Symbolic Learning of Rules for Semantic Relation Types Identification in French Genitive Postnominal Prepositional Phrases

Hani Guenoune, Mathieu Lafourcade
University of Montpellier, LIRMM
161 Rue Ada, 34095 Montpellier, FRANCE
{hani.guenoune, mathieu.lafourcade}@lirmm.fr

Abstract

We are interested in the semantic relations conveyed by polylexical entities in the postnominal prepositional noun phrases form "*A de B*" (A of B). After identifying a relevant set of semantic relations types, we proceed, using generative AI, to build a collection of phrases, for each semantic relation type identified. We propose an algorithm for creating rules that allow the selection of the relation between A and B in noun phrases of each type. These rules correspond to selecting from a knowledge base the appropriate neighborhood of a given term. For the phrase "*désert d'Algérie*" carrying the *location* relation, the term "*désert*" is identified as a geographical location, and "*Algérie*" as a country. These constraints are used to automatically learn a set of rules for selecting the *location* relation for this type of example. Rules are not exclusive as there may be instances that fall under multiple relations. In the phrase "*portrait de sa mère - the portrait of his/her mother*", all of *depiction*, *possession*, and *producer* types are a possible match.

Keywords: Genitive, Postnominal prepositional noun phrases, Semantic relations

1. Introduction

Beyond the necessity of identifying polylexical entities for automated language analysis, it is important for various applications to also understand the nature of relation binding the different components of polylexical terms. Our focus is on the genitive case of the nominal complement "de N" in French. In other words, compound noun phrases (NP) formed through the use of the preposition "*de*" introducing a syntactic complement B to a head A in "A de B", where A and B are nouns (and variations "*A d'B*", "*A du B*", etc.). We aim to automatically identify the semantic relation between terms A and B. Generally, such an approach contributes to a richer interpretation of discourse in textual content and leads to better semantic representations. Among the tasks benefiting from this study, we can mention the Question Answering task (Kapanipathi et al., 2020; Ben Abacha, 2012), which requires a rich semantic representation of text and relations between mentioned entities. Another task is the resolution of anaphors triggered by possessive determiners which involves the transformation of genitive forms into anaphoric phrases ("*le vélo de Julie* $\rightarrow$ *son vélo*"), and in which the resolution is based on constraints of the relation type between the anaphor and its antecedent (Guenoune, 2022).

In our project's specific context, these efforts also contribute to consolidating a common-sense knowledge base, firstly through the identification of semantic relation types relevant to our various inference mechanisms and Natural Language Processing (NLP) applications that leverage the knowledge base. Another way of improving the quality of the knowledge base is to develop a classification system that serves as a control tool. The analysis of the correctness of the results of such a tool would bring insights into the overall quality of the knowledge used. In order for this to be possible, an emphasis has to be put on the explainability of the methods to use, at is essential in identifying potential gaps and highlighting appropriate ways of consolidating the knowledge base.

In "A de B," the nominal head (A) plays a crucial role in maintaining the underlying sense with its complement (B). The type of nouns linked by the preposition in a genitive construction conditions the semantic relations that bind them. Regarding nominal possessives such as "*John's friend*" (which translate to a "A de B" NP in French (*l'ami de John*), the distinction mentioned in (Barker et al., 2019) differentiates the use of *sortal* nouns from (two-place) *relational* nouns. The contrast in their definition is analogous to that between *unary* and *binary* predicates in first-order logic. The *relationality* of a noun (De Bruin and Scha, 1988) concerns whether a referent must be identified in relation to another entity. In constructing nominal phrases, relational nouns only make sense when related to exactly two arguments, as seen in the example of *family/kinship* names *[father, mother, sister, brother...]* considered as archetypal relational nouns (see Löbner classification system (Löbner, 2011)).

The semantic relation in the phrase "*la mère de Lucie - the mother of Lucie*" relates solely to the sens of the nominal head (the *relational* noun "*mère - mother*"). Thus, a genitive noun phrase with a *relational* nominal head allows for a *lexical interpretation* of the semantic relation (De Bruin and Scha, 1988), contrasting with pragmatic readings that re-

The Workshop on Cognitive Aspects of the Lexicon, pages 32–41

sult from the use of certain *sortal nouns*, requiring extra-lexical information to identify the nature of the semantic link in postnominal NPs. Mentioning "*le nuage de Lucie - Lucie's cloud*" requires the introduction of pragmatic elements to fully grasp the type of relation (*the cloud she was looking at, drawing, dreaming of, etc.*).

The meanings conveyed by this type of NPs are therefore diverse, even though automatic interpretation efforts often reduce the types of semantic relations to one of member-collection/possession types (one explanatory element for this simplification could be the importance of theses relations and their predominant role in the standard semantic typologies considered in NLP works).

Beyond the typological framework of nominal heads, specific nouns (whether relational or sortal) introduce a multitude of possible relations between the terms of the NP. This work aims to study the nature of semantic links in this configuration and proposes a semantic typology for these links. Furthermore, we introduce a symbolic learning algorithm which serves as a basis for an explainable system of classification of semantic types in genitive NPs. We note that since figurative meanings are revealed to be challenging to determine computationally, this work does not address the identification of the overall meaning of the form "*A de B*" when it has acquired an idiomatic/figurative sense (*e.g.* "*homme de paille / écran de fumée - straw man / smoke screen*") .

This paper covers several aspects:

1. Proposal of a typology of semantic relations in postnominal genitive phrases, followed by the creation of an associative corpus between examples of genitive constructions and corresponding relation types. Data are collected using a generative AI, cautiously validated by hand. A portion of the corpus serves as training data while the rest serves as a test set.

2. Introduction of *GRASP*-it, a learning algorithm calculating decision rules for probable relation types.

3. Evaluation of the quality of produced constraints by implementing a second algorithm for classifying semantic types in "A de B" forms.

We begin with an overview of the resources used in this project, namely the knowledge base that we seek to improve through this work, then the data used in developing *GRASP*-it. We also provide examples of each relation type resulting from the use of "de N" constructions. We then describe the learning mechanism implemented to synthesize semantic relations between terms of each type. Finally, we conduct an evaluation of the quality of the produced rules applied to a portion of the corpus.

2. Data

This project required the use of external resources to successfully carry out the study in general, the learning process, as well as the evaluation of the learned rules and the classification system.

2.1. Knowledge Base Used

The world knowledge that supports our study and the algorithms developed is built from the latest issue of the *JeuxDeMots* (*JDM*) project data collection (dated February 11, 2024) (Lafourcade, 2024).

JeuxDeMots (*JDM*) (Lafourcade, 2007) is a lexical-semantic network represented by a directed graph. Graph nodes represent terms, while arcs signify typed, weighted, and potentially annotated relations between terms. The graph tackles lexical polysemy by specifying hierarchical sense "*refinements*", where a specific sense is affiliated with the general sense of the term. *JDM* is based on practical tools, principles, and concepts (*e.g.* the notion of refinement, the diversity of semantic types, inverse relations, such as *r_isa* and *r_hypo*, and a series of collaborative tools). The *JDM* network is designed to be used as a knowledge support for AI solutions (semantic text analysis, reasoning, decision-making, automatic summarization, etc.). A weighting and symbolic valuation system (meta-information annotation, *e.g.* *rare, relevant, non-relevant*, etc.) has been implemented to facilitate graph traversal and its exploitation (Lafourcade and Le Brun, 2023). As of February 1, 2024, *JDM* contained approximately $540\ million$ relations between over $9\ million$ terms and $24\ million$ nodes.

One central challenge for this project is to enrich our knowledge base with semantic information, particularly information regarding relations in genitive prepositional noun phrases. This helps text analysis and knowledge extraction. Indeed, When encountering a genitive form in a text it is desirable to guess the relation(s) between A and B.

2.2. Corpus of Genitive Constructions

We present a small-sized corpus for the learning and evaluation of semantic type determination rules. This corpus is to be seen as a starting point for the creation of larger scale collections. Despite the significance of small corpora (Landragin, 2018), below several thousand examples, it reveals challenging to apply resource-intensive procedures such as neural learning algorithms. However, we aim to integrate this effort into a longer-term project where data augmentation procedures can be implemented, such as automatic semantic enrichment mechanisms or manual annotation completion.

In the following, we detail semantic types identified, then discuss the data acquisition and validation protocol. In order to avoid introducing any bias, we chose to collect data from a source independent of the *JDM* knowledge base.

2.2.1. Semantic Typology

In Table 1, we list the considered semantic types along with an explanation and examples, and the corresponding relation type in *JDM* (with the appropriate orientation, where relations with names in the form '*r_x-1*' denote the conversive relation of '*r_x*').

Relation Type	***JDM* Relation**
Consequence (Co): *Term A is a consequence* *of (caused by) term B.*	*r_has_causatif*
dégâts de la tempête - retards de la circulation *(EN) storm damage - traffic delays*	
Possession (Po): *A is possessed by B*	*r_own-1*
fusil du soldat - vélo du cycliste *(EN) the soldier's rifle - the cyclist's bike*	
Material/composition (M): *Term A is composed of* *or is made of material B.*	*r_objet>matière*
cuillère de bois - trône de fer *(EN) wooden spoon - iron throne*	
Origin (O): *Term A originates from* *the location B.*	*r_lieu>origine*
vin de France - café du Brésil *(EN) wine from France - Coffee from Brazil*	
Topic *(T)*: *Term A has term B* *as its theme (or subject).*	*r_topic*
restaurant de sushis - film d'horreur *(EN) sushi restaurant - horror movie*	
Quantification (Q): *A serves as a measure for B.*	*r_quantificateur*
brin d'herbe - minute d'attente *(EN) blade of grass - waiting minute*	
Depiction (D): *Term A is a depiction* *of term B.*	*r_depict*
peinture d'un paysage - photo d'une famille *(EN) painting of a landscape - photo of a family*	
Characterisation (Ca): *Term A is a property* *or the noun of an adjective* *that can qualify term B.*	*r_has_property-1*
sournoiserie du politicien - sagesse du viellard *(EN) the politician's cunning - the elder's wisdom*	
Holonymy(H): *Term A is part* *of term B.*	*r_holo*
coque du bateau - écaille du poisson *(EN) boat hull - fish scale*	
Location (L): *Term A can have term B* *as its location.*	*r_lieu*
tour de Pise - sahara d'Algérie *(EN) tower of Pisa - sahara of Algeria*	
Agent (A): *Term A is an action* *in which the actor is term B.*	*r_processus* *_agent*
travail de l'ouvrier - cours du professeur *(EN) worker's job - the teacher's class*	
Patient (P): *Term A is an action* *that term B undergoes.*	*r_processus* *_patient*
travail du bois - ouverture de la porte *(EN) woodworking - opening of the door.*	
Instrument (I): *Term A is an instrument of B* *or an action that B undergoes.*	*r_processus* *>instr-1*
clé d'ouverture - clé de la porte *(EN) opening key - key of the door*	
Kinship/Social tie (LS): *A plays a role of 'A'* *in relation to B.*	*r_social_tie*
avocat d'une femme battue - chef du groupe *(EN) lawyer for a battered woman - group leader*	
Producer (AC): *A is produced by B.*	*r_product_of*
portrait de Van Gogh - gâteau du pâtissier *(EN) Van Gogh's portrait - pastry chef's cake*	

Table 1: List of semantic types considered and correspondences with semantic relations types in the lexical-semantic network *JDM*.

It should be noted that this particular list is the one we have established as basis for our study. Choices regarding granularity and the number of types were made to align this typology with the requirements of the resources and tools we use, as well as the needs of the applications that will take advantage of this typology. The number and type of semantic relation is thus very knowledge-base directed and strictly semantically expressed rather than pragmatically expressed as was the case with previous examples for English documented in the literature as state of the art (Nastase and Szpakowicz, 2003). It is by no means an exhaustive list of all possible types of relations between terms A and B in a "A de B" form.

Some types may be added to this list. It is also possible to specify/generalize certain types so that they more or less precisely correspond to theoretical frameworks that are different from ours. This is especially relevant when another knowledge base (than *JDM*) is used, potentially one defining a different set of relation types. Among these types, we can mention examples involving absolute temporal semantic relations (carried by class names), such as *repas de midi - brise du matin - bus de nuit* (*midday meal - morning breeze - night bus*) or spatial and temporal relative links (carried by relational nouns) such as *milieu/droite/gauche de la pièce - bas de page* (*middle/right/left of the room - bottom of the page*). Another case is that of nominations: "*Théorème de Pythagore - Râle de Wallace - Kappa de Fleiss*" (*Pythagoras' theorem - Wallace's rail - Fleiss' kappa*) which could be the subject of a separate category. For the mentioned cases, we choose to include them in types of similar semantics; for example, the first two cases are included in the "topic" type, while the case of nominations is classified among instances of "author/creator" (even though it may not necessarily involve a creation *per se*).

2.2.2. Data Collection and Validation

For each type of semantic relation mentioned above, we employed a generative AI (LLM conversational agent[1]) to generate a set of examples. We limited each type to 80 examples, with 50 dedicated to *training* and 30 for *test* purposes.

The strategy for constructing queries to the conversational agent varied depending on the types of relations. Obtaining satisfactory examples proved more or less challenging, depending on the cases. For types where the generated examples were less exploitable, we chose to guide the model through examples. We provided about ten examples composed by us, then explained commonalities at the level of underlying semantic relations over several conversational turns with the chatbot.

Although this iterative approach yielded examples of the desired type, it had the drawback of excessively "influencing" the responses generated by the chatbot, resulting in a set of polylexical entities with low diversity (strong alignment with the examples presented to the agent). Consequently, for the sake of diversification, *after an initial generation of examples*, we emphasized the need to diversify the generation in subsequent queries. This strategy was repeated until we considered the set convincing. However, the generated instances still contained approximately 10% misclassified or duplicated examples and remained imperfect in terms of diversity. Therefore, we conducted a manual validation of all examples produced by the chatbot. Specifically, the validation involved replacing duplicated cases and very similar entities, as well as reclassifying misclassified examples.

2.2.3. Data Formatting and Post-Processing

The produced collection includes examples of variable morphology. Concerning the presence or absence of a determiner for the nominal complement (term B), one might assume that a morphological normalization step would be beneficial. However, this criterion constitutes a morpho-syntactic marker that can be crucial for classification, which is why we choose not to perform morphological or lexical transformations. Nevertheless, it should be noted that polysemy of corpus terms may be observed. Therefore, exploiting the corpus will require data preparation, specifically a phase of semantic disambiguation to select the appropriate senses of the terms.

3. Presentation of *GRASP*-it

The *GRASP*-it (*Genitive Relations And Semantic Pattern Identification Tool*) algorithm aims to produce a set of constraint pairs for each type of relation based on input data. These constraints are based on the semantic types of the nominal head and the complement. They can be considered a synthesis of semantic attributes regarding the content of a knowledge base. The purpose of this set of constraints is to guide a classification process of semantic relations in genitive NPs. Another objective is to produce "*interpretable*" constraints that can easily be read and explained. In general, the first step of *GRASP*-it involves storing, for each example of a certain type, semantic information that could allow to classify the example in the relevant type:

- *Hypernyms* of terms A and B: The goal is to capture, as precisely as possible, the semantic *"types"* of both terms. An hypernym is a *term* (lexical entity) in *JDM* attainable through the relation *r_isa*.
- *Target for Relation Types* (*TRT*): A selection of relation types leading to the term. For example, a term frequently targeted by the *location*

[1] ChatGPT. Model version gpt-4-0613. 2023-06-13

relation is considered, by this approach, as a location. This enables the reinforcement of the relevance of this semantic class for a specific term. The selection of relations leading to the terms can be viewed as a means to supplement the list of hypernyms for a given term.

- *The Standard Semantic Type* (*SST*): through the relation *_INFO-SEM*, the standard type associates a lexicalized term with a standard ontological (conceptual) type.

The result of this step is a set of weighted pairs, referred to here as *signatures* of terms A and B. The number of pairs at this stage corresponds to the number of examples for each type, which, in the case of our corpus (training portion), amounts to 50 NP units of the form "A de B."

A signature is defined as an unordered set of symbols. Each symbol takes a value of a specific entry of JDM. For example, the signature s associated to the term "*véhicule*" would be as follows.

```
s(véhicule) = {
véhicule, transport urbain, partie de l'espace, Transport urbain ,
mode de transport, instrument, lieu,
transport, moyen, machine, moyen de
transport,

r_isa, r_hypo, r_has_part, r_holo,
r_agent, r_patient, r_lieu, r_instr,
r_carac-1, r_lieu-1, r_action_lieu,
r_mater>object, r_processus>agent,
r_own, r_is_instance_of,

_INFO-SEM-SUBST, _INFO-SEM-THING-ARTEFACT, _INFO-SEM-PLACE, _INFO-SEM-THING-CONCRETE, _INFO-SEM-PLACE-HUMAN
}
```

For clarity's sake, we divided the symbols into three blocks: *Hypernyms*, *TRT*s and *SST*. It's worth noting that the signature of a term contains the term itself, this aims to capture instances that are hyponyms of the signed term.

In addition to the need for its explainability, this representation of terms is designed to be controllable in terms of its content and size. This allows the *GRASP*-it method to be adaptable to the variable requirements of the application for which it is used.

The second step aims to aggregate *rules* of each type to process the entire set by generalization. As shown in (1), we define a *rule* R as a pair of constraints s_L and s_R (which are signatures, respectively *left* and *right* corresponding to terms A and B) and a semantic relation type rt.

$$R :< s_L, s_R, rt > \tag{1}$$

The aggregation is a *fusion* operation of two rules and is defined in (2).

$$Fusion(R1, R2) =< s1_L \cup s2_L, s1_R \cup s2_R, rt >^{2} \tag{2}$$

A fusion of two rules means that the constraints they respectively associate are sufficiently *similar* to be represented by a single pair of constraints. Formally, as a signature s can been seen as a vector, we adopted the *cosine similarity* (dot product divided by the product of norms), denoted as sim. Two signatures are considered sufficiently similar when their sim value is above a threshold of 0.5 (which has been set empirically). The merged signature is the vector sum of the two signatures (which corresponds to the union set).

A pair produced by one or more successive fusions is considered more general and *reliable* than a pair that has not undergone fusion. Reliability is therefore a measure of coverage of examples of the type and is calculated by assigning a weight to the pair of constraints corresponding to the number of fusions performed to arrive at the final form of the pair. At the output of this step, a set of more or less aggregated pairs of constraints, with a cardinality at most twice the number of examples of the considered type, is assigned to each type of relation (listed in Table 1).

The idea behind merging rules is that the result of successive fusions is a rule that *represents* appropriately a large set of examples of a certain type. A merged rule can thus be considered as a generalised *model* for a given relation type. One relation type can be associated to several models. A good *model* will appropriately associate the relation type between two terms A and B in a genitive NP.

4. Evaluation of *GRASP*-it

In this section, we present the conditions under which our evaluation was conducted and conclude with the performance scores of our system.

4.1. Data Preparation

A minimal phase of data preparation has been undertook before applying the classification algorithm (detailed in section 4.2). We identified two main tasks that have to be done prior to classification in order to take the entirety of the corpus into account.

Compound Words Identification: instances of NPs containing multiple prepositions "*de*" as in "*lunettes de soleil de marque - détecteur de fumée de protection*" raise the problem of choosing the right separating preposition. This has a direct influence on the identification of the terms A and B,

[2] Only two rules with the same rt can be merged.

hence the appropriate relation type to be identified. We proceeded to identify these instances by verifying their existence in the knowledge base. In "*lunettes de soleil de marque*", the candidates for terms A and B would be "*(**A:** lunettes, **B:** soleil de marque)*" and "*(**A:** lunettes de soleil, **B:** marque)*". In the former, the nonexistence of term B in *JDM* allows us to assign the latter candidate's values to A and B. In case both candidates result in known A and B terms, the intended separating preposition is manually annotated.

Generic Named Entities: Instances containing named entities such as person's first or last names are only well represented in our knowledge base when these entities are renowned or a common knowledge in popular culture (e.g. "*Coca-Cola*" - "*Lucie*"). Therefore, it is important to take into account nonexistent instances by carrying out transformations that replace the name by another (of the same type) that we know is well represented in the knowledge base.

4.2. Application Algorithm

To evaluate the pairs of semantic constraints (rules) produced by the learning algorithm, we implement a validation process that seeks to verify the satisfaction of these constraints. The goal is to identify the semantic types derived from the portion of the corpus not involved in the constraints/rules calculation. This involves 450 examples distributed evenly across the 15 possible types of relations (30 each).

4.2.1. Decision Criteria

Conceptually, the constraints validation approach is based on searching a similarity of semantic types between the terms of the input phrase and the set of terms from which the *GRASP*-it system was trained. The idea is to *induce* an *identity* of types if a test's terms are sufficiently *similar* to the semantics synthesizing the inputs of the learning process.

In practice, the search is conducted by calculating a similarity between terms A and B of the input and the respective signature in all the rules $< s_L, s_R, rt >_i$ learned by *GRASP*-it. The two obtained similarities (for term A with s_L and B with s_R) are aggregated by an arithmetic mean. Therefore, the similarity between a form "A de B" and a rule (pair of constraints) $< s_L, s_R, rt >$ is given in (3).

$$^1/_2(sim(s(A), s_L) + sim(s(B), s_R)) \quad (3)$$

One should note that a *signature* for a term and a *constraint* of a rule share an identical structure.

A positive response is returned for the type best ranked in terms of average similarity values with each pair. Note that for the ranking to be done, the verification procedure is carried out for all types and calculated pairs of constraints (once the training, and all possible fusions have been performed for all possible types).

It remains possible to only use the rules that are either the result of a fusion, or that have not been merged. That is to say, we could exclude rules that have been merged into a new rule. We expect that this *trimming* of the set of rules would lead to a shorter execution time without highly degrading the results (this particular point is subject of a study, see experiment 3).

4.2.2. Extra-Semantic Features

The detection of certain types depends more or less on extra-semantic markers, such as the use or not of a determiner, the use of named entities, or the definiteness of nominal complements. An example illustrating this is the phrase "*photo de famille*" as opposed to "*photo d'une famille*", in which the presence or absence of the determiner, conditions the interpretation of the semantic link (*topic* or *depiction*, *respectively*).

Such heuristics are not part of the core component of our solution, as we tackle the problem of highlighting the implication of *semantic* rules especially. Nevertheless, since the representation of terms is as simple as a set of symbols, the solution proposed seems also suitable for taking into account extra-semantic information. This amounts to the inclusion of the relevant traits into the signatures. Therefore, we proceeded to integrate the *definiteness trait* of complements (B) to our representations. This constitutes a separate study in the subsequent section (see experiment 2).

4.3. Evaluation Protocol

In this evaluation, we will conduct three distinct experiments, aiming to assess the following aspects.

Experiment 1: We seek to assess individually the quality gained from every semantic trait included in the signatures. Our baseline in this experiment is the use of *Hypernyms* only. The idea is to proceed contrastively and analyse the cases that we could successfully classify by adding separately the TRT, and SST traits.

Experiment 2: The inclusion of morphological markers could reveal beneficial to the classification process. Without dedicating *ad hoc* elaborate heuristics for these traits, we experiment the effects of constructing signatures with non-semantic traits, namely the *definiteness* trait. We use a straightforward symbolic approach discussed in 4.4.2.

Experiment 3: Here, we are interested in estimating the overall value for additional computational effort (which translates into a cost in execution time for example). The developed system is also to be used as a component of specific *NLP* tasks (the most important among them being relation extraction from texts or anaphora resolution). It is important to study the feasibility of integrating *GRASP*-it seamlessly in such systems. In these applied systems, the requirements for sub tasks often concern computational complexity and cost of execution. Therefore, this experiment is designed to study the gain/loss in performance relative to computing time in two different settings.

In order to maintain equivalence between the number of examples for each class, we do not consider cases of multiple classification in this evaluation. Regardless of the correctness of additional (fortuitous/incidental) predictions, these NPs were not anticipated as belonging to that (extra) class and are therefore not counted in the number of instances. Moreover, the number of these cases of multi-classifications is not predictable, nor evenly distributed across the types we consider.

4.4. Results

In the following, given a type of relation, we consider precision (P) of a class as the proportion of examples for which the class is correctly predicted relative to all instances predicted as belonging to that class. Recall (R) represents the ratio of examples for which the class is correctly predicted to all actual instances of that class.

4.4.1. Experiment 1 - Semantic Traits

An approach solely based on the semantics of A and B yields the results illustrated in Table 2. In order to assess separately the performance gain allowed by every trait included in the signature of terms, we report the results of 4 contrastive setting combinations of *GRASP*-it.

Setting	P (%)	R (%)	F1 (%)
H	67,3	65,9	65,3
H+SST	70,4	69,7	69,1
H+TRT	77,6	77	76,7
H+TRT+SST	**78**	**77,3**	**77,2**

Table 2: Average precision P (%) , recall R (%) , and F1 (%) scores achieved by the different settings of *GRASP*-it.

The components combined in these settings consist of the information stored while computing the signatures (thus, during the learning of rules): Hypernyms (*H*), Target for Relation Types (*TRT*), and Standard Semantic Types (*SST*).

Firstly, as a baseline, Hypernyms being lexical-semantic traits lead to an average *F1* score of $65,3\%$ which we deem satisfactory considering the potential sparsity of certain terms' hypernyms (e.g. we draw attention to terms A of the types *Characterisation (Ca), Agent (A) and Patient (P)* being all abstract entities for which it is delicate to identify un lexical hypernym). Secondly, we observe that both *SST* and *TRT* traits lead to notable improvements, the non-linear gain tends naturally to become smaller as the scores improve. *H* and *TRT* traits together on one hand and the *SST* on the other, are complementary as they each address specific description needs. *SST* provides the covering of standard typologies (e.g. providing the *_INFOSEM-THING-ABSTRACT* type that helps in the scenario discussed above), while *TRTs* and *H* (being terminological entries of lexical nature) bring forth a finer granularity made possible by an abundance in terminology.

With an F1 score of $77,2\%$, the most favourable setting is the one combining all semantic traits. Table 3 reports the results of the *H+TRT+SST* setting for every considered semantic type.

Type	P	R	F1
Origin (O)	100	86	92
Social Link (LS)	83	100	91
Holonymy (H)	78	86	82
Quantification (Q)	82	80	81
Agent (A)	71	93	81
Depiction (D)	88	73	80
Material (M)	78	83	80
Instrument (I)	77	80	78
Location (L)	84	70	76
Topic (T)	68	86	76
Patient (P)	74	76	75
Producer (AC)	76	73	74
Consequence (Co)	76	63	69
Possession (Po)	65	63	64
Characterization (Ca)	71	50	59
Average	**78**	**77,3**	**77,2**

Table 3: Percentages (%) of Precision (P), recall (R), and F1 scores achieved by the best semantic setting of our system (*GRASP-it*$_{(H+TRT+SST)}$), for every considered type of semantic relations.

We observe generally high results, though disparities exist depending on the type of relation being identified.

Specifically, the low recall for the *Characterization* (*Ca*) relation can be attributed to its limited representation in the database (approximately $10,000$ relations compared to the *Holonymy* (H) relation, which has over $10\ million$ relations), resulting in a low proportion of correctly annotated learning examples. The same applies to test examples (up to half of the cases). Additionally, in the case of

(*Ca*), generics are also poorly represented and often associated with a sense of the term that is not a property (see Section 4.5). It is noteworthy that there is maximum recall for the *social link* type (*LS*) carried by *relational* nominal heads, which allows for a lexical interpretation of this type and is well-synthesized by the created constraints. The Origin type (*O*) is precise due to its small set of general rules (a large number of rules could be merged), but it fails for examples that are not well-represented in the corpus. This case is interesting because instances of the Origin relation type are nearly nonexistent in *JDM* (29 relations), with Hypernyms facilitating the synthesis of effective rules.

In many respects, the followed protocol seeks to assess the model rigorously; indeed, the scores should be interpreted as a baseline to be improved through various treatments of classification processes using *GRASP*-it rules. It is worth noting the absence of morpho-syntactic heuristics (extra-semantic). Additionally, instances that were deemed erroneous (according to the corpus) but are actually valid for the predicted type are counted as errors. A multi-label evaluation would likely elevate the scores for each type (for example, the F1 score of the least well-classified type (Ca) would increase to 76%), however such an evaluation would require manual annotation.

4.4.2. Experiment 2 - Definiteness

Setting	P	R	F1
H+TRT+SST	78	77,3	77,2
H+TRT+SST+DEF	**80,3**	**79,8**	**79,5**
Origin (O)	100	90	95
Social Link (LS)	85	100	92
Holonymy (H)	81	100	90
Instrument (I)	88	80	84
Quantification (Q)	86	83	84
Material (M)	83	83	83
Depiction (D)	85	80	82
Location (L)	85	76	80
Agent (A)	67	96	79
Producer (AC)	79	76	77
Topic (T)	70	86	77
Patient (P)	72	70	71
Consequence (Co)	76	63	69
Possession (Po)	76	63	69
Characterization (Ca)	71	50	59

Table 4: Scores after inclusion of the definiteness trait (*GRASP-it*$_{(H+TRT+SST+\textbf{DEF})}$), for every considered type of semantic relations compared to average scores in the solely semantic setting.

When considering definiteness, we include in the signature of the term B two distinct symbols corresponding to the presence (*resp.* absence) of a definite or indefinite determiner (*Det*, *NoDet*, *Def*, *NoDef*). One special case concerns named entities where none-withstanding a *NoDet* the *Def* attribute is "*forced*". Some examples : *chat du rabbin* => *Det* + *Def*: - *écran de cinéma* => *NoDet* + *NoDef* - *tableau de Chagall* => *NoDet* + *Def*.

As seen in Table 4, taking into account the definiteness trait improves the overall F1 score (compared to the exclusively semantic setting). Nevertheless, we observe some variability in improvement, and in two particular cases (agent and patient) a decrease in performance. The reason for this is that the trait of definiteness isn't typical of one unique type, rather of a subset of types. Its inclusion helps when deciding between two types for which the definiteness trait is decisive. On the contrary, it brings some confusion between types within the same subset of rules (for which the definiteness is not decisive).

4.4.3. Experiment 3 - Rules Trimming

In table 5, we highlight the differences in terms of the number of rules applied ($\#$**r**) and execution times (T) for the *Trim* and *No Trim* settings, which correspond respectively to a complete set of rule and a reduced one. The reduced rule set contains only rules that have not been used for a fusion operation. (T) is the duration for 450 test instances.

Setting	P	R	F1	$\#$r	T (s)
Trim	79,6	77,6	77	49	*25.42*
No Trim	80,36	80	79,8	1384	*92.78*

Table 5: Effects of the reduction of rules on performance scores and total execution times.

The execution time depends on the number of rules and on the size of the signatures. The more fusion occurs, the longer the signatures are. The trim setting leads to a tremendous improvement in computing time (a bit less than 4-fold) with a slight decrease in quality. This means that, as expected, systems that require responses in particularily low execution times could benefit from trimming the set of rules without suffering a significant degradation of the overall performance.

4.5. Analysis of Failure Cases

As a reminder, this work also serves as a tool to control the content of the knowledge base, and helping consolidate appropriate types of relations. The potential gaps in knowledge are highlighted through the analysis of the algorithm's *failure of classification cases*. Among the failure cases, approximately 75% of occurrences are directly attributed to the polysemy of term A and/or term B (combined). For instance, term A in "richesse du royaume" is interpreted in the sense of an *object/artifact* rather than the property of *wealth*, leading the system to classify it as a *location* (an object can be found in a place). Excluding cases of multiple classes (such as "*ombre d'un arbre*" or "*travail du réalisateur*", classified respectively as *(Ca)* and *(AC)*, and predicted by the system as *(D)* and *(A)* (which is also correct). The reasons of remaining errors include knowledge deficiencies (stemming from gaps in JDM) as seen in "*restaurant de cuisine végétalienne*" (where term B was unknown). The use of multiple prepositions is the cause of some failures ("*restaurant de fruits de mer*") due to misidentification of terms A and B (which is not always trivial). Furthermore, a few skill deficiencies (stemming from lack of processing) account for a small number of cases. Specifically, the management of semantic attribute dispersion in *JDM* across various morphological variants of a term: it happens, for intentional/valid reasons or due to the lack of knowledge, that some relations have not been propagated to all derivations of a given term. Regarding our system, this deficiency can be observed particularly at the level of the standard semantic type (SST). In the case of "*liste de films*" (*list of films*) supposed to be classified under the Quantification (Q) type, the standard types of the lemma "film" were missing in its plural variant. Although this failure case proved useful in indicating the state of the database concerning this type of relation, a normalization phase (specifically, a lemmatisation phase) would have circumvented this dispersion case.

5. Conclusion

We have presented the results of a study on genitive noun phrases in the form "A de B". The contributions of this work involved defining a non-exhaustive typology of semantic relations between terms A and B, producing a small corpus of examples annotated by these types, and proposing a learning algorithm for classification rules in the form of aggregated and semantically ordered constraint pairs. The goal is to achieve automatic classification of types that can be used in semantic text analysis and semantic lexicon consolidation.

The insights gained from this project span theoretical and applied domains. Firstly, we recognize the challenge posed by the possibility of including examples in multiple different types. Furthermore, although type identification heavily relies on semantics, it is not the sole criterion for classification decision-making. This implies that a high-performing algorithm should include a series of treatments, including data preparation and analysis (syntactic analysis, compound word identification, polysemy management, etc.).

The adopted symbolic approach has the advantage of being easily explainable. However, its results depend on the types considered, the quality of the data (balance of *weak/strong signals*, good representation of cases), and the richness and quality of the world knowledge base used.

Finally, the fact that rules are not mutually exclusive (and, in some cases, almost identical) presents a challenge that could not be resolved *a priori* – even for a human speaker – without sufficient context, such as in the examples: "*présentation de l'élève - portrait de Van Gogh - travail du ciment*" (*student presentation - Van Gogh portrait - cement work*) (*resp. could be one of : (A)/(P) - (A)/(D) - (A)/(P)*).

As with any corpus construction, the issue of distribution balance and coverage of traits/patterns in the learning set arises. This issue is particularly pronounced for a relatively small dataset. The question of coverage is challenging: intuitively, one might think that increasing the number of examples is necessary. However, the coverage of relation types in genitive forms follows a power law, meaning there are numerous specific cases in the long tail. These cases are difficult to quantify as they may correspond to figurative forms and are often non-prototypical examples (quasi-hapaxes) from a signature perspective. Additionally, prototypical forms corresponding to the beginning of the power law distribution are often abundant (for example, an impressive number of "<animal> of <place>"). Increasing the number of training examples with nearly identical signatures (represented by the same pattern model) has little long-term effect on learning quality. Furthermore, given the computational cost constraint posed by the need to integrate the solution into applied systems, it would be counterproductive to add examples related to an already known pattern. However, one could consider a strategy that exploits a fusion criterion (whether a rule is fusible or not) through incremental learning.

On the other hand, regarding test examples, increasing the number of instances may prove to be useful. We can expect, for a larger test set, a more or less significant decrease in performance as uncertainty increases, but this would allow for a better assessment and thus provide a more accurate idea of the system's quality. In the same vein, it would be useful to establish an evaluation that would assess the real impact of the *GRASP*-it results on applied NLP systems. Looking ahead, potential extensions of this work might include enriching the corpus through the application of our algorithm. A larger collection would allow a thorough evaluation and enable the utilization of various methods, particularly those requiring very large numbers of instances, such as neural learning approaches.

Distribution

Both the data collection and the algorithm are available at (Guenoune and Lafourcade, 2024). For demonstration and experimentation purposes, a webpage hosting a sample implementation of *GRASP*-it is available. A number of settings is proposed, namely the inclusion or not of the traits (*H, TRT, SST and DEF*), the trimming of rules and the setting of the fusion threshold are also possible. Both the training and the test can be done either with the proposed corpus or with another collection. Furthermore, it is not necessary for the semantic types to be restricted to those defined in this study (more or less types can be defined and used within the webpage). However, since the signatures built for this implementation are based on our knowledge base and its structure (the *JDM* structure), we should note that if another data collection is used, it is necessary to ensure that the examples comply with the requirements discussed in the *data preparation* section (4.1).

6. Bibliographical References

Chris Barker, Claudia Maienborn, Klaus von Heusinger, and Paul Portner. 2019. Possessives and relational nouns. *Semantics-noun phrases and verb phrases*, pages 177–203.

Asma Ben Abacha. 2012. *Recherche de réponses précises à des questions médicales: le système de questions-réponses MEANS*. Ph.D. thesis, Paris 11.

Jos De Bruin and Remko Scha. 1988. The interpretation of relational nouns. In *26th Annual Meeting of the Association for Computational Linguistics*, pages 25–32.

Hani Guenoune. 2022. *Résolution des anaphores dans la communication électronique médiée: heuristiques et apport d'informations de sens commun*. Ph.D. thesis, Université de Montpellier.

Pavan Kapanipathi, Ibrahim Abdelaziz, Srinivas Ravishankar, Salim Roukos, Alexander Gray, Ramon Astudillo, Maria Chang, Cristina Cornelio, Saswati Dana, Achille Fokoue, et al. 2020. Leveraging abstract meaning representation for knowledge base question answering. *arXiv preprint arXiv:2012.01707*.

Mathieu Lafourcade. 2007. Making people play for lexical acquisition with the jeuxdemots prototype. In *SNLP'07: 7th international symposium on natural language processing*, page 7.

Mathieu Lafourcade and Nathalie Le Brun. 2023. Apport du jeu pour la construction de connaissances: le projet jeuxdemots. *Technologie et innovation*, 8(4).

Frédéric Landragin. 2018. Étude de la référence et de la coréférence: rôle des petits corpus et observations à partir du corpus mc4. *Corpus*, 2(18).

Sebastian Löbner. 2011. Concept types and determination. *Journal of semantics*, 28(3):279–333.

Vivi Nastase and Stan Szpakowicz. 2003. Exploring noun-modifier semantic relations. In *Fifth international workshop on computational semantics (IWCS-5)*, pages 285–301.

7. Language Resource References

Guenoune, Hani and Lafourcade, Mathieu. 2024. *GenitiveNPS.txt.zip*. repository hosted at jeuxdemots.org. PID https://www.jeuxdemots.org/rezo-GEN1.php.

Lafourcade, Mathieu. 2024. *20240211-LEXICALNET-JEUXDEMOTS-FR-NOHTML.txt.zip*. repository hosted at Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier – LIRMM. PID https://www.jeuxdemots.org/jdm-about.php.