French -age suffixation versus verb to noun conversion: quantitative approaches on surface and underlying properties

Alice Missud, Florence Villoing
Université Paris Nanterre, MoDyCo (UMR 7114)

1 Introduction

There are many rival morphological schemas that derive nouns from verbs using the same kinds of semantic operations. In French, the most extensively studied rivalry is between *-age* and *-ment* suffixations (*battage / battement* 'beating' (of the heart), *décollage / décollement* 'unsticking'). However, another productive rivalry that has remained unexplored is that of *-age* suffixation and verb to noun conversion (henceforth N-*age* and VN, respectively) (for example, the doublets in (1)). This will be the focus of our study.

(1) *accrochage / accroche* 'hook, hanging', *jetage / jet* 'throwing', *pesée / pesage* 'weighing', *rééquilibrage / rééquilibre* 'rebalancing'

In recent years, the work on rivalry has taken the turn of the new approaches in morphology against a representation of morphological relations by means of unidirectional rules and in favor of a surface-to-surface perspective (for example, Burzio 2002, and Hathout 2009, 2011, Roché 2011, Plénat 2011 for work on French) according to which the lexicon itself originates from a set of constraints involving several levels of linguistic analysis. This work led to the recognition that some rival suffixes organize their competition around a certain linguistic specialization.

While studying the productivity of English rival suffixes, Lindsay & Aronoff (2013) identified that a less productive suffix (*-ical*) is able to compete against a more productive one (*-ic*) thanks to its ability to select bases under certain conditions that only apply in its particular case. Thus, if phonological constraints apply for both suffixes and favor the shortest one, the disadvantaged suffix will find a way to remain productive by specializing in a different way, on a phonological, morphological, semantic or pragmatic level. We propose to study the rivalry between V→N conversion and *-age* suffixation from this perspective. Hence, a first point to note is that, compared to *-age* suffixation, V→N conversion does not add any new element to the selected base. Therefore, size constraints (Plénat & Roché 2003; Plénat, 2009) might play a role to a lesser extent on V→N conversion than on *-age* suffixation. In that case, on average, we expect *-age* derived noun bases to be shorter than converted noun bases. On a semantic level, V→N conversion mostly derives action nouns, as well as result, agent, instrument and locative nouns (Tribout, 2010), whereas *-age* suffixation prototypically constructs action and result nouns (Namer, 2009). Therefore, V→N conversion seems less restricted in base selection and deverbal interpretation than *-age* suffixation. In that case, we would expect the latter to become specialized in order to remain productive.

Two quantitative approaches can help reveal some of the constraints that might explain the coexistence of rival schemas. The first one, introduced by Arndt-Lappe (2014), considers analogy as an explanatory mechanism. Using Analogical Modeling (Skousen, 2002) on English *-ity* and *-ness* derivatives, Arndt-Lappe shows that *-ity* suffixation specialized in selecting adjectives that display different phonological patterns from those that select *-ness* on the surface level. The second one, proposed by Bonami & Thuilier (2019), uses logistic

regression to observe the statistical effects of multiple predictors in the selection of a base. In their study, these predictors are underlying properties identified for each base in the case of French *-iser* and *-ifier* rivalry. As these approaches have proven their worth, we propose to reinvest both of them in our study. We hypothesize that the interaction of surface level properties and underlying constraints could help shed light on what motivates the coexistence of rival V→N conversion and *-age* suffixation.

2 Predictors

We mapped several predictors reflecting relevant constraints on multiple linguistic levels in order to use them as variables in our statistical model.

Stem length: We hypothesize that *-age* suffixation will favor short stems because of the ability of V→N conversion not to add any exponent to the selected base.

Lexical aspect of the verb: Although the lexical aspect of the base verb is not necessarily transmitted to the derivative (Haas et al. 2008), *-age* derivatives can be expected to preferably select activity, accomplishment or achievement verbs, since they usually denote action or result nouns. V→N preferences in that matter need further investigation.

Argumental structure: The bases of *-age* derivatives were identified as prototypically agentive, sometimes ergative or unaccusative (Ferret et al. 2010). However, no preference was identified for V→N conversion. A constraint on the argumental structure of the base might apply for *-age* suffixation. As agentivity can be found amongst dynamic verbs that are frequently used in the first person (usually denoting an animated subject) (Lapraye, 2017 with *-age / -ment*), measuring the frequency of the use of the first person could possibly help to detect agentivity for a base verb automatically. Furthermore, following Hathout (2009) and Bonami & Strnadová's (2016) proposals in favor of a paradigmatic analysis of morphological schemas, we hypothesize that the presence of an agent noun in the morphological family of a verb could also be a predictor for agentivity.

Verb group: Lapraye (2017) has shown that *-ment* suffixation is able to select verbs from the 2nd and 3rd group whereas *-age* suffixation almost exclusively selects verbs from the 1st group. Such preferences need to be investigated in our case.

Reflexivity: As observed in the case of *-age / -ment* (Fradin, 2014), reflexivity of the base verb can sometimes be a criterion to distinguish two connotations of a single verb that derives doublets (*emballer* 'to wrap' → *emballage* 'wrapping', *s'emballer* 'to get excited' → *emballement* 'hype'). We aim to test if the same applies for our case, without taking doublets into account.

Surface form: Finally, surface form could also play a role. If analogical mechanisms are operating, phonological clues could favor one or the other suffix if it happens to be recurrent amongst multiple lexemes derived using the same morphological process.

3 Data and methodology

Because we wanted to test highly frequent derivatives and see the strongest constraints emerge, our hypotheses were tested using quantitative methods on a dataset limited to 200 derived nouns consisting of 100 converted nouns (from Tribout, 2010) and 100 *-age* derivatives extracted from frWaC (Baroni et al. 2009, 1.6 billion words). We selected derivatives with a frequency range from 211 000 (most frequent) to 800 (less frequent) as documented in frWaC. Doublets were excluded, except when frequencies were extremely unequal (*chauffage* 'heat' → *chauffe* 'warm up'). The corresponding verbs were paired manually.

Predictors were annotated for each base-derivative pair. Stem length, group and argumental structure of the base verb were annotated manually. The presence or absence of an agent noun in the morphological family was identified manually from Démonette (Hathout & Namer, 2014). We used Nomage (Balvet et al. 2011) and Haas & Marin's annotated lexicon (in preparation) to annotate the lexical aspect of some of the verbs; missing verbs were annotated manually. The frequency of use of the 1st and 3rd persons and the frequency of reflexive forms were extracted automatically from frWaC, dependency-parsed with Mind The Gap (Coavoux, 2017).

4 Experiments and results

4.1 Logistic regression

Binomial logistic regression was used to observe the relation between a dependent variable (V→N conversion / *-age* suffixation) and 6 variables. The 1st / 3rd person ratio was used as a normalized quantitative variable. The other 5 variables are categorical: verb group (1st, 2nd or 3rd), transitivity of the verb (transitive, labile, inaccusative or inergative), lexical aspect of the verb (state, activity, accomplishment or achievement), reflexivity of the verb (more or less than 5 occurrences) and the presence or absence of an agent noun in the morphological family of the derivative.

Our model showed that, compared to V→N conversion, *-age* suffixation significantly favors short stems (1 or 2 syllables, except for *paramétrer* 'to set' that has 3 syllables) and verbs that do not belong to the 3rd group. A slight preference for accomplishment verbs and non-reflexive verbs was also identified for *-age* suffixation. Longer stems are favored by V→N conversion (mostly 2 but up to 4 syllables), and even though it has a preference for verbs of the 1st group, verbs of the 2nd and 3rd group are selected as well (~30% of the data). The area under the curve (or accuracy) of the model attained 87%.

4.2 Analogical Modeling

Analogical Modeling was given each phonemes of a base at the infinitive form (except for the last one that indicates the verbal group) as categorical variables and applied commutation on each of them, in order to match other bases in the dataset and assimilate the base to either V→N conversion or *-age* suffixation. *-age* suffixation obtained an F-score of 66% whereas V→N conversion reached 57%. Predictions were slightly better than the ones presented in a random baseline (F-score of 57% for *-age* suffixation and 45% for V→N conversion). Although some phonemes can assist classification, the analogical gangs did not reveal any regular pattern of association in the surface form. Examining the matching phonemes did not provide insight into the formal properties of verbs that can distinguish V→N conversion and *-age* suffixation, except that verbs selected by *-age* suffixation tend to gather more in analogical niches.

5 Discussion

The underlying properties tested using logistic regression helped in the identification of semantic, morphological and phonological restrictions applying to *-age* suffixation. However, nothing significant came from transitivity, the 1st / 3rd person ratio and the presence of an agent noun in the morphological family of the derivative. The hypothetical correlation between reflexivity and agentivity of a verb needs further investigation. Combining these results with an analysis of the surface properties of the base verbs using Analogical Modeling was unable to reveal any relevant formal specialization for *-age* suffixation. These results can

be explained by the small size of our dataset. As a next step, investigating analogical mechanisms on a semantic level using distributional methods could eventually help reveal the different semantic domains that are invested by converted and *-age* suffixed nouns.

References

Arndt-Lappe, S. (2014). Synchronic and Diachronic Analogy in Suffix Rivalry: The Case of -ity and -ness. English Language and Linguistics, 18(03), 497-548.

Balvet, A., Barque, L., Condette, M.-H., Haas, P., Huyghe, R. et al.. (2011). La ressource Nomage. Confronter les attentes théoriques aux observations du comportement linguistique des nominalisations en corpus. Traitement Automatique des Langues, ATALA, 2011, Ressources Linguistiques Libres, 52 (3), pp.129-152.

Baroni, M., Bernardini, S., Ferraresi A. & Zanchetta E. (2009). The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. Language Resources and Evaluation 43(3), 209-226.

Bonami, O. & Strnadová, J. (2016). Derivational paradigms: pushing the analogy. 49th Annual Meeting of the Societas Linguistica Europaea.

Bonami O. & Thuilier J. (2019). A statistical approach to affix rivalry in lexeme formation: French *–iser* and *ifier*. *Word Structure*, pp. 4-41.

Burzio, L. (2002). Surface-to-Surface Morphology: when your Representations turn into Contrasts. In P. Boucher (ed.), Many Morphologies (142-177). Somerville, MA: Cascadilla Press.

Coavoux, M. (2017). Discontinuous constituency parsing of morphologically rich languages. Thèse de doctorat, Université Paris-Diderot.

Ferret, K., Soare, E., & Villoing, F. (2010). Rivalry between French –age and -ée: the role of grammatical aspect in nominalization. In M. Aloni, H. Bastiaanse, T. De Jager, & K. Schultz (eds.), Lecture Notes in Computer Science: Vol. 6042. Logic, Language and Meaning, 17th Amsterdam Colloquium, The Netherlands, December 2009, Revised Selected Papers (pp. 284–295). Berlin: Springer.

Fradin, B. (2014) Deverbal nominalization and the 'Means' interpretation. Paper read at International Morphology Meeting (IMM16), Budapest, May 29-June 1, 2014.

Haas P., Huyges R., Marin R. (2008) « Du verbe au nom: calques et décalcages aspectuels ». In *Actes du Congrès Mondial de Linguistique Française,* pp.2039-2053.

Hathout, N. (2009). Contributions à la description de la structure morphologique du lexique et à l'approche extensive en morphologie. Toulouse : Université de Toulouse le Mirail. Mémoire d'Habilitation à diriger des recherches.

Hathout, N. (2011) « Une approche topologique de la construction des mots : proposition théoriques et application à la préfixation en *anti-*». In : M. Roché, G. Boyé, N. Hathout, S. Lignon et M. Plénat, *Des unités morphologiques au lexique,* Hermes-Lavoisier, Paris, pp.251-318.

Hathout, N. & Namer, F. (2014). Démonette, a French derivational morpho-semantic network. Linguistic Issues in Language Technology 11(5): 125-168.

Lapraye, A. (2017). Une approche statistique de la concurrence entre procédés constructionnels : la dérivation en -age et en -ment en français. Mémoire de Master, Université Paris-Diderot.

Lindsay, M. & Aronoff M. (2013). Natural selection in self-organizing morphological systems. In: F. Montermini, G. Boyé, J. Tseng (eds.): Morphology in Toulouse: Selected Proceedings of Décembrettes 7. Munich: Lincom Europa, pp. 133–153.

Plénat M. (2009) « Les contraintes de taille ». In : Fradin B., Kerleroux F. & Plénat M. (éds). *Aperçus de morphologie du français.* Saint-Denis : Presses Universitaires de Vincennes, 47-64.

Plénat M. (2011) « Enquête sur divers effets des contraintes dissimilatives en français ». In : M. Roché, G. Boyé, N. Hathout, S. Lignon et M. Plénat, *Des unités morphologiques au lexique,* Hermes-Lavoisier, Paris, pp.145-190.

Plénat, M., Roché M. (2003) "Prosodic constraints on suffixation in French". In: Booij G., DeCesaris J., Ralli A. & Scalise S. (eds). *Topics in Morphology. Selected Papers from the Third Mediterranean Morphology Meeting (Barcelona, September 20-22, 2001).* Barcelona, I.U.L.A.: Universitat Pompeu Fabra, 285-299.

Roché M. (2011) « Quel traitement unifié pour les dérivations en *-isme* et en *-iste* ? ». In : M. Roché, G. Boyé, N. Hathout, S. Lignon et M. Plénat, *Des unités morphologiques au lexique,* Hermes-Lavoisier, Paris, 69-143.

Skousen, R., Lonsdale, D. W. & Dilworth, B. (2002). Parkinson (eds). Analogical modeling. Amsterdam: John Benjamins.

Tribout, D. (2010). Les conversions de nom à verbe et de verbe à nom en français, Thèse de doctorat, Université Paris Diderot-Paris 7.