

Derivational Morphology and Proper Names

Rafael Antonio Gutiérrez Martínez
Universitat Pompeu Fabra

Proper names are often combined with derivational affixes to create new words whose meanings are related in some way to the proper name's referent. However, the problem with this kind of derived word is that it is not easy to predict the (exact) meaning (or meanings) it will take. Consider, for instance, the Spanish word *rafaelista* (*Raphael-ist* from Raphael Sanzio). What is the meaning of this form? The morphology provides some hints about its possible meanings. It suggests that the word, in addition to a use as a relational adjective (*obra rafaquista*, 'work by Raphael'), can have several interpretations as a noun: It can refer to (i) a specialist in the work of Raphael, (ii) an admirer of Raphael, (iii) a member of an organization associated with Raphael or (iv) a follower of the ideas of Raphael (RAE-ASALE, 2009, 6.9a, f, k; 7.7h). But which of these interpretations is adequate is not clear *a priori*, nor is it clear which one is unlikely or even inadmissible. A (slightly simplified) adaptation of the analysis of relational adjectives in Arsenijević, et al. (2014) that treats *rafaquista* by saying that it has a logical translation like $\lambda x[Rel(x, \mathbf{Raphael})]$, where x is human and Rel is a contextually-specified relation, fails to give us the set of possible meanings that the morphology suggests for it provides a too weak interpretation (for example, being Raphael's neighbor could in principle be a possible value for Rel but does not seem like a valid value). For obvious reasons, this analysis also fails to tell us which interpretations among those that the morphology licenses are probable and which ones are improbable (in the Corpus del Español: Web/Dialects (Davies, 2016) *rafaquista* is never used with meanings (iii) and (iv), which might suggest that these meanings are unlikely for that word¹). A semantic theory of word meaning should aspire to say something about this².

In this work, I address this problem; to do so, I follow some ideas suggested by McNally et al. (2023). Building on the proposal of Erk & Herbelot (2024), these authors propose that developing a theory that predicts word meaning in context requires considering the probabilistic interaction of two kinds of knowledge: linguistic knowledge (from morphology) and world knowledge (from the word stem). How do these ingredients help to face the problem? *World knowledge* here means *scenario knowledge*. Scenarios are chunks of information and knowledge about situations and everyday events in the world (Sanford & Garrod, 1998). Specifically, they are groupings of concepts corresponding to entities and happenings that often appear together in everyday situations (Fillmore, 2006) and belong together (Erk & Heberlot, 2021). For example, for a situation like the sale and purchase of bakery products, the corresponding scenario consists of a system that includes concepts for the production activity (BAKING), what is made (BREAD), who does it (BAKER), and where it happens (BAKERY) (example taken from McNally et al. (2023)). Information of this kind helps to narrow down the interpretations of words in the following way. Suppose that a person who does not know the meaning of the Spanish word *panadero* ('baker') listens to that word in a given situation. Two

¹ It is important to emphasize that the claims made here and in what follows are restricted to the data found in the Corpus del Español: Web/Dialects. Uses of the word *rafaquista* with meanings (iii) and (vi) could, for example, be found in other sources. Although this is relevant in the long run, for the time being, what is of most interest is to examine the plausibility of the theoretical ideas considered here, for which the data chosen and analyzed are sufficient. Empirical improvements (like considering as well other corpora) can be implemented later once the theoretical ideas prove to be successful for a specific amount and type of data.

² It is worth noting that this problem is more acute for words derived from proper names than for words derived from common nouns. One reason for this is that in the former case, there is more indeterminacy about the semantic content of the word with which the derived word starts than in the latter.

cues are given to her in this context. First, the presence of the stem *pan* gives her access to the baked goods making-and-selling scenario that includes the concepts of BAKING, BREAD, BAKER and BAKERY. Second, the presence of the suffix *-ero* allows her to infer that the word names an entity of a specific type (in this case, an individual that sells or makes something). These two pieces of knowledge allow her to interpret the word's meaning probabilistically and predict that *panadero* is related to the person who sells or makes bread and denotes the baker.

To use these ideas to treat the semantics of words derived from proper names, we first need to determine what kinds of scenarios we can associate with proper names. The problem here is that even if we believe that knowledge of a proper name involves knowledge about some properties of the proper name's referent (if we, for example, adopt the notion of entity representations (Kamp, 2015) to theorize about proper names and their derived forms), it is clear that these properties can be of any kind and constitute whatever kind of information we have about the referent. Thus, we do not have a set of concepts with the feature of being highly salient in experience that can be *a priori* selected to be the concepts of scenarios associated with proper names. To address this issue, I analyzed 277 Spanish derived words of 40 names from 6 different classes (anthroponyms; toponyms; and names of organizations, human creations, brands, and temporal periods) collected from El Corpus del Español: Web/Dialects (Davies, 2016) to see if some generalizations could be found. From this analysis, I observed, first, that 9 different sorts of entities are systematically introduced through derivations from proper names (intellectual movements, the referent itself, qualities, actions, individuals, appreciations, periods of time, historical events, and units of language) and, second, that those entities appear to form clusters in some ways depending on the general class of the referent. For example, in my dataset, names of artists (e.g., *Dante*) and human products (e.g., *Gioconda*), when combined with derivational morphemes of different sorts (*-esco* and *-ina*), tend to create words that introduce qualities (*dantesco* and *giocondina*) and not words that introduce other types of entities. Likewise, names of artists rarely take affixes (like *-ismo*) that form words referring to intellectual movements (*rafaelismo*, for instance, is not found).

We then need to adapt the probabilistic semantic framework for polysemy resolution proposed by Erk & Heberlot (2024) to word meaning.³ Informally stated, the idea is the following. I assume (1) that entities of the sort identified in my corpus study are part of the scenarios for proper names, (2) that the knowledge of the morphology provides the set of possible semantic sorts of each derived word and (3) that the clusterings of interpretations mentioned in the last paragraph are what set the probability of the entities of the scenario for each derived word, which in turn determines the interpretation of that word. Thus, all derived words from proper names are potentially associated with scenarios that include the same sorts of entities, but depending on the class of the referent of the name some entities will receive more or less probability. To take an example, consider again the word *rafaelista*. The 9 sorts of entities mentioned above are potentially part of the scenario associated with this word. However, given that *Rafael* refers to an artist and names of artists tend to introduce qualities and avoid introducing intellectual movements, qualities receive a higher probability, and intellectual movement receive a lower probability. This, in turn, predicts that *rafaelista* will more likely refer to a property related to Raphael and not to a follower of Raphael's ideas.

³ In this framework, probabilities are understood in the Bayesian sense as degrees of subjective belief or reasonable expectations.

Overall, the present work shows that integrating scenario knowledge and morphology in distinct ways provides promising insights to face the problem of the semantics of derivation from proper names that can be operationalized.

References

- Arsenijević, Boban; Gemma Boleda; Berit Gehrke; and Louise McNally. 2014. Ethnic adjectives are proper adjectives. R. Baglini, et al. (eds.) *Proceedings from the 46th Annual Meeting of the Chicago Linguistic Society*, 17-30.
- Davies, Mark (2016). *Corpus del Español: Web/Dialects*. Available online at <http://www.corpusdelespanol.org/web-dial/>.
- Erk, Katrin and Aurélie Herbelot (2024). *How to marry a star: probabilistic constraints for meaning in context*. In: *Journal of Semantics*, pp. 1-35
- Fillmore, Charles (2006). "Frame Semantics". In: *Cognitive Linguistics: Basic Readings*. Berlin: De Gruyter Mouton.
- Kamp, Hans (2015). "Using Proper Names as Intermediaries Between Labelled Entity Representations". In: *Erkenn* 80, pp. 263-312.
- McNally, Louise; Bonami, Olivier; Paperno, Denis (2023). "It's time for a complete theory of partial predictability in language". To appear in *Theoretical Linguistics*, vol. 50.
- RAE and ASALE (Real Academia Española/Asociación de Academias de la Lengua Española. (2009). *Nueva gramática de la lengua española*. Madrid: Espasa Calpe.
- Sanford, Anthony J. and Simon C. Garrod (1998). "The role of scenario mapping in text comprehension". In: *Discourse Processes* 26.2-3, pp. 159–190