Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 5912–5918
Marseille, 11–16 May 2020

FrSemCor: Annotating a French corpus with supersenses

L. Barque[1], P. Haas[2], R. Huyghe[3], D. Tribout[4], M. Candito[5], B. Crabbé[5], V. Segonne[5]
[1] Université Paris 13 & LLF, [2] Université Paris 13 & Lattice, [3] Université de Fribourg
[4] Université de Lille & STL, [5] Université de Paris & LLF
{lucie.barque, pauline.haas}@univ-paris13.fr, richard.huyghe@unifr.ch
delphine.tribout@univ-lille.fr, {marie.candito, vincent.segonne}@linguist.univ-paris-diderot.fr

Abstract

French, as many languages, lacks semantically annotated corpus data. Our aim is to provide the linguistic and NLP research communities with a gold standard sense-annotated corpus of French, using WordNet Unique Beginners as semantic tags, thus allowing for interoperability. In this paper, we report on the first phase of the project, which focused on the annotation of common nouns. The resulting dataset consists of more than 12,000 French noun tokens which were annotated in double blind and adjudicated according to a carefully redefined set of supersenses. The resource is released online under a Creative Commons Licence.

Keywords: semantic annotation, supersenses, corpus

1. Introduction

This paper presents the construction of a French corpus whose nouns are manually annotated with semantic classes, hereafter "supersenses". The project aims at compensating for the lack of semantically annotated data that can be used in both NLP and linguistics research. Our goal is to provide a gold standard resource, based on a careful semantic annotation designed and supervised by lexical semantics experts. In this paper, we focus on the annotation of more than 12,000 occurrences of common nouns in a French treebank corpus.

Labels used in semantically annotated corpora are of two types: they can either be linked to sets of predefined word senses, as in a dictionary, or correspond to bare semantic classes that are not pre-associated to a given lexicon. In Semcor (Landes et al., 1998), a well-known sense annotated corpus of English, occurrences of a word are associated with a predefined list of senses (represented by synsets) from the Princeton WordNet (Miller et al., 1990). WordNet being known to include very fine-grained lexical distinctions, its hierarchical organization has been seen as different levels of semantic granularity. The top level of WordNet ontology, namely the "Unique Beginners" (UBs) have been taken as coarse-grained categories for both manual and automatic annotation, which has been proved to limit the difficulty of making fine-grained sense distinctions (Palmer et al., 2007; Navigli, 2006; Navigli, 2009). First used by Ciaramita and Johnson (2003) and Ciaramita and Altun (2006) as "supersenses", WordNet UBs are adopted in a growing number of studies devoted to coarse semantic tagging (Johannsen et al., 2014; Flekova and Gurevych, 2016).

For semantic annotation projects based on WordNet UBs[1], the corpus may or may not be pre-tagged with UBs, depending on whether or not a WordNet lexicon is available with sufficient coverage for the target language. Pretagging was used for instance for Danish in the SemDax project (Pederson et al., 2016), but not used for Arabic in (Schneider et al., 2012)[2]. French has a freely availaible Wordnet-like resource, the *Wolf*, that has been automatically built from the Princeton WordNet and other resources, using translation techniques (Sagot and Fišer, 2008). We decided not to use it to pre-tag the corpus we annotated, because the Wolf has only been very partly manually validated and the gain of noisy data for speeding up annotation process is not obvious. Instead, following (Schneider et al., 2012), we used UBs as lexicon-independent semantic classes to annotate nouns in our corpus.

As already pointed out in previous manual supersense-annotation projects (Schneider et al., 2012; Pederson et al., 2016), WordNet UBs in their original version can hardly be exploited and require adjustments. We had to refine the UB tagset, clarify the definition of linguistically motivated ontological categories, and introduce complex types to account for often ignored lexical and contextual phenomena. We manually annotated all noun tokens in a French corpus of about 3100 sentences, with a careful methodology ensuring good annotation quality (73% of the noun tokens were double annotated and adjudicated, and all noun tokens were validated by an expert).

The paper is organized as follows. In Section 2, we present our semantic tagset for nouns, detailing the adjustments we made to the WordNet UBs, and the introduction of operators generating complex semantic tags, to account for contextual or lexical phenomena. We then focus on our annotation methodology (Section 3) and provide both quantitative and qualitative analysis of the resulting data (Section 4). Finally, we conclude with some perspectives on further developments of the resource, both for NLP and lexical semantics studies (Section 5).

2. Description of the tagset

The WordNet UBs from which our supersenses are derived were not initially designed for annotation purposes. They were intended to facilitate lexicographic work by allowing groupings of synsets sharing the same general hypernym (Miller et al., 1990). In order to implement UBs in an operational annotation tagset, we had to make several

[1] See Petrolito and Bond (2014) for an overview.

[2] Note that in the former case, UBs are generalizations over senses, whereas in the latter case, UBs are taken as bare semantic classes. We will use the term supersense in both cases though.

FrSemCor Supersenses	Generalization
Act, Event, ~~Process~~, Phenomenon	Dynamic_situation
Attribute, State, Feeling, Relation	Stative_situation
Animal, Person	Animate_entity
Body, Object, Plant	Natural_Object
Cognition, Communication	Informational_object
Part, Quantity, Group	Quantification
Artifact, Food, **Institution**, ~~Location~~, ~~Motive~~, Possession, ~~Shape~~, Substance, Time, Tops	Other

Table 1: Supersenses used in annotation (left), grouped into more general classes (right). Crossed out supersenses are the UBs that we did not retain, and bold supersenses are the ones we added.

adjustments[3], resulting in differences in semantic class delineation and supersense definition. Moreover, in the light of certain phenomena encountered in target corpus, we decided to introduce supersense operators generating complex supersenses.

2.1. Supersense inventory

Table 1 lists the 24 supersenses that were finally kept in our tagset, and their specificities with regard to the original WordNet UB set. New supersenses appear in bold text, and UBs that we did not retain are crossed out. Generalizations given in column 2 are intended to cluster coherent groups of supersenses and clarify their presentation below. We did not use them for annotation but only to assess agreement scores at various levels of granularity (see below, Section 4.2.).

Removed classes Four original classes have been removed from our inventory because they seemed problematic in terms of definition and/or scope. As far as dynamic situations are concerned (Table 1, line 2), we excluded the Process UB because of the recurrent confusion within dynamic situations, and decided to maintain a distinction between Act and Event only, based on the (non-)agentivity of the denoted situations. We also excluded the Location UB, considering it a semantic role rather than a semantic class. Nouns denoting locations were tagged as Artifact when built by people (e.g. *maison* 'house'), and as Object when corresponding to natural grounds (e.g. *campagne* 'country'). The fact that these physical entities can be contextually used as localizers is not meant to be captured at the lexical level but through semantic parsing. Finally, we also removed the Motive and Shape UBs. The former has already been described as problematic in previous supersense annotation projects, and concerns very few words in WordNet. Shape was excluded because of its heterogeneity and fuzzy boundaries, and because its members could be easily dispatched in other existing categories, such as Artifact and Cognition.

Novel classes Two supersenses absent from the WordNet inventory were used in the annotation. First, we adopted the Institution tag proposed in the SemDax corpus (Pederson et al., 2016), except that we used this tag to classify not only organizations but also their premises and agent facets. In other words, nouns like *banque* 'bank' or *école* 'school' have been tagged as Institution, whether they denote in context the institution itself, the building, or the people who make them work. We also introduced the Part supersense, as a missing counterpart of the Group sense, to annotate nouns like *extrémité* 'end (of something)' or *morceau* 'piece (of something)'.

2.2. Supersense definitions

Removing and adding supersenses to the tagset implies changes in the content of the supersenses. It has to be noted that UBs are very briefly defined in the WordNet documentation. No linguistic indication is given as to how UBs are specifically attributed, and no linguistic test is provided to discriminate between UBs. As a consequence, we had to clarify and refine WordNet definitions of UBs, which may have resulted in modifications in supersense extensions. Detailed definitions of supersenses used in FrSemCor are presented in an annotation guide[4]. Guidelines distinguish each supersense from close supersenses, and provide linguistic tests when possible.

Delineating supersense scope Since supersenses are understood as forming a flat hierarchy, where classes are mutually exclusive, we made the semantic scope of each supersense explicit in the guidelines, by listing both which kinds of nouns pertain to each class and which do not. For instance, several supersenses denoting concrete entities share semantic properties. Plant-denoting nouns denote natural objects, and had to be distinguished from closely related supersenses such as Object (see Table 1, line 5). Consequently, we refined definitions so as to avoid overlaps in supersense extensions. The annotation guide provides explicit guidelines to delineate as clearly as possible the boundaries between competing supersenses.

The major adjustment as concerns class definitions is about Communication and Cognition: in WordNet (and in corpora annotated from it), Communication includes nouns denoting dynamic situations (e.g. *phone call*), and Cognition includes nouns denoting stative situations, namely properties (e.g. *stupidity*). In the FrSemCor dataset, Cognition is limited to nouns denoting informational objects, and Communication is limited to nouns denoting linguistic objects and communication media.

Linguistic tests Whenever possible, linguistic tests are given to help annotators selecting the appropriate tag. For

[3] All other annotation projects using UBs as tags we know of also had to make adjustments, but to a more limited extent. In Schneider et al. (2012), extended definitions of WordNet UBs are proposed, illustrated by English examples, along with rule decisions such as "all man-made structures (buildings, rooms, bridges, etc.) are to be tagged as Artifact". Pederson et al. (2016) extend the WordNet UB set by including more specific classes. For instance, they use Vehicle, Building and Container as subclasses of the Artifact supersense (Martinez Alonso et al., 2016). In both projects though, the definition and lexical scope of WordNet UBs are preserved.

[4] Guidelines for annotation, written in French, can be found on the project page: `https://frsemcor.github.io/FrSemCor/`

instance, it is often difficult to distinguish in French between nouns denoting inherent properties (Attribute) and nouns denoting episodic states (State or Feeling). Tests that have been proposed to account for that distinction (Flaux and Van de Velde, 2000; Kerleroux, 2008) are described in the guide: nouns that combine with predicate in (1-a) (e.g. *taille* 'size') or with predicate in (1-b) (e.g. *autorité* 'leadership') will be tagged as Attribute.

(1) a. *être d'un(e) grand(e) N* 'to be of great N'
b. *faire preuve de N* 'to show N'

Nouns that can appear in phrases such as those in (2) will be tagged as Feeling (e.g. *douleur* 'pain', *espoir* 'hope').

(2) a. *un sentiment / une sensation de N* 'a feeling / sensation of N'
b. *éprouver / ressentir du N* 'to feel / experience N'

Note that these tests are sufficient but not necessary. In many cases, they do not apply and coders have to rely on guideline descriptions and on their intuition to make their decision. Nevertheless, when available, tests may prove useful for tricky distinctions or class identifications.

2.3. Complex Supersenses

One of the original features of the annotation scheme we propose is that it allows for the use of complex supersenses. These are intended to account for two distinct phenomena: functional properties of certain classes of nouns on the one hand, contextual ambiguity or lexical sense multiplicity on the other.

The x operator is used for some nouns pertaining to Group, Part or Quantity classes. Nouns from these classes share a relational meaning involving an argument that can either be included in nominal semantics (e.g *family* denotes a group of people) or realized as a complement in *de* 'of' (e.g. a *class* of something).
The x operator is then used in two different cases: when the argument of the function is included in the semantics of the noun, as for *famille* 'family' (3-a), and when the argument of the function is normally expressed as a *complement* but is not in the target context, as illustrated in (3-b), where the context (medical leaflets) suggests that a class of drugs (Substance) is denoted. In both cases, the aim of the operator is to make explicit the argument of the function. Otherwise, when nouns are contextually used with their argument in a complement introduced by *de*, such as *de systèmes d'organes* in (3-c), they are tagged with the corresponding simple supersense.

(3) a. *Guy était arrivé ici tout jeune avec sa famille*$_{\mathbf{Group\ x\ Person}}$. 'Guy arrived here with his family when he was very young'
b. *Effet de classe*$_{\mathbf{Group\ x\ Substance}}$ *: anomalie de la fonction rénale* 'Class effect: abnormality of kidney function'
c. *Les évènements indésirables sont listés par classes*$_{\mathbf{Group}}$ *de systèmes d'organes dans le tableau 1.* 'Adverse events are listed by system organ classes in table 1'

The / and + operators are used to account for type disjunction resulting either from contextual indeterminacy or from properties of certain nouns known as multifaceted nouns (Cruse, 2002).
Indeterminacy occurs when two meanings are possible in a given context, and yet exclude each other. In (4-a), *arrêt* 'stop' can both refer to an act if someone stops the perfusion, or to an event if the perfusion stops when empty. With more explicit context or knowledge on medical protocols, the word meaning could be disambiguated. Multifaceted nouns have different meanings that are not mutually exclusive. A multifaceted noun like *exposé* 'presentation' in (4-b) can refer to an action, to an informational content, or to both of them.

(4) a. *L'introducteur artériel peut être retiré deux heures après l'arrêt*$_{\mathbf{Act/Event}}$ *de la perfusion de bivalirudine.* 'The arterial initiator can be removed two hours the perfusion is stopped/ends.'
b. *Le visiteur pourra aussi découvrir un exposé*$_{\mathbf{Act/Cognition}}$ *sur l'histoire et le fonctionnement des planètes.* 'The visitor will discover a presentation on the history and the functioning of planets.'

The + operator is used when the context simultaneously gives access to every facet of a word, a phenomenon known in the literature as copredication (Ježek and Melloni, 2011). In (5), the Artifact facet of *contrat* is triggered by the left context and the Cognition facet by the right context.

(5) *Ils trouvent un contrat*$_{\mathbf{Artifact+Cognition}}$ *de 1 million de francs destiné à l'achat d'un appartement pour sa compagne.* 'They found a 1 million francs contract for the purchase of an apartment for his girlfriend.'

3. Annotation design

As mentioned in the introduction, the annotation was conducted entirely manually, without any automatic pretagging inferred from an existing lexicon. In this section, we describe the selected data and the annotation process.

3.1. Selected data

We annotated occurrences of common nouns from the Sequoia Treebank, a French Corpus that includes morphological and syntactic annotations (Candito and Seddah, 2012). The Treebank contains 3,099 sentences from four different sources: a regional newspaper, narrative historical pages from the French Wikipedia, Europarl transcriptions, and two medical reports from the European Medicine Agency[5]. We used some existing annotations in the Sequoia corpus to identify nouns to annotate: parts of speech to focus on common nouns, lemmas to group together the occurrences of the same lemma into a single file. We also used the recently added named entity (NE) and multi-word expression

[5]More precisely, each report is a public assessment of why the marketing authorization has been granted or refused for a given medicine.

Total number of noun tokens	15,197
Total number of noun types	2974
Number of noun tokens in NE or MWE	3689

Table 2: Statistics of common nouns in the Sequoia corpus.

(MWE) annotations (Candito et al., in preparation) to filter out some of the nouns, as explained below in Section 3.2. Statistics about common nouns in the corpus are given in Table 2.

We left aside nouns included in named entities or in MWEs, in case they were not MWE heads (see below, Section 3.2.). The resulting number of tokens we had to annotate is 12,971.

3.2. Annotation process

To perform the manual annotation, we hired three students, who were French native speakers with no prior experience or expertise.

During a first step, one of them isolated a list of non-ambiguous lemmas among the most frequent nouns of the Sequoia corpus. The occurrences of these lemmas were then mono-annotated by another annotator. Every occurrence of all other lemmas was double-blind annotated and adjudicated.

The annotation was carried out in WebAnno, a freely available web tool (Eckart de Castilho et al., 2016). Noun instances were grouped by lemmas, and highlighted in the web tool. Contexts were limited to sentences, but annotators could have access to broader contexts when needed.

Annotators had to assign a supersense tag to each nominal token by selecting one of the 24 tags in the list, or by typing a complex tag when appropriate. As mentioned earlier, no automatic pre-annotation was proposed for nouns and no wild card was allowed in case of difficulty. However, annotators could consult the description of the target noun's English equivalent on the WordNet website – using the "show lexical info" option – in order to guide their decision, while keeping in mind the differences between the two tagsets.

Nouns included in NEs were not selected for annotation, because they were already annotated with PARSEME-FR semantic tags[6]. Only head nouns included in nominal MWEs[7] were tagged, for which annotators had to choose a semantic tag corresponding to the meaning of the whole expression. For instance, occurrences of the noun *cordon* 'cord' were not annotated in a verbal MWE such as *couper le cordon (lit. cut the cord)* '∼leave the nest', contrary to occurrences in *cordon bleu* (lit. *blue cord*) 'good cook' and *cordon ombilical* 'umbilical cord'–tagged respectively as Person and Body.

During the training phase, short annotation sessions were followed by adjudication sessions with an expert of the team to answer coders' questions and clarify guidelines whenever needed[8]. Annotators then performed manual tagging on their own, without communicating with one another. Training and double annotation of the whole dataset required approximately 600 hours.

4. Analysis

In this section, we analyze the annotated corpus, and evaluate annotation quality by studying both inter-annotator agreement and the confusion between supersenses for the part that was double annotated after training phases.

4.1. Analysis of the resulting annotations

The corpus comprises 12,917 annotated noun tokens[9]. 3481 tokens correspond to lemmas that were judged monosemic before corpus annotation, and directly assigned their unique supersense. 3882 tokens were double annotated during the training phases, and 5554 were double annotated after the training phases (see Section 3.2.). All double annotated tokens have been adjudicated.

Supersense distribution The corpus contains annotation for 24 simple supersenses and 64 complex supersenses (see Section 2.3.). Table 3 lists the 20 most frequent supersenses, whether they are simple or complex. Although semantic type representation depends on the kind of corpus we annotated, it shows that most nouns, both types and tokens, belong to the Act class. This result was not expected, since nouns are often conceived as prototypically referring to objects. The list also includes four complex supersenses, that can be viewed as lexical classes in their own right. Informational contents (Cognition) often constitute one facet of multifaceted nouns, the other facet being Act (e.g. *remerciement* 'acknowledgement') or Artifact (e.g. *notice* 'notice'). The Act/Possession supersense concerns the act of paying a certain amount of money, denoted by nouns like *aumône* 'alms' or *virement* 'transfer'. Finally, the Group×Person supersense reveals the most represented semantic class among French collective nouns, an information that is not captured in the original WordNet inventory since collective nouns are all gathered in the Group UB.

Lexical information is summed up in Table 4. Among the 2697 lemma types of the annotated corpus, 400 have been attributed several supersenses: 1 lemma has been tagged with 8 distinct supersenses (*élément* 'element'), 3 lemmas with 5 supersenses (*perfusion* 'infusion', *système* 'system', *source*, 'source'), 31 lemmas with 4 supersenses, 84 lemmas with 3 supersenses.

Nouns having several supersenses are either polysemous as illustrated in (6) with *intérêt* 'interest', or multifaceted as illustrated in (7) with two occurrences of the same word-sense of *contrat* 'contract'.

(6) a. *L'enquête avait mis en évidence l'intérêt*$_{\textbf{Feeling}}$ *des adultes pour ces ac-*

[6]The semantic types for NEs chosen by the PARSEME-FR team are Person, Organization, Location, Artifact and Event. A mapping has to be done between these tags and our supersenses.

[7]MWEs were made visible in WebAnno.

[8]Annotations performed during the training phases are not taken into account in inter-annotator agreement scores (4.2.).

[9]The 54 remaining tokens, with respect to the total number of instances that were supposed to be annotated, are mainly acronyms and items wrongly tagged as noun, that could not be annotated (e.g. *ST-, 2-7113-0389-6*).

Supersense	Nb lemma tokens	Nb lemma types
Act	2079	613
Person	1853	396
Time	1052	108
Cognition	906	252
Institution	735	155
Event	712	149
Substance	711	129
Quantity	687	93
State	621	160
Attribute	565	184
Artifact	530	236
Possession	366	71
Act/Cognition	314	97
Group×Person	222	54
Artifact/Cognition	213	51
Body	203	54
Object	159	56
Feeling	81	46
Part	67	28
Phenomenon	60	18

Table 3: Distribution of the 20 most frequent supersenses in the annotated corpus (= those used for more than 50 tokens).

Lemmas	Nb types	Nb tokens	Avg. ambig. (types)	Avg. ambig. (tokens)
All	2697	12,917	1.2	1.6
Multi-sense	400	4293	2.4	2.8

Table 4: Supersense ambiguity in the annotated corpus, both for all lemmas ("All") and for lemmas with more than one supersense ("Multi-sense"). Columns 2-3: nb of distinct lemma types and nb of lemma tokens. Columns 4-5: average supersense ambiguity for lemma types / for lemma tokens.

tivités. 'The survey revealed French' interest for these activities.'

b. *Le budget de la France ne peut rembourser 16 milliard de francs plus intérêts*$_{\textbf{Possession}}$ 'France's budget cannot repay 16 billion of francs plus interest.'

(7) a. *Il a signé un contrat*$_{\textbf{Artifact/Cognition}}$ *de deux ans.* 'He signed a two-year contract'

b. *Thomson est le maître d'œuvre du nouveau contrat*$_{\textbf{Cognition}}$. 'Thomson is the prime contractor for the new contract'

A lemma can also be associated with several supersenses not because of polysemy, nor because of micro-senses, but as a result of our reference-driven annotation method. For instance, the noun *membre* 'member' can be annotated as Person (*un membre de la famille* 'a member of the family') or Institution (*un membre de l'ONU* 'a member of the UN')[10] or Animal (*un membre de la meute* 'a member of the pack'), depending on its contextual denotation. In short, the set of lemmas whose occurrences have been annotated with distinct supersenses will need to be carefully examined in order to distinguish between cases of polysemy, cases of multifaceted nouns, and cases of general nouns having heterogeneous referents.

Anno 1	Anno 2	Partial match
Group×Person	Group	Group
Act+Cognition	Act/Cognition	Act_Cognition
Act/Cognition	Artifact/Cognition	Cognition

Table 5: Examples of partial matches

4.2. Inter-Annotator agreement

We calculated inter-annotator agreement for the 5554 tokens annotated after the training phases.

Agreement scores We have evaluated the inter-annotator agreement using both the plain F-score among the two sets of annotations and Cohen's kappa.

Four types of agreement have been considered according to the matching type on the one hand, and to the granularity of the tagset on the other hand. Table 6 reports the results. In "exact" match, agreement is achieved if the two annotated categories are identical (whether simple or complex), whereas in "partial" match, we partially neutralized differences between complex categories. In the latter case, complex categories are split into a list of simple supersenses. Then, in case there is a non-empty intersection between both lists, each decision is replaced by the intersection list (if this intersection contains several simple types, they are concatenated with a meaningless operator "_"). Examples of partial matches are illustrated in Table 5.

We also computed exact and partial match scores using a coarser set of semantic tags, in which we have manually grouped some supersenses into coarser tags. They are those proposed as generalization in Table 1, column 2. So for instance, in coarse agreement, Act, Event, Phenomenon tags are all generalized as Dynamic_situation.

The agreement levels presented in Table 7 are satisfactory since categories are numerous (and actually form an open set because of the possibility to build complex supersenses). The likelihood that the agreement is due to chance being very low (near to 0.1 in both types of match), kappa's values are close to agreement values. Moreover, agreement scores take into account all cases, given that no wild card was allowed. As a comparison, Schneider et al. (2012) report similar scores (to Exact match) calculated only on data judged as unproblematic by annotators.

Disagreement The confusion matrix in Table 7 summarizes confusions between classes. Diagonal scores (agreement per supersense) reveal a great variability within the

[10]Note that, when referring to an institution, *membre* 'member' can refer to the organisation in itself, like a state in l'ancien membre de l'Europe 'the former member of Europe', or to its agent facet (*les membres de l'ONU ont voté*, 'members of UN voted), see Section 2.1.

	Full set	Coarse set
Exact match	κ: 0.649 ag: 0.683	κ: 0.700 ag: 0.742
Partial match	κ: 0.734 ag: 0.763	κ:0.790 ag: 0.822

Table 6: Inter-Annotator agreement: κ stands for Cohen's kappa and *ag* for agreement Fscore.

seventeen more frequent classes, between those that generate the greatest consensus (like Person or Substance) and those that trigger many alternative decisions (like Event or Attribute).

The most prominent symmetrical confusions are between State and Attribute, and between Body and Object, which is not unexpected. State and Attribute are used for nouns denoting stative situations (along with Feeling, mainly confused with State), and are hardly distinguished on a linguistic basis. Guidelines explicit the delimitation between Body and Object–these can be confused because parts of the body are natural objects. Nevertheless a close look at the annotated data shows that disagreement cases for Body/Object concentrate on one lexical item, namely *voie* 'route' (for example in *voie orale* 'oral route').

Confusions among dynamic situations are not symmetrical, contrary to what is observed in the case of stative situations. Event is the most often selected tag in case of disagreement upon Act, but still in limited proportion (7.7%), whereas Act is selected in almost a third of the disagreement cases upon Event.

Disagreements related to the most frequent complex types clearly appear in the matrix, calculated from partial matches (see examples in Table 5). Most frequent alternatives to Cognition are Act, followed by Artifact, revealing the difficulty to identify multifaceted nouns pertaining to Act/Cognition or Artifact/Cognition.

Finally, the results highlight the three facets encapsulated in the Institution supersense, namely the institution's members (confusion with Group×Person, 7.8%), the institution's place (confusion with Artifact, 8.5%), and the institution's activity (confusion with Act, 6.2%). We also observe that most of the alternative decisions for Group×Person involve Institution (22.9%), showing how delicate it is to assess whether a group of people forms an institution or notâdisagreement cases concern words like *pouvoir* 'lit. power / the authorities', *majority* 'majority'.

It can be noted that each of the above mentioned pairs of classes belongs to the same supersense group (as defined in Table 1).

5. Conclusion

In this paper, we have presented the FrSemCor resource, a French corpus in which nouns are manually annotated with supersenses. The annotation was carried out using a carefully defined semantic tagset that refines WordNet Unique Beginners, clarifies conceptual contents and ontological class boundaries, and accounts for complex lexical meanings. We tried to find the best trade-off between abiding by the original WordNet inventory, in order to produce a resource compatible with other supersense annotated corpora (for multilingual purposes), and defining the most linguistically reliable tagset, as well as the more explicit and intuitive for annotators. Another interest of the resource is the coverage of manual annotation, since 100% of the occurrences of the nouns were doubly manually annotated, providing a gold standard for French. Agreement scores (kappa 0.64 to 0.79, according to semantic granularity) range from moderate to substantial. In short term, we intend to use this dataset to test supervised or semi-supervised methods for word sense disambiguation of French nouns. Data can also be generalized over verbal argument structures, and be used as a cue for French verbal disambiguation (Segonne et al., 2019). The resource can also be used in a theoretical perspective and offer empirical supports for studies dedicated to syntactic aspects of lexical semantics, such as co-predication for multifaceted words (Ježek and Melloni, 2011), or complement optionality for relational nouns (Partee and Borschev, 2003).

6. Acknowledgements

We are grateful to Lila Entressangle, Louise Fouquet and Lucie Miskic, who are now expert annotators, and to Héctor Martínez Alonso for his valuable insights on the project.

This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the program "Investissements d'Avenir" (reference: ANR-10-LABX-0083). It contributes to the IdEx Université de Paris - ANR-18-IDEX-0001

7. Bibliographical References

Candito, M. and Seddah, D. (2012). Le corpus sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical. In *Proceedings of TALN'2012*.

Candito, M., Constant, M., C., R., Savary, A., Guillaume, B., and Cordeiro, S. ("in preparation"). A french corpus annotated with named entities and multi-word expressions.

Ciaramita, M. and Altun, Y. (2006). Broad-coverage sense disambiguation and information extraction with a supersense sequence tagger. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 594–602.

Ciaramita, M. and Johnson, M. (2003). Supersense tagging of unknown nouns in wordnet. In *Proceedings of the EMNLP*, pages 168–175.

Cruse, D. A., (2002). *Aspects of the micro-structure of word meaning*, pages 30–51. Oxford:Oxford University Press.

Eckart de Castilho, R., Mújdricza-Maydt, É., Yimam, S., Hartmann, S., Gurevych, I., Frank, A., and Biemann, C. (2016). A web-based tool for the integrated annotation of semantic and syntactic structures. In *Proceedings of the LT4DH workshop at COLING 2016*.

Flaux, N. and Van de Velde, D. (2000). *Les noms en français. Esquisse de classement*. Ophrys, Paris.

Flekova, L. and Gurevych, I. (2016). Supersense embeddings: A unified model for supersense interpretation,

	Act	State	Cognition	Person	Artifact	Attribute	Event	Institution	Substance	Quantity	Possession	Time	Body	GroupxPerson	Group	Feeling	Object	Other
Nb annotations	2735	1097	1009	958	938	642	532	498	451	394	282	228	190	172	137	131	126	588
Nb instances	1600	728	684	493	545	447	387	306	246	227	173	138	111	105	96	87	89	411
Act	**70.9**	11.1	14.9	0.0	1.8	6.3	31.8	6.2	1.2	0.4	13.3	5.1	0.0	6.7	1.0	14.9	0.0	11.7
State	5.1	**50.7**	4.7	1.4	1.8	22.4	12.7	1.0	4.1	1.3	1.7	5.8	0.0	1.0	2.1	18.4	0.0	8.5
Cognition	6.4	4.4	**47.5**	0.0	10.5	10.7	1.3	3.9	0.0	2.2	13.3	6.5	0.0	4.8	30.2	5.7	0.0	6.7
Person	0.0	1.0	0.0	**94.3**	0.6	0.2	0.0	1.6	0.0	0.0	0.6	0.0	2.7	0.0	6.2	0.0	0.0	0.5
Artifact	0.6	1.4	8.3	0.6	**72.1**	0.4	0.0	8.5	2.8	0.9	0.6	0.0	0.0	1.0	0.0	0.0	9.0	6.2
Attribute	1.8	13.7	7.0	0.2	0.4	**43.6**	6.5	0.7	0.0	3.5	2.9	5.8	0.0	0.0	2.1	10.3	0.0	3.5
Event	7.7	6.7	0.7	0.0	0.0	5.6	**37.5**	0.7	0.4	0.4	0.6	7.2	0.0	0.0	0.0	0.0	0.0	6.2
Institution	1.2	0.4	1.8	1.0	4.8	0.4	0.5	**62.7**	0.0	0.9	0.0	0.0	0.0	22.9	8.3	0.0	11.2	0.2
Substance	0.2	1.4	0.0	0.0	1.3	0.0	0.3	0.0	**83.3**	4.0	0.0	0.0	1.8	0.0	0.0	0.0	7.9	0.5
Quantity	0.1	0.4	0.7	0.0	0.4	1.8	0.3	0.7	3.7	**73.6**	1.2	1.4	0.0	0.0	1.0	0.0	1.1	5.7
Possession	1.4	0.4	3.4	0.2	0.2	1.1	0.3	0.0	0.0	0.9	**63.0**	0.0	0.0	0.0	0.0	0.0	0.0	1.2
Time	0.4	1.1	1.3	0.0	0.0	1.8	2.6	0.0	0.0	0.9	0.0	**65.2**	0.0	0.0	0.0	0.0	0.0	1.0
Body	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0	**71.2**	0.0	5.2	0.0	24.7	0.0
Group x Person	0.4	0.1	0.7	0.0	0.2	0.0	0.0	7.8	0.0	0.0	0.0	0.0	0.0	**63.8**	0.0	0.0	0.0	0.0
Group	0.1	0.3	4.2	1.2	0.0	0.4	0.0	2.6	0.0	0.4	0.0	0.0	4.5	0.0	**42.7**	0.0	0.0	0.2
Feeling	0.8	2.2	0.7	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**50.6**	0.0	0.0
Object	0.0	0.0	0.0	0.0	1.5	0.0	0.0	3.3	2.8	0.4	0.0	0.0	19.8	0.0	0.0	0.0	**41.6**	1.0
Other	2.9	4.7	3.9	0.4	4.6	3.1	6.5	0.3	0.8	10.1	2.9	2.9	0.0	0.0	1.0	0.0	4.5	**46.6**

Table 7: Confusion matrix for the 17 most frequent supersenses, in partial match. The tag "Other" groups together the 42 other supersenses, each concerning less than 100 annotations (including Part, Phenomenon, Animal, etc.). First line: number of annotator decisions for each category. Second line: number of instances annotated for each category by at least one annotator. Bottom part: each column for class c provides the proportion, among the instances annotated as c by at least one annotator, of each class for the other annotator. For instance, among the 728 instances annotated as State at least by one annotator, the other annotator chose Act for 11.1% of the cases, and State for 50.7% of the cases.

prediction, and utilization. In *Proceedings of the ACL*, pages 2029–2041.

Ježek, E. and Melloni, C. (2011). Nominals, polysemy and co-predication. *Journal of cognitive science*, 12:1–31.

Johannsen, A., Hovy, D., Alonso, H. M., Plank, B., and Søgaard, A. (2014). More or less supervised super-sense tagging of twitter. In *Proceedings of the Third Joint Conference on Lexical and Computational Semantics*.

Kerleroux, F., (2008). *Des noms indistincts*, pages 113–132. Amsterdam: John Benjamins.

Landes, S., Leacock, C., and Tengi, R. I., (1998). *Building semantic concordances*, pages 1999–2016. Fellbaum, C.

Martinez Alonso, H., Johannsen, A., Olsen, S., Nimb, S., and Pedersen, B. S. (2016). An empirically grounded expansion of the supersense inventory. In *Proc. of the Global Wordnet Conference*.

Miller, G., Beckwith, R., Fellbaum, C., D., G., and Miller, K. (1990). Wordnet: An online lexical database. *International Journal of Lexicography*, (3):235–244.

Navigli, R. (2006). Meaningful clustering of senses helps boost word sense disambiguation performance. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 105–112.

Navigli, R. (2009). Word sense disambiguation: A survey. *ACM Computing Surveys*, 41(2):1–69.

Palmer, M., Dang, H. T., and Fellbaum, C. (2007). Making fine-grained and coarse-grained sense distinctions, both manually and automatically. *Natural Language Engineering*, 13(2):137–163.

Partee, B. H. and Borschev, V. (2003). Genitives, relational nouns, and argument-modifier ambiguity. *Modifying adjuncts*, 4:67–112.

Pederson, B. S., Braasch, A., Johannsen, A., Martinez Alonso, H., Nimb, S., Olsen, S., Søgaard, A., and Hartvig Sørensen, N. (2016). The semdax corpus – sense annotations with scalable sense inventories. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 842–847.

Petrolito, T. and Bond, F. (2014). A survey of wordnet annotated corpora. In *Proceedings of the Seventh Global WordNet Conference*.

Sagot, B. and Fišer, D. (2008). Building a free French wordnet from multilingual resources. In *OntoLex*, Marrakech, Morocco, May.

Schneider, N., Mohit, B., Oflazer, K., and N., S. (2012). Coarse lexical semantic annotation with supersenses : an arabic case study. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, Jeju Island, Korea.

Segonne, V., Candito, M., and Crabbé, B. (2019). Using wiktionary as a resource for wsd: the case of french verbs. In *Proceedings of the 13th International Conference on Computational Semantics*, pages 259–270.