine (hundred)	nuée (cloud)
fleurs (flowers)	choses (things)	années (years)	mètres (meters)	oiseaux (birds)
détails	gens (people)	jours (days)	francs	sauterelles (locusts)
colonnes (columns)	histoires (stories)	mètres (meters)	pas (steps)	moucheron (swat)
chevaux (horses)	bêtises (faults)	minutes	années (years)	solliciteurs (petitioners)
mosaïques (mosaics)	livres (books)	hommes (men)	hommes (men)	moineaux (sparrows)
ornements (ornaments)	monde (people)	pas (steps)	personnes (people)	étincelles (sparks)
couleurs (colors)	trucs (things)	personnes (people)	pieds (plants)	pierres (stones)
dentelles (laceworks)	idées (ideas)	kilomètres (kms)	millions	flèches (arrows)
mets (dishes)	questions	fois (times)	pages	hannetons (cockchafers)
roses	raisons (reasons)	pages	écus (crowns)	copeaux (shavings)

Table 2: List of the ten most frequent quantified nouns for each of the five most prominent quantifiers of the innovative cluster.

evident reason. Here, I have argued that we can achieve a fine-grained understanding of the process by looking at the interactions between the members of a constructional schema. I have offered a computational method to track and visualize such interactions, a method that can be perfected and further automatized, e.g. with the use of clustering algorithms. The picture that emerges remains highly complex, but the large cluster-to-cluster competitions that the method can evidence may lead to a fascinating linguistic insight into the fine-grained processes of language change.

When considering schematic constructions, that is, constructions that can host a variety of fillers, similarly to an ecological niche, change can occur on three levels: 1) on the level of the individual fillers (whenever they undergo functional change); 2) on the level of the construction, typically through the recruitment or loss of new fillers, that is, through changes in its syntactic productivity (Sánchez-Marco and Evert, 2011); and finally 3) in the way the fillers are organized within the construction, that is, on the paradigm level, leading to a new way to categorize the semantic space on which the construction applies. Although all three changes are expected to occur to an extent in a given process, such a process can be better characterized by one or the other of these changes.

All these changes are instances of semantic shifts; while it is clear that the first two changes affect meaning (that of the individual filler in 1, that of the construction as a whole in 2), the third kind of change is less obvious. Yet, as it redefines the categories in which the arguments of the construction are partitioned, it evidences different features of these arguments. This kind of semantic shift is especially prevalent when considering tight categories applying to a broad class of words, such as determiners, classifiers, or auxiliaries. Examples

of this shift are the emergence of French demonstratives (Marchello-Nizia, 2006) or the change in the auxiliaries in Old Spanish (Mateu, 2009).

In our example of the French paradigm of quantifiers, we have shown that the construction underwent a significant paradigmatic change in the 1801-1950 period. This paradigmatic change can be related, at least on a qualitative level, to semantic considerations regarding the logic underlying the different partitioning of nouns by the two competing clusters. Individual change does not seem to play a large role in this picture, even though *un nombre de* is more versatile than the other quantifiers and is likely to have undergone a semantic shift. Constructional change does occur to some extent (the nouns covered by the innovative cluster do not all overlap with the nouns covered by the entrenched cluster). Among the innovative quantifiers, those tied to a precise quantity such as *une douzaine de* ('a dozen') are the more likely to be associated with new nouns, indicative of a semantic opening of the quantifier construction towards a 'measure' meaning, while it was more closely associated to a 'count' one (Unterbeck, 1994). This latter change may also be due to an increasing proportion of scientific texts in the corpus. Yet, the remarkable cluster competition that unfolds throughout the nineteenth century is testimony enough that the paradigm level is the most suited to understand the change phenomenon in this period.

This study is a first step, an invitation to consider more systematically language change from a systemic perspective, especially with the help of automated tools, that are most needed to deal with the intrinsic complexity of these systems. Although the analysis presented here could be refined with the use of a wider range of methods, I hope that the results are intriguing enough to foster further interest in changes unfolding on the paradigm level.

References

- Jean Aitchison. 2001. *Language change: Progress or decay?* Cambridge university press.
- ATILF. 1998. [Base textuelle frantext \(en ligne\)](#).
- Richard A Blythe and William Croft. 2012. S-curves and the mechanisms of propagation in language change. *Language*, pages 269–304.
- Kathie Carpenter. 1992. Two dynamic views of classifier systems: Diachronic change and individual development. *Cognitive Linguistics*, 3(2):129–150.
- Corpus and language statistics for corpora of the Leipzig Corpora Collection. 2021. [Corpus: fra_news_2021 - the most frequent 50 words](#).
- William Croft. 2000. *Explaining language change: An evolutionary approach*. Pearson Education.
- Thomas Eckart and Uwe Quasthoff. 2013. Statistical corpus and language comparison on comparable corpora. In *Building and using comparable corpora*, pages 151–165. Springer.
- Nick C Ellis and Dave C Ogden. 2017. Thinking about multiword constructions: Usage-based approaches to acquisition and processing. *Topics in Cognitive Science*, 9(3):604–620.
- Benjamin Fagard and Michel Charolles. 2018. Ailleurs, d’ailleurs, par ailleurs: De l’espace à l’humain, de l’humain au discours. *Journal of French Language Studies*, 28(3):351–375.
- Benjamin Fagard and Bernard Combettes. 2013. De en à dans, un simple remplacement? une étude diachronique. *Langue française*, 178(2):93–115.
- Quentin Feltgen, Benjamin Fagard, and J-P Nadal. 2017. Frequency patterns of semantic change: corpus-based evidence of a near-critical dynamics in language change. *Royal Society open science*, 4(11):170830.
- Stefan Th Gries and Anatol Stefanowitsch. 2010. Cluster analysis and the identification of collexeme classes. In *Empirical and experimental methods in cognitive/functional research*, pages 73–90. CSLI Publications.
- Martin Hilpert and Florent Perek. 2015. Meaning change in a petri dish: constructions, semantic vector spaces, and motion charts. *Linguistics Vanguard*, 1(1):339–350.
- Michael Israel. 1996. The way constructions grow. *Conceptual structure, discourse and language*, 217:230.
- Alexander Koplenig. 2017. A data-driven method to identify (correlated) changes in chronological corpora. *Journal of Quantitative Linguistics*, 24(4):289–318.
- Alexander Koplenig. 2018. Using the parameters of the zipf–mandelbrot law to measure diachronic lexical, syntactical and stylistic changes—a large-scale corpus analysis. *Corpus Linguistics and Linguistic Theory*, 14(1):1–34.
- Anthony S Kroch. 1989. Reflexes of grammar in patterns of language change. *Language variation and change*, 1(3):199–244.
- Christiane Marchello-Nizia. 2000. Les grammaticalisations ont-elles une cause? le cas de beaucoup, moult et tres en moyen français. *L’information grammaticale*, 87(1):3–9.
- Christiane Marchello-Nizia. 2006. Du subjectif au spatial: l’évolution des formes et du sens des démonstratifs en français. *Langue française*, 152(4):114–126.
- Jaume Mateu. 2009. Gradience and auxiliary selection in old catalan and old spanish. In *Historical syntax and linguistic theory*, pages 176–193. Oxford University Press Oxford.
- William B McGregor. 2018. The history of verb classification in nyulnyulan languages. In *The Diachrony of Classification Systems*, pages 315–351. John Benjamins Publishing Company.
- Tomáš Mikolov, Wen-tau Yih, and Geoffrey Zweig. 2013. Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies*, pages 746–751.
- Steve Nicollet. 1998. A relevance theory perspective on grammaticalization. *Cognitive Linguistics*, 9(1):1–35.
- Florent Perek. 2016. Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics*, 54(1):149–188.
- Florent Perek. 2018. Recent change in the productivity and schematicity of the way-construction: A distributional semantic analysis. *Corpus Linguistics and Linguistic Theory*, 14(1):65–97.
- Peter Petré and Hubert Cuyckens. 2008. The old english copula weorðan and its replacement in middle english. In *English historical linguistics 2006: selected papers from the fourteenth international conference on English historical linguistics (ICEHL 14), Bergamo 21-25 August 2006*, pages 23–48. John Benjamins Publishing Company.
- Cristina Sánchez-Marco and Stefan Evert. 2011. Measuring semantic change: The case of spanish participial constructions. In *Proceedings of Quantitative Investigations in Theoretical Linguistics*, pages 79–83.
- Sali A Tagliamonte and Jennifer Smith. 2021. Obviously undergoing change: Adverbs of evidentiality across time and space. *Language Variation and Change*, 33(1):81–105.

Elizabeth Closs Traugott and Graeme Trousdale. 2013. *Constructionalization and constructional changes*, volume 6. Oxford University Press.

Barbara Unterbeck. 1994. Korean classifiers. In *Theoretical issues in Korean linguistics*, pages 367–385. Center for the Study of Language (CSLI).

Freek Van de Velde and Isabeau De Smet. 2021. Markov models for multi-state language change. *Journal of Quantitative Linguistics*, pages 1–25.