Check for updates

A Simple Set of Rules for Characters and Place Recognition in French Novels

Cyril Bornet* and Frédéric Kaplan

Digital Humanities Laboratory, DHI, CDH, EPFL, Lausanne, Switzerland

This article describes a simple unsupervised system for automatic extraction and classification of named entities in French novels. The solution presented combines a set of different standalone classifiers within a meta-recognition system. The system is tested on 35 classic French novels, representing 5 million words and 3,700 names of people and places. The results demonstrate that although none of the standalone methods clearly outperforms the others, their combined classification offers a robust solution in this context.

Keywords: text processing, named entity recognition, NER, literature, French, distant reading

1. INTRODUCTION

The systematic extraction of characters and places from novels could potentially contribute to new analyses and ultimately better understanding of fictional stories and the way they were written. The ability to view the authors' narrative strategies by plotting the interleaved presence of recurring characters and places opens new forms of distant visualization and facilitates a *distant reading* comparison of novels (Moretti, 2013). Similarly, connecting distinct novels by the places and characters they share may offer interesting and innovative navigation and exploration means, as already suggested by some real-world applications (Albanese, 2011).

This paper presents a simple system for automatically extracting named entities within French novels. Named Entity Recognition (*NER*) has become reasonably well developed, driven notably by the needs for machine translation and automatic information extraction out of flows of news articles (Hirschberg and Manning, 2015). However, only a small number of systems perform well on languages other than English, and in particular, as in our case, on French contents (Azpeitia et al., 2014). Previous initiatives targeted specifically old French newspapers corpora (Mosallam et al., 2014); however, fictional novels are challenging in a different way since their referents are typically not listed in databases commonly used by many *NER* systems, like DBPedia,[1] Wikidata,[2] or YAGO.[3] One additional difficulty with non-English languages is the limited amount of readily available sets to train machine learning recognizers. Therefore, solutions that can achieve good results without the need of training beforehand are highly relevant.

Our solution[4] combines a set of different standalone unsupervised recognition strategies feeding a meta-recognition system. This approach, whose concept was first described aiming for improved accuracy needs in biometrics, has proven to be effective in various other fields (Scheirer et al., 2011), and among them notably *NER* (Si et al., 2005). The rationale behind meta-recognition is to reach

---

[1]http://wiki.dbpedia.org.
[2]http://www.wikidata.org.
[3]http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago.
[4]Made available at https://github.com/dhlab-epfl/3n-tools/.

out for a set of independent classifiers, each looking at different features and possibly using different strategies. One would then expect that although none of the standalone method clearly outperforms the others, their combined judgment offers a robust solution to the tackled recognition problem, as their individual errors will tend to cancel out.

The next sections describe the performances obtained with state of the art tools compared to each of our classifiers on a 5M tokens test corpus, and the overall gain brought by averaging their results using a fixed weight and a meta-recognition algorithm. The conclusion outlines some applications for analyses of novels that can be performed on the outputted categorization.

1.1. Test Procedure

We built a test corpus made out of 35 classic works (5M tokens) in French, stemming from digitized books part of Project Gutenberg.[5] The works were handpicked in a way to get a mix of different genres and narration types. They include 3 sagas, 6 classical books that were translated from other languages and 25 French classical novels. For each piece of work, we produced a corresponding reference file by extracting the proper nouns occurring more than five times and manually labeling each of them with the category they fall in (character, place, or other). The resulting database references about 3,700 categorized named entities, along with the work they come from and their number of occurrences.

Given our further needs, we considered names referring to pets, companies, and groups of persons as regular characters whenever they serve the same narrative purpose (people mostly interacting with corporations or groups of people). Conversely, they were considered places in the case of names used to qualify where actions take place (as with corporations considered as locations), or ethnical groups used to name their originating place (as in "*le pays des Oreillons*"). This differs a lot from most knowledge databases, as the very bounds between characters and places might be fuzzy and context dependent. For instance, if the word "*France*" is part of the expression "*Roi de France,*" we do not want it to be labeled as a place, but the whole entity as a *character* instead.

For the actual test procedure, we ran each classifier on our corpus and computed the averagely yield precision (positive predictive value), recall (true positive rate) and F_1 score of the classification task. Section 2 shortly discusses state of the art of existing tools under the same constraints, and sections 3.2 to 3.6 give the details of the implementations of each classifier, and the obtained performances.

2. NER IN NOVELS

The field of *NER* is evolving quickly, as big data approaches and access to increasingly large corpora brought up several projects promising interesting results. However, these tools remain mainly purposed to summarize, categorize, and extract meaning out of

short texts that relate to the outside world (e.g., news articles, encyclopedias, or otherwise informative texts). Analyses on big portions of self-contained texts like entire chapters or full fiction books, whose named entities are difficult to look up and interpret out of their context, remain a challenging endeavor. Novels in particular are different from traditional applications for *NER* techniques for at least two reasons:

- Novels create a world of their own, with own recurring characters and places that can be hard to understand out of their context. Many traditional *NER* methods do not exploit this characteristic as they on the contrary rely on external databases to identify and classify named entities (Jovanovic et al., 2014).
- Each novel tends to be characterized by specific stylistic features whose purpose precisely is to give the narration a unique taste. On the one hand, a novelist's writing diverges in nature from journalistic, academic, or other forms of non-fiction writings, and on the other hand, it is usually purposely meant to introduce noticeable differences compared to the style of other authors. This phenomenon is obvious enough that it was shown to allow to confidently attribute texts to their authors in many contexts (Stamatatos, 2009). This intrinsic diversity makes a one-size-fits-all approach to *NER* difficult and is problematic for pattern-inducing machine learning algorithms.

In order to establish a baseline for our study, we conducted an initial evaluation with one of the current widely used *NER* systems. *OpeNER*[6] is a multilingual and powerful *NER* pipeline, promising state of the art results on French corpora (Azpeitia et al., 2014). Even if it was not designed to handle big amounts of text,[7] its processing routine is efficient and produces XML result files that can easily be processed and compared to our reference data. Since *OpeNER* is extracting a wider range of entities than the ones we considered (such as date and time information), we only considered the relevantly tagged words and manually resolved naming differences in order to keep the comparison fair and accurate.

3. A SET OF SIMPLE RULES

3.1. Preprocessing

In general, spotting proper nouns in French texts is a rather easy task due to capitalization rules, that are quite similar to English if not simpler due to less false positives (e.g., words related to languages, or days and months are written in lower case) (Geno, 1992). Still, the task is not entirely trivial, because we need to filter out capitalization due to sentences starts and miscellaneous stylistic effects (such as subtitles, quotes, and verses). Additionally, we want full names extracted as single entities and remove false positives, like honorific titles or named time periods (Grevisse and Lenoble-Pinson, 2009). In order to do that, we designed a two-pass method:

---

[5]Project Gutenberg | Browse By Language: French—http://www.gutenberg.org/browse/languages/fr.

[6]The OpeNER project—http://www.opener-project.eu.

[7]The JVM heap size limit set by JRuby needed to be increased to over 8 GB in order to process our bigger samples (yet only about 6 MB in size).

- First, our scripts extract a list of all proper nouns that are not leading sentences. To do this, we compute all 1, 2, and 3 g of each sentence while letting out the first word. We then keep all isolated proper nouns (i.e., capitalized words surrounded by lower cased ones), pairs of juxtaposed proper nouns, and triplets of proper nouns joined by an hyphen or a nobiliary particle,[8] each of them as if they were single entities.
- Then, we run a second pass on the full text, where we look for all sentences containing each noun we identified in the first run to include legitimate nouns leading sentences we had left out. For each of them, we store an index of the sentences they appear in for further processing.

This approach allows to recover all occurrences of each capitalized word, as long as they are not systematically at the start of sentences. In practice, the resulting list needs to be refined thereafter, as some capitalized common nouns still happen to end up in it. The reasons for this may be multiple, ranging from sentence tokenization errors, typos in the source text or other stylistic effects that may influence punctuation and case. This refining can be done accurately and efficiently by combining three strategies:

- by computing and comparing the mean positions of each word in the sentences they were found in. As we may expect, the distribution of words that are usually capitalized only when leading the sentences will concentrate toward 0, whereas the real proper nouns' will tend to even out. **Figure 1** illustrates that the typical distributions often allow for easy separation, with very few outliers.
- The said outliers being usually connection words (e.g., pronouns and prepositions) that are often capitalized in different contexts, they can be filtered out using a simple French stopwords list.
- Third, words related to nationalities or ethnic groups can be easily filtered out as they are usually present in both their singular and plural forms, whereas genuine proper nouns will usually not follow this characteristic. Hence, we remove the words ending in "s" that can also be found without the "s" throughout the same text.

Once identified, proper nouns usually fall in three main categories that serve different purposes to narration: they can namely be *characters*, *places*, or *others* (brands, abstract concepts, acronyms…). We designed and evaluated six independent classifiers. Each classifier gets one word at a time as an input as well as the context that is necessary and relevant for its way of processing data, and returns the predicted category (namely *character*, *place*, or *other*). We first present the implementation characteristics of each component before looking in details at the resulting scores.

3.2. Classifier #1: Obvious Context, Titles, and Predicates

When one encounters a proper noun in a sentence, a good guess on its nature can sometimes easily be taken due to the immediate context. The simplest case, which we will refer here as *obvious context*, would be if the noun is immediately preceded by a title or a predicate that hints at what it refers to. For instance, the name *Vilquin* could probably refer to anything without clear preference when out of any context, but it would be very easy to classify, respectively, as a person or as a place, if at least one sentence was mentioning "[…] *monsieur Vilquin* […]" or, conversely, "[…] *rue Vilquin* […]."

For this classifier, we compiled a simple list of obvious context classifiers that allow to make good guesses about the nature of the immediately or next to following proper noun:

- In the case of characters, the list is basically the usual name titles of the French language, such as *madame* or *docteur*.[9]
- For places, we considered the most common predicates used to qualify toponyms, as for instance *ville*, *avenue*, or *rue*.
- For other words we wanted specifically to filter out, we added a short list of terms referring to deities (*dieu, jésus, marie, vierge, saint*).

3.3. Classifier #2: Naive Position

In French, like in many other languages, the grammatical structure makes it more likely for sentences to follow a pattern that puts the subject of the action at the beginning, and the location toward the end. This characteristic can be used when one looks at enough examples to make a simple, yet quite powerful guess about the global roles of the proper nouns.

The accuracy of this classifier is indeed strongly dependent on the writing style of the author, as the frequent use of specific figures of speech may break its work hypothesis, and longer sentences may narrow the gap between the categories or blurry the bounds. This can be seen clearly in **Figure 2**, where we show the relative positions of identified classes of names for three different stories. In those examples, we can see the effectiveness of a separation guess at around 45% of sentences' length, which is expected to yield quite good results for the first two books, yet a bit more disappointing ones for the third one.

3.4. Classifier #3: Semantics out of Neighboring Words

Inspired by Latent Semantic Analysis (Dumais, 2005), in which it is hypothesized the meaning of words are the result of their neighbors', a slightly more sophisticated approach consists of a broader look at neighboring words semantics. This approach is very different from section 3.2 in that we are not expecting the surrounding words to qualify the noun *per se* but to hint at its nature due to the actual meaning of the near context. For this implementation, we compiled lists of words that are more likely (but not exclusively) to appear, respectively, nearby characters, places, or abstract concepts. For instance, we expect names of characters to be more often surrounded by words related to emotions, body functions, speech, or professions, whereas names of places would be more closely related to motion verbs, place features, and prepositions.

---

[8] Since nouns can be in any language despite the text being in French, we considered the list of nobiliary particles found on the related English Wikipedia page: http://en.wikipedia.org/wiki/Nobiliary_particle.

[9] Wikipédia: Titres et prédicats—http://fr.wikipedia.org/wiki/Titres_et_prédicats.

FIGURE 1 | Typical mean positions of uppercased words in their respective tokenized sentences vs. their number of occurrences (on a logarithmic scale).

FIGURE 2 | Relative mean position of characters and places names for three classical French novels.
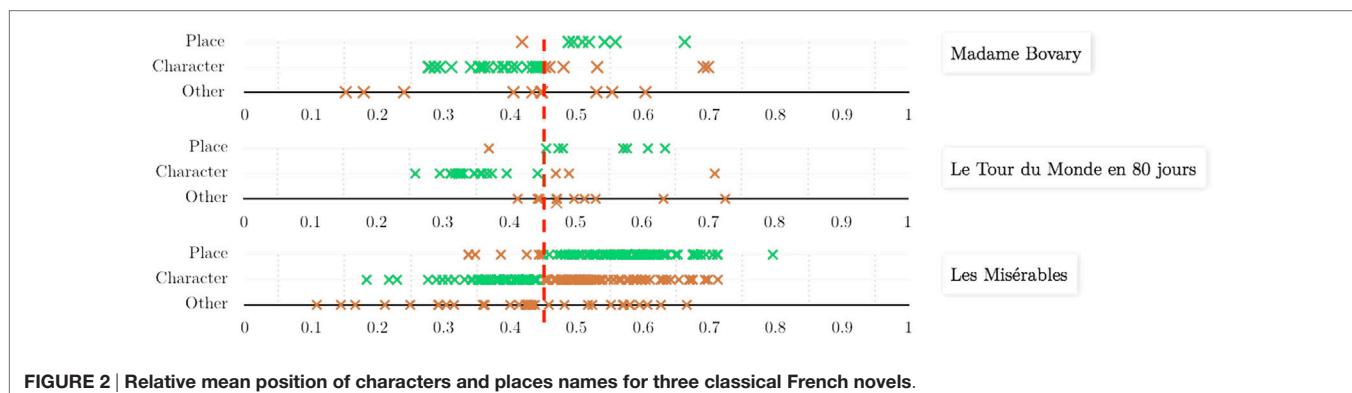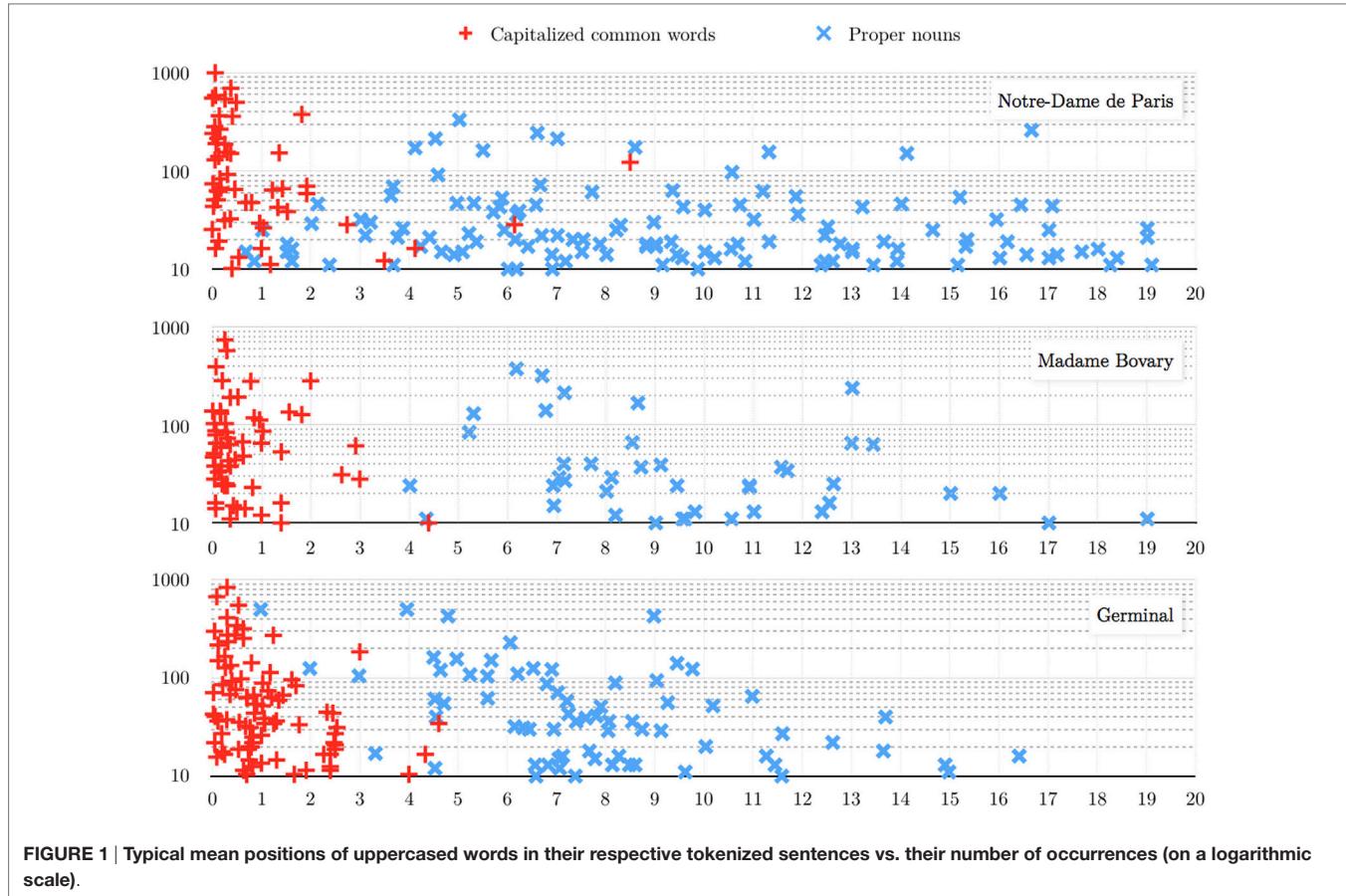
Starting from common nouns that are unambiguously related to one of the categories we are interested in, we used a French synonyms dictionary service[10] to put up a list aiming to be as extensive as possible. The final files resulted in 4,500 words for characters, 670 for places and 50 for concepts (see Appendix B for the complete list). The script then looks for these words in the neighborhood of the nouns to be disambiguated and returns the most probable category.

3.5. Classifier #4: Grammatical Structure

As characters and places serve different narrative purposes, one may expect the grammatical constructs surrounding them to differ in a significant way. For instance, place names are often preceded with prepositions or determiners, whereas it is expected for character names to be more often directly followed by verbs. We thus introduced a script classifying names based on its knowledge of the full text, grammatically tagged using TreeTagger,[11] and

---

[10]http://www.synonymo.fr.

[11]http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger.

tokenized in sentences. To guess the nature of the names, it then matches all sentences containing them against a set of rules that are typical constructions one uses when writing about a person or a place.

We tried out a set of seven manually established rules covering the most straightforward grammatical constructs (described in details in **Table 1**), plus two that help filter out tokenization errors at a sentence level by flagging words that are preceded by a punctuation mark or that are alone in their sentence. Formally, rules are patterns possibly matching the set $\{t_{i-2}, t_{i-1}, t_{i+1}, t_{i+2}\}$ of grammatical tags attributed to up to four neighboring tokens $\{w_{i-2}, w_{i-1}, w_{i+1}, w_{i+2}\}$ for each proper noun w_i. When matched, they increase or decrease the probability score for one or several classifications, and the category yielding the highest score is chosen and returned in the end.

3.6. Classifier #5: Online-Querying

A lot of proper nouns can be non-ambiguously or with a high probability related to one or several categories based on general knowledge. For instance, a human could make a guess that "*Elisabeth*" will most likely refer to a person, whereas "*Manhattan*" is likely to be a place and "*Amour*," a concept. But the same knowledge may equivocally tell that those same words could also potentially be related to the ship (*RMS Queen Elisabeth*), an abstract concept (*project Manhattan*), or a place (*Amur river*), probably with a lower likelihood if no other context is available.

For many nouns, the knowledge we are looking for is well captured in the categorization of their related Wikipedia pages. Using categories instead of the text of the articles also presents the advantages of being very straightforward and reduces a lot noisy signals related to text processing techniques. To test this idea, we implemented a simple algorithm that gathers the categories of the page whose name is closest to the noun we are looking for and looks for ones tagging people, places, or abstract concepts. In the case no category gives a hint (which tends to happen both with very complex or very precise pages), it tries to recursively walk up the hierarchy until the necessary clues are found.

3.7. Classifier #6: Quotes-Based

Several works already showed the relevance of locating direct and indirect speech parts to identify characters in novels (Glass and Bangay, 2007; Goh et al., 2012; Karsdorp et al., 2012). Most of

these approaches rely heavily on the lexical database WordNet[12] to find out speech-related verbs and refine their accuracy, but for performance reasons and since we wanted the classifiers to remain efficient even on very long texts we implemented a simpler version that simply checks the proximity of detected proper nouns to quotation marks. For each proper noun w appearing m_w times, the system would essentially count the number q_w of mentions that appear near quotations. It then computes:

$$r_w = \frac{m_w}{q_w} \qquad (1)$$

the quotation ratios for each noun, and:

$$\tilde{t} = \text{median}\{r_w\} = \begin{cases} r_{\frac{n}{2}} & \text{for an odd number } n \text{ of words} \\ \frac{1}{2}\left(r_{\frac{n}{2}} + r_{\frac{n+1}{2}}\right) & \text{for an even number } n \text{ of words} \end{cases} \qquad (2)$$

the median value of all ratios, to be considered the differentiation threshold for this book. Each noun w is then assigned the *character* class if r_w is higher than $\tilde{t}$ or *place* otherwise.

4. META-CLASSIFICATION

Once all classifiers returned their answer for a given word, the last step is to compare these results and to decide on a final answer. This meta-classification step can be done by voting systems, choosing the final result according to the majority of predictions using various strategies, or by a meta-recognition system, aiming to discard classifiers that seem to have encountered a problem on the considered text file. We implemented and discussed the performance of four distinct meta-classification methods.

4.1. Simple Vote

The easiest and most obvious solution to average the different classifications is a simple voting system (i.e., the classification who gets the majority of results wins). However, since there is an even number of classifiers, ties are to be expected. This situation is quite unlikely since it would require exactly three classifiers deciding correctly, and the three others agreeing incorrectly on a wrong categorization. Still, in case, this situation occurred the final choice would be non-deterministic by lack of model to support one option or the other. For this reason, we introduced a second meta-classification, which involves for each classifier to compute a confidence self-assessment score.

4.2. Self-Assessed Confidence

For most classifiers, their internal mechanics allow themselves to evaluate to which extent the strategy they are using seems likely to return reliable results, given the current work context. Hence, a simple strategy to help the voting process in the case of ties is for each classifier to return a confidence index, between 0 and 1.

The used self-assessing strategies are as follows:

- For *Obvious Context*, *Semantic*, and *Wikipedia* classifiers (sections 3.2, 3.4, and 3.6), we use $\frac{C_{max} - C_{min}}{\sum C_i}$ where C_{max} is the count of most and C_{min} of least represented categories for which we found clues. For instance, if the Semantic classifier

TABLE 1 | Rules for grammatical structure classifier.

t_{i-2}	t_{i-1}	t_{i+1}	t_{i+2}	Result
(At least one is a verb)				Character
	VER			Character
	NAM			Character
PRP	PRP		–	Place
	VER			Place
	DET			Place
	PRP	(Empty)	(Empty)	Concept
	PUN			Unknown[a]
	(Empty)	(Empty)		Unknown[a]

[a]*Probable tokenization error.*

VER, verb; NAM, proper noun, PRP, preposition; DET, determiner; PUN, punctuation.

---

[12]WordNet | A lexical database for English—https://wordnet.princeton.edu/.

FIGURE 3 | Comparison between precision and recall for each classifier, on each book.
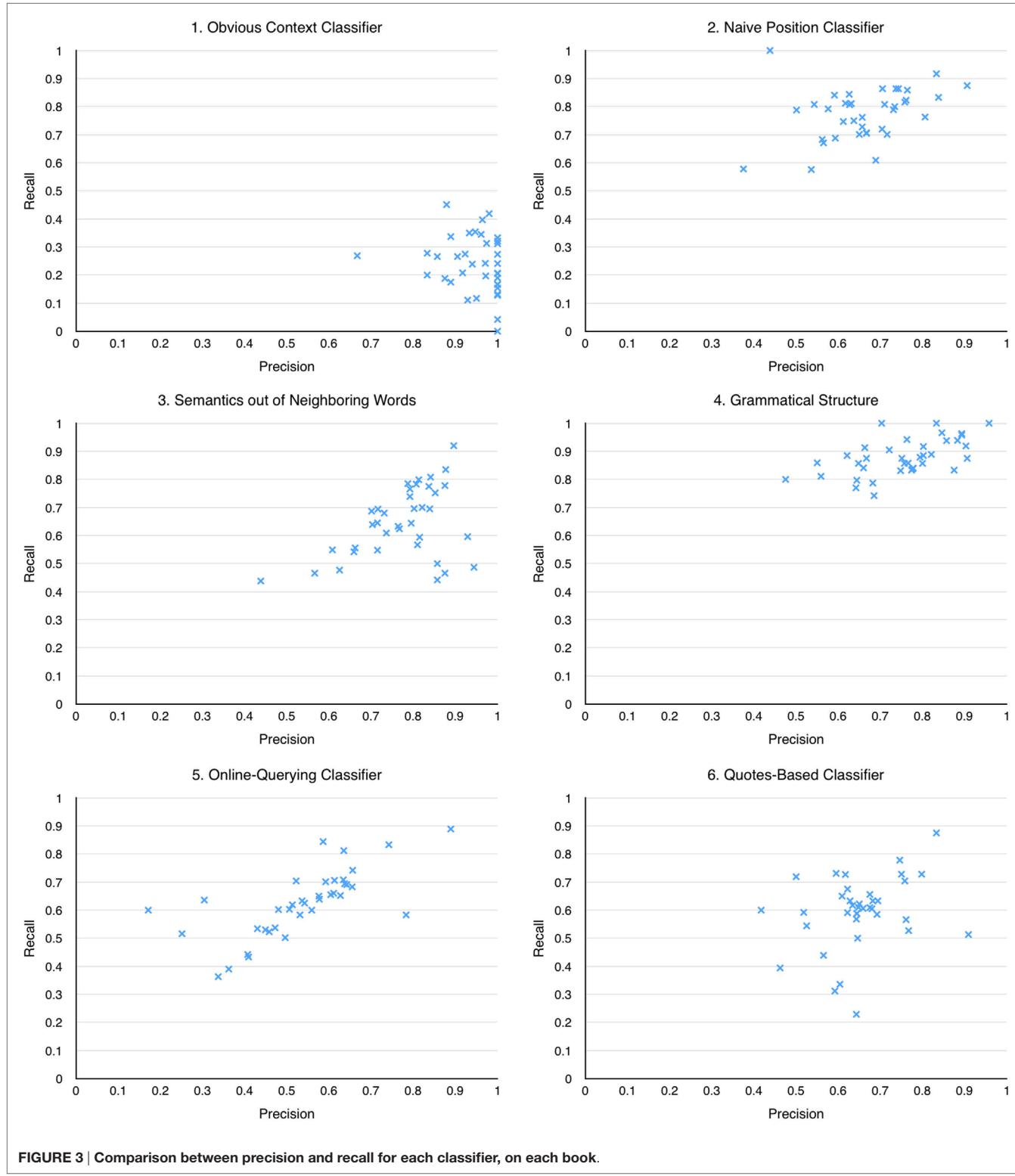
finds 6 neighboring words related to places, 3 to people and 1 to abstract concepts, the confidence index will be $\frac{6-1}{(6+3+1)} = 0.5$. This index is thus expected to equal 1 if the decision was made with no ambiguity and 0 if the clues were equally distributed.

- For *Naive Position*, *Grammatical*, and *Quotes* (sections 3.3, 3.5, and 3.7) classifiers, we use the difference between the splitting threshold and observed decision value, normalized between 0 and 1. For instance, considering the *Quotes* classifier computes

TABLE 2 | Precision and recall scores, per classifier and per class.

	1. Obvious context			2. Naive position			3. Semantics			4. Grammar			5. Wikipedia			6. Quotes		
	P	*R*	*F₁*	*P*	*R*	*F₁*	*P*	*R*	*F₁*	*P*	*R*	*F₁*	*P*	*R*	*F₁*	*P*	*R*	*F₁*
20000 Lieues sous les mers	**0.667**	0.269	0.3828	0.806	0.763	0.7839	0.736	0.609	0.6666	0.800	0.857	0.8275	0.627	0.652	0.6394	0.594	0.557	0.6567
Au bonheur des dames	0.947	0.354	0.5153	0.737	0.864	0.7956	0.839	0.695	0.7607	0.778	0.840	0.8077	0.592	0.701	0.6422	0.658	0.611	0.6033
Belle rose	0.980	0.419	0.5868	0.773	0.859	0.8087	0.835	0.812	0.8059	0.757	0.831	0.7876	0.347	**0.363**	0.3498	0.693	0.637	0.6404
Bijou	0.905	0.266	0.4107	**0.411**	0.578	**0.4549**	0.792	0.766	0.7784	0.565	0.859	0.6710	0.262	0.516	0.3378	0.508	0.719	0.5897
Blanche et Bleue	1.000	0.312	0.4762	0.761	0.823	0.7906	0.659	0.542	0.5946	0.802	0.885	0.8418	0.634	0.708	0.6689	0.602	0.406	**0.3379**
Boule de Suif	1.000	0.154	0.2667	0.543	0.808	0.6493	0.857	0.442	0.5835	0.621	0.885	0.7299	0.407	0.442	0.4240	0.595	0.731	0.6561
Candide	0.917	0.208	0.3395	**0.906**	0.875	**0.8904**	**0.944**	0.487	0.6428	0.906	0.875	0.8904	0.783	0.583	0.6684	**0.917**	0.551	0.6557
Colomba	1.000	0.167	0.2857	0.637	0.750	0.6892	0.662	0.556	0.6040	0.802	0.917	0.8556	0.577	0.639	0.6063	0.771	0.889	0.7615
La Comedie Humaine	0.882	**0.451**	**0.5959**	0.635	0.811	0.7098	0.703	0.641	0.6780	0.656	0.857	0.7379	0.552	0.633	0.5806	0.626	0.698	0.6478
Consuelo	0.940	0.239	0.3813	0.657	0.728	0.6905	0.802	0.696	0.7451	0.682	0.787	0.7311	0.576	0.650	0.6108	0.759	0.715	0.7299
David Copperfield	1.000	0.333	0.5000	0.593	0.688	0.6369	0.787	0.785	0.7859	0.763	0.942	0.8432	0.635	0.812	0.7126	0.624	0.605	0.6059
Don Quichotte	0.974	0.313	0.4733	0.579	0.792	0.6669	0.762	0.596	0.7258	0.564	0.811	0.6621	0.477	0.593	0.5342	0.613	0.659	0.6284
Germaine	0.889	0.337	0.4888	0.536	**0.576**	0.5550	0.608	0.549	0.5773	0.685	**0.742**	0.7126	0.604	0.655	0.6286	0.735	0.527	0.6243
Germinal	1.000	0.128	0.2266	0.731	0.789	0.7587	0.772	0.617	0.7120	0.821	0.889	0.8538	0.637	0.694	0.6646	0.469	0.411	0.4256
Germinie Lacerteux	0.833	0.200	0.3226	0.562	**1.000**	0.6087	0.857	0.500	0.6316	**0.500**	0.800	**0.5961**	**0.208**	0.600	**0.2667**	**0.417**	0.600	0.4918
Jane Eyre	1.000	0.205	0.3396	0.705	0.864	0.7760	0.896	0.920	0.9080	0.846	0.966	0.9021	0.613	0.706	0.6559	0.598	0.506	0.4943
La Bête Humaine	1.000	0.321	0.4858	0.833	0.917	0.8730	0.875	0.796	0.8050	0.857	0.938	0.8955	0.559	0.600	0.5789	0.661	0.588	0.6146
La Femme du mort	0.964	0.397	0.5624	0.508	0.788	0.6122	0.695	0.644	0.7046	0.741	0.905	0.8025	0.332	0.652	0.4111	0.544	0.673	0.5525
L'Assommoir	0.971	0.242	0.3880	0.743	0.864	0.7989	0.815	0.594	0.6871	0.794	0.879	0.8343	0.656	0.742	0.6964	0.617	0.603	0.6146
Le Mort vivant	1.000	0.207	0.3429	0.667	0.707	0.6862	0.810	0.567	0.6669	0.766	0.858	0.8093	0.496	0.502	0.4989	0.682	0.623	0.6404
Le Parfum de la dame en noir	1.000	0.274	0.4304	0.689	0.609	0.6467	0.703	0.639	0.6693	0.903	0.919	0.9112	0.655	0.683	0.6686	0.767	0.583	0.6494
Les Malheurs de Sophie	0.833	0.125	0.3088	0.617	0.812	0.7012	0.688	0.688	**0.4375**	0.750	0.938	0.9091	0.433	0.500	0.5804	0.833	**0.938**	**0.8537**
Les Miserables	0.933	0.350	0.5091	0.739	0.800	0.7655	0.820	0.775	0.8241	0.784	0.833	0.8024	0.424	0.442	0.4208	0.711	0.667	0.6624
Les Tribulations d'un Chinois	1.000	0.132	0.2340	0.710	0.808	0.7556	0.792	0.739	0.7646	0.757	0.857	0.8036	0.506	0.603	0.5500	0.708	0.718	0.6344
Les trois mousquetaires	0.923	0.275	0.4238	0.657	0.762	0.7061	0.852	0.752	0.7988	0.751	0.875	0.8085	0.449	0.530	0.4859	0.696	0.682	0.6653
Les Vacances	0.889	0.175	0.2924	0.562	0.683	0.6171	0.807	0.783	0.7950	0.667	0.875	0.7568	0.531	0.583	0.5562	0.667	0.683	0.6308
Le Tour du Monde en 80 jours	1.000	0.190	0.3200	0.724	0.701	0.7086	0.738	0.610	0.6925	0.894	0.939	0.9103	0.652	0.691	0.6654	0.771	0.685	0.7389
Le voluptueux voyage	1.000	0.042	0.0800	0.671	0.701	0.6744	0.875	0.466	0.6081	0.923	0.958	0.9244	0.420	0.436	0.3756	0.679	0.617	0.6256
L'Iliade	0.950	0.117	0.2082	0.620	0.747	0.6728	0.716	0.554	0.6202	0.651	0.797	0.7121	0.526	0.626	0.5608	0.571	0.342	0.4083
L'Odyssée	0.929	0.111	0.1985	0.627	0.807	0.7056	**0.532**	**0.430**	0.5109	0.642	0.770	0.7004	0.522	0.704	0.5995	0.593	**0.328**	0.4318
Madame Bovary	1.000	0.241	0.3881	0.758	0.816	0.7863	0.707	0.662	0.7047	0.893	0.963	0.9266	0.611	0.660	0.6343	0.505	0.439	0.6315

(Continued)

TABLE 2 | Continued

	1. Obvious context			2. Naive position			3. Semantics			4. Grammar			5. Wikipedia			6. Quotes		
	P	R	F₁	P	R	F₁	P	R	F₁	P	R	F₁	P	R	F₁	P	R	F₁
Les Mystères de Paris	0.961	0.345	0.5081	0.565	0.671	0.6133	0.877	0.835	0.8555	0.660	0.841	0.7397	0.472	0.537	0.5024	0.646	0.658	0.6287
Nana	1.000	0.212	0.3282	0.574	0.836	0.6941	0.638	0.644	0.6936	0.639	0.906	0.7682	0.420	0.513	0.4765	0.549	0.670	0.5344
Notre-Dame de Paris	0.857	0.266	0.4063	0.704	0.720	0.7118	0.768	0.631	0.6880	0.776	0.837	0.8055	0.742	0.833	0.7848	0.748	0.711	0.7614
Récits d'un soldat	1.000	**0.000**	**0.0000**	0.667	0.705	0.6851	0.625	0.477	0.5412	**0.958**	**1.000**	**0.9787**	0.458	0.523	0.4884	0.629	0.773	0.6674
Tarass Boulba	1.000	0.167	0.2857	0.626	0.844	0.7192	0.821	0.700	0.7559	0.703	**1.000**	0.8258	0.586	0.844	0.6918	0.650	0.622	0.6358
Thérèse Raquin	0.833	0.278	0.4167	0.838	0.833	0.8354	0.875	0.778	0.8235	0.875	0.833	0.8537	**0.889**	**0.889**	**0.8889**	0.733	0.556	0.5636
Average	0.949	0.239	0.369	0.656	0.774	0.706	0.771	0.643	0.695	0.751	0.881	0.808	0.525	0.617	0.562	0.652	0.598	0.615
SD	0.070	0.101	0.133	0.110	0.091	0.088	0.105	0.124	0.102	0.111	0.065	0.085	0.143	0.119	0.130	0.099	0.129	0.103

Lowest and highest scores are highlighted in bold.

a ratio of 0.4 for some noun with a threshold of 0.2, the result will be classified as a person with a confidence index of $\frac{0.4-0.2}{1-0.2} = 0.25$. Again, this index is expected to see its value tend toward 0 for ambiguous cases and toward 1 for the more definite ones.

On top of that, some classifiers are given the possibility to return 0 to mark their results as known to be invalid, and thus irrelevant at voting time. This can happen for instance when we do not find any known title preceding a word throughout the text, if no grammatical rule could be matched, or if Wikipedia does not have any result for the searched word.

The improved voting algorithm then first discards all classifications that have a confidence mark of 0 and proceeds to a simple vote between the remaining ones for each noun. In case of a tie, the results rated with the highest confidence will be privileged.

4.3. Fixed Weighting

Not all classifiers exhibit the same behavior regarding precision and recall. It thus can be justified to put more confidence on some of them in cases when we know they are more likely to succeed. For this test, we used manually set weights putting more importance to the *obvious context* classifier (section 3.2), due to its high precision rate, all others being treated equally. With the help of confidence rating (section 4.2), we know the low recall rate will not impact negatively on the other classifiers because it will return a confidence score of 0 if it could not find any classification clue. Hence, those cases will be discarded regardless of the coefficient. A good compromise can be reached by giving 3 times more weight to the *obvious context* classifier, allowing the others to still easily overpower it in the unlikely case a majority of them reach a contradictory agreement.

4.4. Meta-Recognition Approach to Optimized Weighting

A meta-recognition algorithm follows the idea of improving its accuracy by entirely removing one classifier if it detects it is consistently failing, typically due to stylistic biases or other broken assumptions on the considered book. Given the global classification results, one can easily compute an agreement score between the different classifiers, for instance using the Fleiss' Kappa method (Fleiss and Cohen, 1973). In case this indicator hints at discrepancy, we can simply recompute all averages by systematically letting out one of the classifiers, until the new Fleiss' Kappa value increased. Our hypothesis here is that since the remaining classifiers reached a higher agreement, the discarded one must have globally failed in some way and needs to be put aside.

5. RESULTS

Let us consider in **Figure 3** the precisions vs. recall results that each of the six classifiers achieved on our test corpus. One can immediately see a typical pattern in any information retrieval system: one parameter is detrimental to the other, and no two classifiers behave in a similar way. We can also see that for each
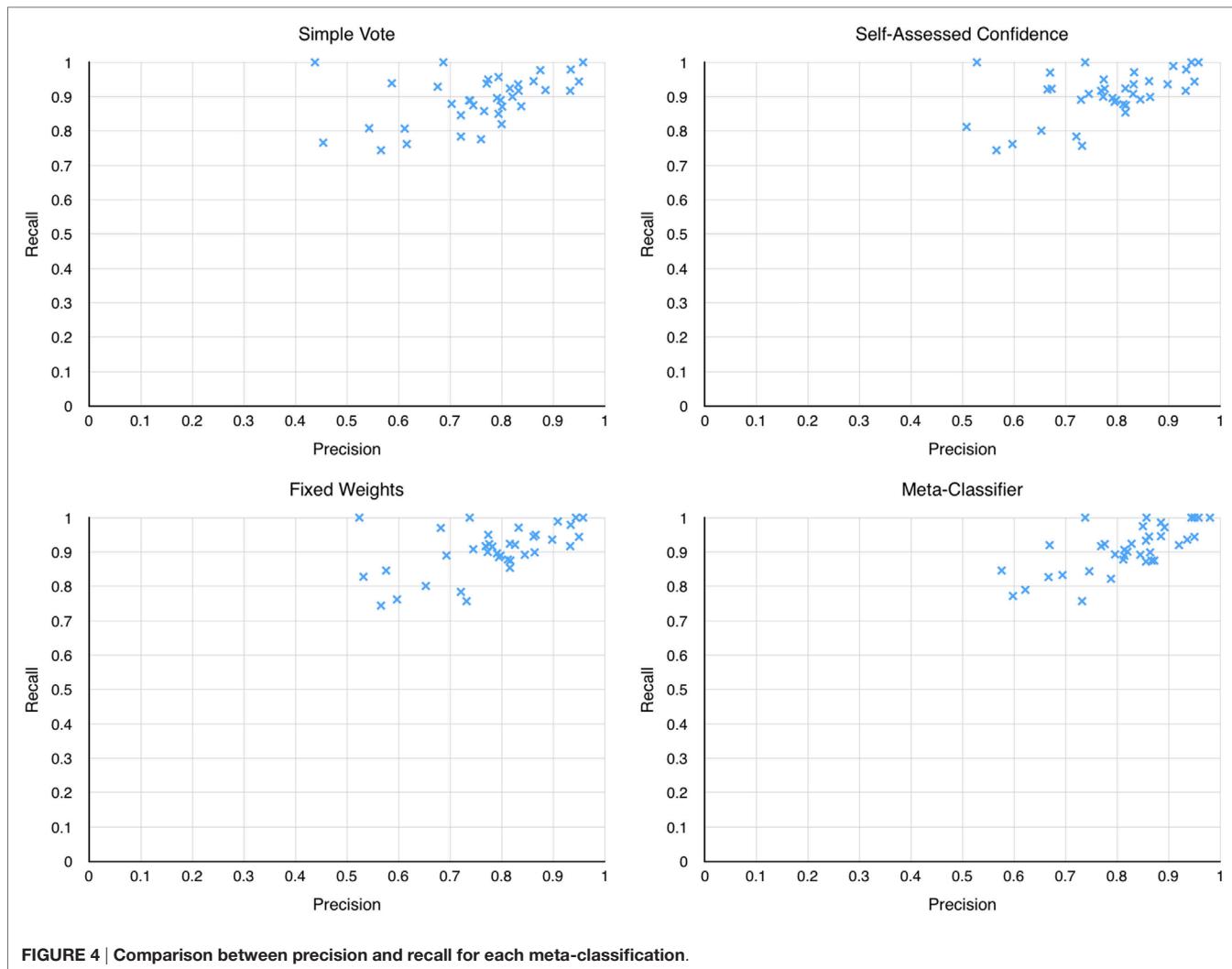
FIGURE 4 | Comparison between precision and recall for each meta-classification.
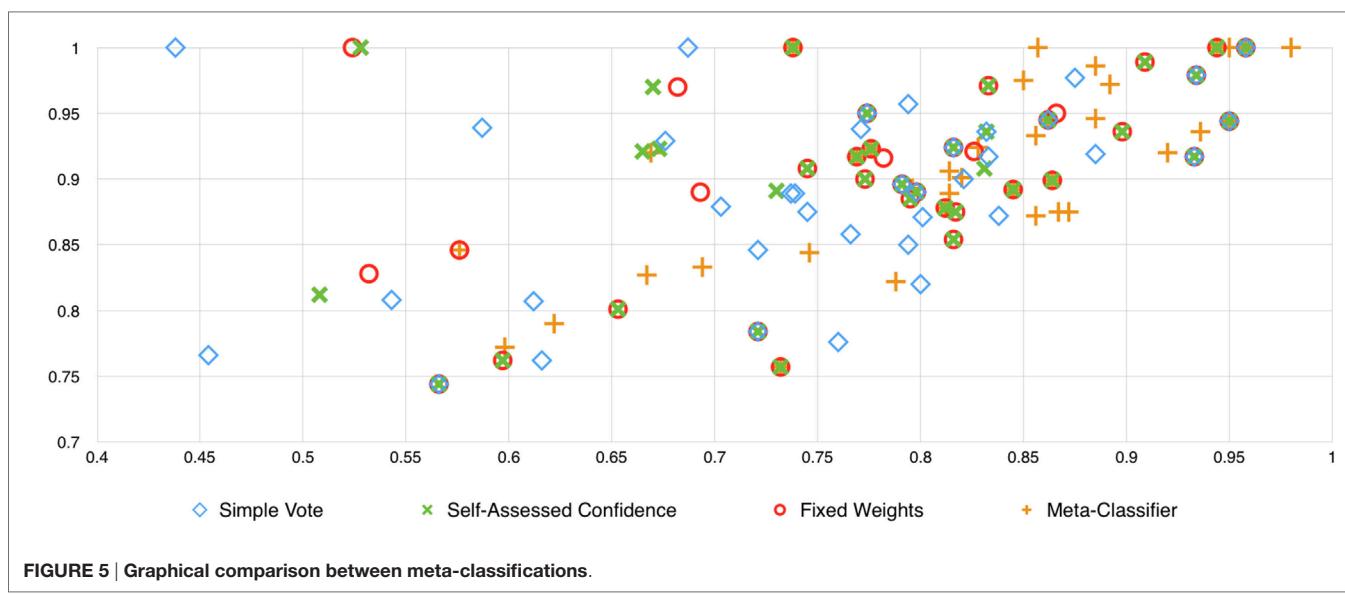
FIGURE 5 | Graphical comparison between meta-classifications.

TABLE 3 | Overall precision, recall, and F_1 scores.

	1. Obvious context				2. Naive position				3. Semantics				4. Grammar				5. Wikipedia				6. Quotes			
	P_{char}	P_{plac}	P_{char}	R_{plac}	P_{char}	P_{plac}	P_{char}	R_{plac}	P_{char}	P_{plac}	P_{char}	R_{plac}	P_{char}	P_{plac}	P_{char}	R_{plac}	P_{char}	P_{plac}	P_{char}	R_{plac}	P_{char}	P_{plac}	P_{char}	R_{plac}
20000 Lieues sous les mers	0.500	0.833	0.037	0.500	0.862	0.750	0.926	0.600	0.933	0.538	0.519	0.700	0.957	0.643	0.815	0.900	0.826	0.429	0.704	0.600	0.833	0.545	0.741	0.600
Au bonheur des dames	1.000	0.895	0.312	0.395	0.556	0.919	0.938	0.791	0.875	0.804	0.438	0.953	0.667	0.889	0.750	0.930	0.385	0.800	0.938	0.465	0.524	0.793	0.688	0.535
Belle rose	1.000	0.960	0.222	0.615	0.581	0.966	1.000	0.718	0.778	0.892	0.778	0.846	0.548	0.966	0.944	0.718	0.216	0.478	0.444	0.282	0.552	0.833	0.889	0.385
Bijou	1.000	0.810	0.000	0.531	0.100	0.722	0.750	0.406	0.750	0.833	0.750	0.781	0.308	0.821	1.000	0.719	0.115	0.409	0.750	0.281	0.231	0.786	0.750	0.688
Blanche et Bleue	1.000	1.000	0.375	0.250	0.577	0.944	0.938	0.708	0.545	0.773	0.375	0.708	0.652	0.952	0.938	0.833	0.571	0.696	0.750	0.667	0.455	0.750	0.312	0.500
Boule de Suif	1.000	1.000	0.000	0.308	0.286	0.800	1.000	0.615	1.000	0.714	0.500	0.385	0.333	0.909	1.000	0.769	0.100	0.714	0.500	0.385	0.333	0.857	1.000	0.462
Candide	1.000	0.833	0.000	0.417	0.812	1.000	1.000	0.750	1.000	0.889	0.308	0.667	0.812	1.000	1.000	0.750	0.565	1.000	1.000	0.167	0.833	1.000	0.769	0.333
Colomba	1.000	1.000	0.000	0.333	0.375	0.900	1.000	0.500	0.500	0.824	0.333	0.778	0.667	0.938	1.000	0.833	0.308	0.846	0.667	0.611	0.667	0.875	1.000	0.778
La Comedie Humaine	0.857	0.908	0.231	0.671	0.271	1.000	1.000	0.621	0.533	0.872	0.308	0.975	0.329	0.983	1.000	0.714	0.183	0.920	0.769	0.497	0.288	0.964	0.731	0.665
Consuelo	1.000	0.880	0.000	0.478	0.400	0.913	1.000	0.457	0.750	0.854	0.500	0.891	0.486	0.879	0.944	0.630	0.378	0.774	0.778	0.522	0.636	0.882	0.778	0.652
David Copperfield	1.000	1.000	0.000	0.667	0.185	1.000	1.000	0.377	0.636	0.938	0.700	0.870	0.526	1.000	1.000	0.884	0.270	1.000	1.000	0.623	0.357	0.891	0.500	0.710
Don Quichotte	1.000	0.947	0.286	0.340	0.219	0.939	1.000	0.585	0.667	0.857	0.286	0.906	0.259	0.868	1.000	0.623	0.147	0.806	0.714	0.472	0.312	0.914	0.714	0.604
Germaine	1.000	0.778	0.091	0.583	0.500	0.571	0.818	0.333	0.667	0.550	0.182	0.917	0.643	0.727	0.818	0.667	0.571	0.636	0.727	0.583	0.636	0.833	0.636	0.417
Germinal	1.000	1.000	0.056	0.200	0.583	0.878	0.778	0.800	0.700	0.844	0.389	0.844	0.643	1.000	1.000	0.778	0.405	0.870	0.944	0.444	0.250	0.688	0.333	0.489
Germinie Lacerteux	1.000	0.667	0.000	0.400	0.500	0.625	1.000	1.000	1.000	0.714	0.000	1.000	0.250	0.750	1.000	0.600	0.167	0.250	1.000	0.200	0.333	0.500	1.000	0.200
Jane Eyre	1.000	1.000	0.000	0.409	0.409	1.000	1.000	0.727	0.818	0.974	1.000	0.841	0.692	1.000	1.000	0.932	0.267	0.958	0.889	0.523	0.333	0.862	0.444	0.568
La Bête Humaine	1.000	1.000	0.100	0.542	0.667	1.000	1.000	0.833	0.800	0.950	0.800	0.792	0.714	1.000	1.000	0.875	0.368	0.750	0.700	0.500	0.421	0.900	0.800	0.375
La Femme du mort	1.000	0.929	0.400	0.394	0.286	0.731	1.000	0.576	0.625	0.765	0.500	0.788	0.692	0.789	0.900	0.909	0.294	0.370	1.000	0.303	0.421	0.667	0.800	0.545
L'Assommoir	1.000	0.941	0.000	0.485	0.526	0.960	1.000	0.727	0.667	0.963	0.400	0.788	0.588	1.000	1.000	0.758	0.370	0.941	1.000	0.485	0.400	0.833	0.600	0.606
Le Mort vivant	1.000	1.000	0.000	0.414	0.333	1.000	1.000	0.414	0.667	0.952	0.444	0.690	0.571	0.960	0.889	0.828	0.227	0.765	0.556	0.448	0.455	0.909	0.556	0.690
Le Parfum de la dame en noir	1.000	1.000	0.000	0.548	0.490	0.889	0.960	0.258	0.682	0.724	0.600	0.677	0.806	1.000	1.000	0.839	0.621	0.690	0.720	0.645	0.867	0.667	0.520	0.645
Les Malheurs de Sophie	1.000	0.667	0.000	0.250	0.400	0.833	1.000	0.625	0.500	0.875	0.500	0.875	0.500	1.000	1.000	0.875	0.200	0.667	0.500	0.500	0.667	1.000	1.000	0.875
Les Miserables	1.000	0.867	0.267	0.433	0.527	0.950	0.967	0.633	0.810	0.831	0.567	0.983	0.697	0.871	0.767	0.900	0.273	0.575	0.500	0.383	0.564	0.857	0.733	0.600
Les Tribulations d'un Chinois	1.000	1.000	0.111	0.154	0.692	0.727	1.000	0.615	0.833	0.750	0.556	0.923	0.680	0.833	0.944	0.769	0.600	0.412	0.667	0.538	0.750	0.667	0.667	0.769
Les trois mousquetaires	1.000	0.846	0.000	0.550	0.475	0.840	1.000	0.525	0.917	0.787	0.579	0.925	0.594	0.909	1.000	0.750	0.342	0.556	0.684	0.375	0.625	0.767	0.789	0.575
Les Vacances	1.000	0.778	0.000	0.350	0.250	0.875	0.667	0.700	0.667	0.947	0.667	0.900	0.333	1.000	1.000	0.750	0.154	0.909	0.667	0.500	0.333	1.000	0.667	0.650
Le Tour du Monde en 80 jours	1.000	1.000	0.000	0.381	0.755	0.692	0.974	0.429	0.840	0.636	0.553	0.667	0.925	0.864	0.974	0.905	0.784	0.520	0.763	0.619	0.773	0.769	0.895	0.476

(Continued)

TABLE 3 | Continued

	1. Obvious context				2. Naive position				3. Semantics				4. Grammar				5. Wikipedia				6. Quotes			
	P_{char}	P_{plac}	R_{char}	R_{plac}	P_{char}	P_{plac}	R_{char}	R_{plac}	P_{char}	P_{plac}	R_{char}	R_{plac}	P_{char}	P_{plac}	R_{char}	R_{plac}	P_{char}	P_{plac}	R_{char}	R_{plac}	P_{char}	P_{plac}	R_{char}	R_{plac}
Le voluptueux voyage	1.000	1.000	0.000	0.083	0.643	0.700	0.818	0.583	0.182	0.750	1.000	0.750	0.846	1.000	1.000	0.917	0.385	0.455	0.455	0.417	0.643	0.714	0.818	0.417
L'Iliade	1.000	0.900	0.000	0.234	0.333	0.907	0.857	0.636	0.238	0.807	0.625	0.870	0.404	0.898	0.905	0.688	0.262	0.791	0.810	0.442	0.438	0.737	0.333	0.364
L'Odyssée	1.000	0.857	0.000	0.222	0.351	0.902	0.929	0.685	0.286	0.756	0.308	0.574	0.342	0.943	0.929	0.611	0.286	0.759	1.000	0.407	0.444	0.769	0.286	0.370
Madame Bovary	1.000	1.000	0.000	0.481	0.600	0.917	0.818	0.815	0.545	0.913	0.500	0.778	0.786	1.000	1.000	0.926	0.421	0.800	0.727	0.593	0.261	0.750	0.545	0.333
Les Mystères de Paris	1.000	0.922	0.118	0.573	0.227	0.903	1.000	0.341	0.706	0.898	0.857	0.963	0.386	0.933	1.000	0.683	0.183	0.761	0.647	0.427	0.400	0.893	0.706	0.610
Nana	1.000	1.000	0.143	0.281	0.233	0.915	1.000	0.672	0.429	0.902	0.375	0.859	0.333	0.945	1.000	0.812	0.100	0.741	0.714	0.312	0.217	0.870	0.714	0.625
Notre-Dame de Paris	1.000	0.714	0.167	0.366	0.508	0.900	1.000	0.439	0.433	0.723	0.812	0.829	0.617	0.935	0.967	0.707	0.651	0.833	0.933	0.732	0.741	0.756	0.667	0.756
Récits d'un soldat	1.000	1.000	0.000	0.000	0.833	0.500	0.909	0.500	0.455	0.250	1.000	0.500	0.917	1.000	1.000	1.000	0.750	0.167	0.545	0.500	0.857	0.400	0.545	1.000
Tarass Boulba	1.000	1.000	0.000	0.333	0.364	0.889	0.800	0.889	0.400	0.643	1.000	1.000	0.714	0.692	1.000	1.000	0.444	0.727	0.800	0.889	0.800	0.500	0.800	0.444
Thérèse Raquin	1.000	0.667	0.333	0.222	0.875	0.800	0.778	0.889	0.556	0.750	1.000	1.000	1.000	0.750	0.667	1.000	0.889	0.889	0.889	0.889	0.800	0.667	0.444	0.667
Average	0.983	0.908	0.088	0.389	0.475	0.858	0.936	0.615	0.480	0.796	0.747	0.816	0.600	0.909	0.950	0.805	0.380	0.694	0.753	0.493	0.526	0.786	0.680	0.567

of them, some books get incredibly good results, and few others turn out very bad. Interestingly and as backed up by the full numerical values shown in **Table 2**, those are almost never the same, confirming our hypothesis that some methods may work way better (or worse) on some texts, giving a strong justification for the multi-classifier Mcapproach. The averaged results seem to confirm this intuition. In **Figures 4** and **5**, we can see that all meta-classification schemes overall pushed the results toward the top, and at the same time made the clustering denser, hence reducing the differences between the books and output more consistant results by removing the worse outliers.

By looking at the numerical results (**Tables 2** and **3**), several interesting facts can be stated about each classifier:

- Classifier #1: as one may expect with this kind of simple implementation, the resulting predictions usually translate to a very high precision (100% in nearly 2/3 of the cases, and on average 0.949) but a poor recall rate (only 0.239), as the presence of enough context words leaves little doubt on the categorization, though unlikely to happen for most of the nouns we would want to disambiguate. It is also interesting to see that its highest score was achieved on *La Comédie Humaine*, one of our longest samples, with an F_1 score of 0.596. Actually, all three sagas ranked quite high compared to the shorter books (*Les Misérables* and *Les Mystères de Paris* both got an F_1 score higher than 0.5).
- Classifier #2: unsurprisingly, this classifier is one of those that have the overall worst performance when compared to the others. However, one has to note that with a precision and recall of respectively 0.654 and 0.773, this very simple method still outperforms possible more complex ones.
- Classifier #3: the semantics approach surprisingly did not perform as well as one may have expected. Its precision of 0.769 is the second highest, but it comes with a cost of a recall of 0.651, which brings its F_1 score down even below the one of #2.
- Classifier #4: the grammar classifier resulted in the overall highest F_1 score (0.807), with a precision of 0.750 and a recall of 0.879. One can notice it performed best, with a very satisfying 0.909 F_1 score, on *Les Malheurs de Sophie*, which could be explained by since the book was written for children, its syntax might be more regular than other novels.
- Classifier #5: the quotes classifier did not perform very well. Its F_1 score (0.568) is the one to lowest and it has the lowest precision rate of all (0.528). Yet it still achieved a very good score on the novel *Thérèse Raquin*, where it actually got the highest precision, recall, and F_1 scores amongst all classifiers.
- Classifier #6: this method resulted in a rather disappointing overall precision and recall of 0.644 and 0.597, respectively, which can be explained by the expected lack of fictional entities represented on Wikipedia, and among those who seem to be there, many would actually relate to real-world counterparts that may be very different from the fictional use of the same word.

Regarding the meta-classification schemes (**Table 4**), we first wanted to compare them to our the *OpeNER* baseline. On our test corpus and using similar evaluation *OpeNER* averaged a

TABLE 4 | Averaged precision, recall, and *F*₁ scores, for each classifier.

	OpenNER			Simple vote			Self-assessment			Fixed weights			Meta-recognition		
	P	*R*	*F₁*	*P*	*R*	*F₁*	*P*	*R*	*F₁*	*P*	*R*	*F₁*	*P*	*R*	*F₁*
20000 Lieues sous les mers	0.696	0.858	0.769	0.760	0.776	0.768	0.732	0.757	0.744	0.732	0.757	0.744	0.732	**0.757**	0.744
Au bonheur des dames	0.572	0.795	0.665	0.862	0.945	0.902	0.862	0.945	0.902	0.862	0.945	0.902	0.862	0.945	0.902
Belle rose	0.645	0.638	0.641	0.832	0.936	0.881	0.832	0.936	0.881	0.826	0.921	0.871	0.892	0.972	0.931
Bijou	0.569	0.788	0.661	0.454	0.766	**0.570**	**0.508**	0.812	**0.625**	0.532	0.828	0.648	0.872	0.875	0.874
Blanche et Bleue	0.616	0.714	0.661	0.833	0.917	0.873	0.795	0.885	0.838	0.795	0.885	0.838	0.814	0.906	0.858
Boule de Suif	0.524	0.857	0.650	0.543	0.808	0.649	0.673	0.923	0.779	0.576	0.846	0.685	**0.576**	0.846	0.685
Candide	0.895	0.778	0.832	0.933	0.917	0.925	0.933	0.917	0.925	0.933	0.917	0.925	0.920	0.920	0.920
Colomba	0.632	0.864	0.730	0.739	0.889	0.807	0.769	0.917	0.836	0.769	0.917	0.836	0.769	0.917	0.836
La Comedie Humaine	0.545	0.721	0.621	0.737	0.889	0.806	0.745	0.908	0.819	0.745	0.908	0.819	0.746	0.844	0.792
Consuelo	0.679	0.787	0.729	0.791	0.896	0.840	0.791	0.896	0.840	0.791	0.896	0.840	0.820	0.901	0.859
David Copperfield	0.638	0.877	0.739	0.794	0.957	0.868	0.833	0.971	0.897	0.833	0.971	0.897	0.885	0.986	0.932
Don Quichotte	0.477	0.782	0.593	0.566	**0.744**	0.643	0.566	**0.744**	0.643	0.566	**0.744**	**0.643**	0.598	0.772	**0.674**
Germaine	0.610	0.694	0.649	0.721	0.784	0.751	0.721	0.784	0.751	0.721	0.784	0.751	0.788	0.822	0.805
Germinal	0.683	0.692	0.687	0.794	0.850	0.821	0.812	0.878	0.844	0.812	0.878	0.844	0.812	0.878	0.844
Germinie Lacerteux	0.321	0.600	0.418	**0.438**	1.000	0.609	0.528	1.000	0.691	**0.524**	1.000	0.688	0.857	1.000	0.923
Jane Eyre	0.619	0.891	0.731	0.875	0.977	0.923	0.909	0.989	0.947	0.909	0.989	0.947	0.950	1.000	0.974
La Bête Humaine	0.683	0.711	0.697	0.934	0.979	0.956	0.934	0.979	0.956	0.934	0.979	0.956	**0.980**	**1.000**	**0.990**
La Femme du mort	0.713	0.822	0.764	0.587	0.939	0.723	0.670	0.970	0.793	0.682	0.970	0.801	0.669	0.920	0.775
L'Assommoir	0.462	0.623	0.531	0.816	0.924	0.867	0.816	0.924	0.867	0.816	0.924	0.867	0.814	0.889	0.850
Le Mort vivant	0.718	0.872	0.788	0.766	0.858	0.809	0.816	0.854	0.835	0.816	0.854	0.835	0.856	0.872	0.864
Le Parfum de la dame en noir	0.811	0.878	0.843	0.885	0.919	0.902	0.864	0.899	0.882	0.864	0.899	0.882	0.864	0.899	0.882
Les Malheurs de Sophie	**0.267**	**0.571**	**0.364**	0.771	0.938	0.846	0.944	1.000	0.971	0.944	1.000	0.971	0.944	1.000	0.971
Les Miserables	0.693	0.800	0.743	0.821	0.900	0.859	0.831	0.908	0.868	0.866	0.950	0.906	0.856	0.933	0.893
Les Tribulations d'un Chinois	0.569	0.654	0.609	0.721	0.846	0.779	0.776	0.923	0.843	0.776	0.923	0.843	0.776	0.923	0.843
Les trois mousquetaires	0.510	0.758	0.610	0.745	0.875	0.805	0.773	0.900	0.832	0.773	0.900	0.832	0.828	0.924	0.873
Les Vacances	0.430	0.711	0.536	0.774	0.950	0.853	0.774	0.950	0.853	0.774	0.950	0.853	0.850	0.975	0.908
Le Tour du Monde en 80 jours	0.802	0.903	0.850	0.800	0.820	0.810	0.845	0.892	0.868	0.845	0.892	0.868	0.845	0.892	0.868
Le voluptueux voyage	0.646	0.625	0.635	0.801	0.871	0.835	0.817	0.875	0.845	0.817	0.875	0.845	0.867	0.875	0.871
L'Iliade	0.494	0.583	0.535	0.616	0.762	0.681	0.653	0.801	0.720	0.653	0.801	0.720	0.694	0.833	0.757
L'Odyssée	0.604	0.631	0.617	0.612	0.807	0.696	0.597	0.762	0.669	0.597	0.762	0.669	0.622	0.790	0.696
Madame Bovary	0.471	0.667	0.552	0.838	0.872	0.855	0.898	0.936	0.917	0.898	0.936	0.917	0.936	0.936	0.936
Les Mystères de Paris	0.542	0.800	0.646	0.703	0.879	0.781	0.730	0.891	0.803	0.782	0.916	0.844	0.885	0.946	0.915
Nana	0.524	0.723	0.608	0.676	0.929	0.782	0.665	0.921	0.772	0.693	0.890	0.779	0.667	0.827	0.738
Notre-Dame de Paris	0.721	0.772	0.746	0.798	0.890	0.842	0.798	0.890	0.842	0.798	0.890	0.842	0.796	0.893	0.842
Récits d'un soldat	**0.923**	**1.000**	**0.960**	**0.958**	**1.000**	**0.979**	**0.958**	**1.000**	**0.979**	**0.958**	**1.000**	**0.979**	0.958	1.000	0.979
Tarass Boulba	0.571	0.778	0.659	0.687	1.000	0.814	0.738	1.000	0.849	0.738	1.000	0.849	0.738	1.000	0.849
Thérèse Raquin	0.675	0.667	0.671	0.950	0.944	0.947	0.950	0.944	0.947	0.950	0.944	0.947	0.950	0.944	0.947
Average	0.608	0.756	0.669	0.749	0.889	0.809	0.775	0.904	0.831	0.776	0.903	0.831	0.815	0.907	0.857
SD	0.137	0.103	0.116	0.127	0.071	0.096	0.115	0.069	0.088	0.117	0.070	0.090	0.103	0.067	0.082

Lowest and highest scores are higlighted in bold.

precision and recall of 0.609 and 0.754, respectively. This may seem surprisingly low compared to the standards usually set by this tool, but actually is a good illustration of how difficult it may get to find the correct tagging in fictional texts. The two

worst cases (*Les Malheurs de Sophie* and *Germinie Larcerteux*) are shown in details in **Tables A1** and **A2** (Appendix A), and show that those problems are about as much related to bad classifications as missing entities. Our meta-classification methods

perform on average better than that, and we can see an encouraging trend that the various strategies we tried out tended to get increasingly better results and to close the standard deviation gap. That being said, as far as SD is concerned and taken into account, this improvement was not statistically significative. Yet, the meta-recognition method managed a F_1 score no worse than 0.674 and as high as 0.99 on one of the books, on which it basically found all entities and only misclassified one of them.

6. CONCLUSION AND FUTURE WORKS

The contribution of this paper is to establish a set of efficient and autonomous tools that can be run in a limited environment (such as a web server), on any French novel without the need of training or manual user input, yet keeping reliable results. We showed that combining different classifiers, especially using a meta-recognition technique, allowed to attain an overall better score than each of them would separately, and to outperform some state-of-the-art tools in the very narrow considered use case. Yet, we are aware this process implies several hard assumptions that may break under certain circumstances and cause the

system to fail, like the systematic extraction limited to capitalized proper nouns.

Future works may focus on coreference resolution by merging entities that refer to the same character, or conversely, disambiguation of homonyms, even if this does not happen very often in closed environments like fictional works. In parallel to a reliable extraction of the characters networks, further textual analytics may try to uncover the nature of the relations between them.

AUTHOR CONTRIBUTIONS

CB initiated this research, wrote the software components, and elaborated the test sets and methods described in this report. FK supervised this research and advised on the necessary tests and developments needed to improve its relevance.

ACKNOWLEDGMENTS

REFERENCES

Albanese, A. (2011). Small demons makes big splash at Frankfurt book fair. *Publishers Weekly* 258.43: 6–8.

Azpeitia, A., Cuadros, M., Gaines, S., and Rigau, G. (2014). NERC-fr: supervised named entity recognition for French. In *Text, Speech and Dialogue*, Vol. 8655, Edited by P. Sojka, A. Horák, I. Kopeček, and K. Pala, 158–165. Cham: Springer International Publishing.

Dumais, S.T. (2005). Latent semantic analysis. *Annual Review of Information Science and Technology* 38: 188–230. doi:10.1002/aris.1440380105

Fleiss, J.L., and Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement* 33: 613–9. doi:10.1177/001316447303300309

Geno, M.G. (1992). *A la française: Correct French for English Speakers*. London, New York: University Press of America.

Glass, K., and Bangay, S. (2007). *A Naïve, Salience-Based Method for Speaker Identification in Fiction Books*. Durban, South Africa: PRASA. 1–6.

Goh, H.-N., Soon, L.-K., and Haw, S.-C. (2012). Automatic identification of protagonist in fairy tales using verb. In *Advances in Knowledge Discovery and Data Mining*, Vol. 7302, Edited by D. Hutchison, T. Kanade, J. Kittler, J.M. Kleinberg, F. Mattern, J.C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M.Y. Vardi, G. Weikum, P.-N. Tan, S. Chawla, C.K. Ho, and J. Bailey, 395–406. Berlin, Heidelberg: Springer.

Grevisse, M., and Lenoble-Pinson, M. (2009). Majuscules. In *Le français Correct: Guide pratique des difficultés*, 6th ed, 88–92. Bruxelles: Duculot.

Hirschberg, J., and Manning, C.D. (2015). Advances in natural language processing. *Science* 349: 261–6. doi:10.1126/science.aaa8685

Jovanovic, J., Bagheri, E., Cuzzola, J., Gasevic, D., Jeremic, Z., and Bashash, R. (2014). Automated semantic tagging of textual content. *IT Professional* 16: 38–46. doi:10.1109/MITP.2014.85

Karsdorp, F., van Kranenburg, P., Meder, T., and van den Bosch, A. (2012). *Casting a Spell: Identification and Ranking of Actors in Folktales*. Lisbon, Portugal: Edições Colibri. 39–50.

Moretti, F. (2013). *Distant Reading*. London, New York: Verso.

Mosallam, Y., Abi-Haidar, A., and Ganascia, J.-G. (2014). Unsupervised named entity recognition and disambiguation: an application to old French journals. In *Industrial Conference on Data Mining*, 12–23. St. Petersburg: Springer.

Scheirer, W.J., Rocha, A., Micheals, R.J., and Boult, T.E. (2011). Meta-recognition: the theory and practice of recognition score analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33: 1689–95. doi:10.1109/TPAMI.2011.54

Si, L., Kanungo, T., and Huang, X. (2005). Boosting performance of bio-entity recognition by combining results from multiple systems. In *Proceedings of the 5th international workshop on Bioinformatics, BIOKDD '05* 76–83. Chicago: ACM.

Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology* 60: 538–56. doi:10.1002/asi.21001

APPENDIX

Appendix A

TABLE A1 | Benchmarked *OpeNER* results on "*Les Malheurs de Sophie*."

Entity	*OpeNER* result	Reference
Amérique	Place	Place
AUBERT	Character	N/A
Aubert	Place	N/A
Beau-Minon	Place	Character
Camille	Place	Character
Dieu	N/A	Other
Élisabeth	Place	Character
LAMBERT	Character	Character
Lambert	Place	Character
Louise	N/A	Character
M. de Réan	N/A	Character
Madeleine	Place	Character
Mme d'Aubert	N/A	Character
Mme de Réan	N/A	Character
Paris	Place	Place
Paul	Character	Character
Réan	Place	N/A
Sophie	Place	Character

TABLE A2 | Benchmarked *OpeNER* results on "*Germinie Lacerteux*."

Entity	*OpeNER*	Reference
Adèle	N/A	Character
Afrique	N/A	Place
Dieu	Character	Other
Gautruche	Place	Character
Germinie	Place	Character
Joseph	N/A	Character
Jupillon	Place	Character
M. de Varandeuil	N/A	Character
Mlle	Place	N/A
Mlle de Varandeuil	N/A	Character
Mme Jupillon	Character	Character
Montmartre	N/A	Place
Paris	Place	Place
Révolution	N/A	Other
Sempronie	Place	Character
Varandeuil	Place	N/A

Appendix B
B.1. Characters

titres monsieur, homme, mâle, mec, quidam, seigneur, sieur, madame, dame, mademoiselle, aimée, amie, amoureuse, bien-aimée, commandante, compagne, concubine, copine, dirigeante, dulcinée, employeuse, favorite, femme, fiancée, initiatrice, négrière, pédagogue, primordiale, reine, supérieure, prince, altesse, émir, archiduc, évêque, célébrité, cardinal, dauphin, diadoque, empereur, gloire, grand-duc, hospodar, infant, kronprinz, landgrave, maharadjah, maharajah, maharaja, margrave, monarque, monseigneur, principicule, rajah, rhingrave, roi, sire, souverain, autocrate, césar, chah, despote, dynaste, mogol, kaiser, khan, magnat, majesté, mikado, phénix, pharaon, ponte, potentat, raja, roitelet, sultan, trésor, abbé, capelan, corbeau, curé, curaillon, cureton, dignitaire, ecclésiastique, inquisiteur, inquisitrice, père, pontife, prédicant, prélat, prestolet, prieur, ratichon, révérend, religieux, pape, agréé, collecteur, décimateur, exacteur, fisc, impôt, baderne, belliqueux, caporal, colonel, conscrit, lieutenant, maréchal, martial, polémologique, rengagé, sergent, soldatesque, sous-officier, tactique, troupe, appelé, bachi-bouzouk, bidasse, biffin, bleusaille, bras, cavalier, champion, cipaye, combattant, conquérant, deuxième, pompe, drille, engagé, estafette, evzone, factionnaire, fantassin, franc-tireur, fusilier, garde-voie, goumier, grenadier, griveton, grognard, guérillero, guerrier, jalonneur, janissaire, légionnaire, méhariste, maquisard, mercenaire, mortepaye, palikare, pandour, papalin, partisan, patrouilleur, pertuisanier, pierrot, pionnier, pioupiou, piquier, planton, poilu, pourvoyeur, réquisitionnaire, résistant, reître, recrue, sentinelle, serviteur, soudard, soudrille, spahi, territorial, tirailleur, tommy, tourlourou, tringlot, troufion, troupier, vétéran, fils, aîné, élève, citoyen, descendant, disciple, fieu, fiston, fruit, géniture, garçon, garçonnet, gars, gosse, grand, jeune, jouvenceau, postérité, progéniture, race, rejeton, sang, fille, adolescente, bachelette, bambine, belle-fille, blondinette, bouchon, brebis, brunette, célibataire, catherinette, colombe, coureuse, courtisane, demoiselle, descendante, donzelle, fifille, fillasse, fillette, frangine, gamine, garçonne, gazille, gonzesse, héritière, hétaïre, jeunesse, jouvencelle, louloute, midinette, mignonne, nana, nénette, nonne, nymphette, petite, poulette, prostituée, quille, religieuse, rosière, soeur, soeurette, souris, suivante, tendron, trottin, typesse, bébé, baby, bambin, chiard, enfançon, enfant, gamin, môme, marmaille, marmot, mioche, mouflette, moutard, moutatchou, nourrisson, nouveau-né, petit, petit-salé, poupée, poupard, poupon, têtard, mari, époux, conjoint, épouse, bourgeoise, conjointe, déesse, doudou, ménagère, mariée, mousmé, m, mr, mme, mrs, miss **parole** accent, accoucher, écorcher, émouvoir, énoncer, apprendre, argot, articulation, articuler, avouer, babiller, bafouiller, bégayer, balbutier, baragouiner, bavarder, blaguer, bredouiller, calomnier, causer, chevroter, chuchoter, citer, communiquer, confabuler, conférer, consulter, converser, débattre, débit, débiter, déblatérer, déclamer, dégoiser, dénoncer, deviser, dialecte, dialoguer, diction, dire, discourir, discuter, disserter, enregistrer, entretenir, exprimer, fasciner, franc-parler, frapper, giberner, gueuler, haranguer, idiome, intervenir, jaboter, jacter, jargon, jargonner, jaser, jaspiner, jobelin, langage, langue, médire, marmonner, marmotter, mentionner, murmurer, nasiller, nommer, palabrer, parlementer, parloter, parole, pérorer, parler, patois, phonation, plaire, pratiquer, prononcer, prononciation, rabâcher, raconter, radoter, révéler, relater, retracer, rognonner, abandonner, adresser, épancher, confier, expliquer, style, usage, usance, verbe, vociférer, aboyer, acclamer, accuser, éclater, égosiller, annoncer, apostropher, appeler, attraper, avertir, bêler, beugler, brailler, braire, bramer, bruire, chahuter, chanter, clabauder, clamer, conspuer, couiner, craquer, criailler, croasser, feuler, gémir, glapir, grincer, grogner, gronder, héler, houper, hucher, hululer, hurler, implorer, interpeller, invectiver, invoquer, jurer, manifester, meugler, mugir, pépier, piailler, plaindre, pleurer, pousser, prévenir, prier, proclamer, proférer, protester, publier, réclamer, répandre, réprimander, retentir, rouspéter, rugir, écrier, époumoner, fâcher,

récrier, sermonner, exclamer, signaler, indigner, supplier, tancer, tempêter, tonitruer, tonner, vagir **pensée** faire, convaincu, avis, persuadé, admettre, élaborer, étudier, aviser, évoquer, cogiter, combiner, concevoir, conjecturer, considérer, contempler, délibérer, devoir, envisager, espérer, estimer, examiner, faillir, attention, gamberger, idée, imaginer, juger, méditer, manquer, mûrir, ordonnancer, pensée, peser, présumer, prévoir, prendre, projeter, rêvasser, rêver, réfléchir, raisonner, rappeler, recueillir, regarder, repenser, rouler, saisir, scruter, concentrer, douter, figurer, sentir, préoccuper, proposer, représenter, souvenir, targuer, intéresser, occuper, songer, souhaiter, soupçonner, spéculer, supposer, trouver **nutrition** absorber, alimentation, attaquer, avaler, bâfrer, becqueter, bouffe, bouffer, boulotter, boustifailler, brichetonner, briffer, brouter, chipoter, claquer, collationner, consommer, consumer, corroder, croûte, croûter, croquer, croustiller, cuisine, débrider, décolorer, déguster, déjeuner, dîner, dépenser, déteindre, dévorer, dilapider, dissiper, engloutir, engouffrer, entamer, disparaître, faner, festoyer, fricasser, fricoter, friper, gaspiller, goûter, gober, gobichonner, godailler, goinfrer, grailler, grappiller, grignoter, gruger, gueuletonner, ingérer, ingurgiter, mâcher, mangeotter, mastiquer, mets, mettre, mordre, nourrir, nourriture, oublier, paître, pâture, picorer, pignocher, pitance, prodiguer, repas, ripailler, ronger, ruiner, alimenter, sauter, savourer, bourrer, décarêmer, gaver, goberger, gorger, lester, empiffrer, emplir, piffrer, régaler, rassasier, remplir, repaître, restaurer, sustenter, souper, sucer, tâter, tordre, tortiller, tortorer, toucher, transgresser **émotions** amour, ébullition, agité, agitation, agressif, agressivité, aigreur, algarade, animosité, éréthisme, atrabilaire, atrabile, bile, bilieux, bouffée, bourrasque, chagrin, colérique, courroucé, courroux, crise, déchaînement, effervescence, emportement, exaspéré, exaspération, excitation, explosion, fâcherie, foudre, fulminant, fureur, furie, furieux, grogne, haine, hargne, hargneux, humeur, impatience, impatient, indignation, irascibilité, ire, irrité, irritabilité, irritable, mécontentement, monté, péché, querelle, rage, rageur, rancoeur, représailles, rogne, scène, surexcitation, susceptibilité, tempête, vengeance, violence, accouplement, adoration, adultère, affect, affection, altruisme, amativité, amitié, amourette, ange, aphrodite, archerot, éros, association, attachement, attraction, aventure, babiole, badinage, bagatelle, béguin, baibatifolage, biquet, biquette, bluette, bouillonnement, bricole, caprice, chaleur, charité, coït, coeur, concubinage, concupiscence, conquête, copulation, coquetterie, cupidon, débauche, délicatesse, désir, dévotion, dévouement, dilection, engouement, entente, enthousiasme, estime, faible, fanatisme, fantaisie, ferveur, fièvre, flamme, fleurette, flirt, folie, fréquentation, fraternité, galanterie, goût, grâce, hyménée, hymen, idolâtrie, inceste, inclination, intérêt, intrigue, ivresse, lascivité, libertinage, luxure, maladie, mariage, marivaudage, mouvement, mysticisme, passade, passion, passionnette, penchant, philanthropie, piété, plaisir, pulsion, relation, rut, sens, sensibilité, sentiment, tendance, tendresse, toquade, touche, union, vénération, vénus, agrément, allégresse, amusement, épanouissement, ardeur, avantage, béatitude, bienêtre, bienfait, bonheur, consolation, contentement, délice, douceur, enchantement, enjouement, entrain, euphorie, exaltation, extase, exultation, félicité, fierté, folichonnerie, gaieté, gaité, griserie, hilarité, jouissance, liesplaisir, régal, réjouissance, ravissement, rayonnement, rigolade, sourire, abandon, abattement,

accablement, affliction, amertume, éplorement, épreuve, assombrissement, asthénie, bourdon, cafard, calamité, découragement, dépression, désabusement, désenchantement, désespérance, désespérer, désespoir, désolation, deuil, douleur, grisaille, laideur, lassitude, lypémanie, mal, malaise, maussaderie, mocheté, monotonie, navrement, neurasthénie, noir, noirceur, nostalgie, nuage, pauvreté, peine, platitude, sévérité, sombreur, souffrance, spleen, suspens, uniformité, affres, alarme, attente, crainte, désarroi, incertitude, inquiétude, peur, souci, transe, abattu, adjudicataire, égrotant, aliéné, alité, altéré, anéanti, anormal, atteint, avarié, cacochyme, chétif, cinglé, client, démoli, déprimé, dérangé, désaxé, détraqué, dingue, dolent, fada, fanatique, fatigué, fiévreux, fou, galeux, gâté, gâteux, grabataire, incommodé, indisposé, infirme, ladre, lépreuse, lépreux, maboul, maladif, malsain, mauvais, morbide, pâle, patient, patraque, perturbé, piqué, rachitique, raide, révolutionné, scrofuleux, secoué, sonné, souffrant, souffreteux, tifosi, timbré, toqué, traumatisé, valétudinaire, abhorrer, abominer, éructer, évacuer, chasser, cracher, débecter, dégobiller, dégorger, dégueuler, détester, exécrer, expectorer, expulser, fulminer, gerber, haïr, honnir, huer, lancer, mépriser, maudire, régurgiter, répudier, rejeter, rendre, gorge, restituer, siffler, souffler, apeurer, cajoler, déconcerter, désemparer, effaroucher, effrayer, embarrasser, gêner, geler, glacer, impressionner, inhiber, inquiéter, menacer, paralyser, terroriser, ébauché, angoissé, approximatif, capon, complexé, confus, craintif, démerdard, discret, douteux, effarouché, effarouchable, effrayé, embarrassé, farouche, flottant, flou, frileux, froussard, fumeux, gêné, gauche, hésitant, honteux, humble, imperceptible, incertain, indécis, indéfini, indéterminé, indistinct, intimidé, lâche, modeste, nébuleux, nuageux, obscur, peureux, poltron, pudibond, pusillanime, réservé, rougissant, sauvage, subtil, timoré, transi, vague, émoi, appréhensif, appréhension, circonspection, confusion, effacement, effarouchement, gêne, gaucherie, hésitation, honte, humilité, indécision, modestie, prémonition, prudence, pudeur, réserve, trac, ébranlé, affecté, affolé, alarmé, éméché, émotionné, éperdu, apitoyé, éploré, attendri, attristé, blessé, bondissant, bouleversé, bouleversant, déchiré, déchirant, empoigné, enflammé, excité, frappé, impressionné, inquiété, intéressé, ivre, pantelant, pathétique, remué, retourné, saisi, suffoqué, surexcité, touché, touchant, tremblant, écoeurement, affadissement, allergie, éloignement, accident, affligé, aigre, élégiaque, assombri, bidoche, bougon, boulet, bourru, cafardeux, chagriné, colère, coléreux, consterné, consternation, content, contrarié, contrariété, contrit, couenne, cuir, déboire, déception, déchirement, dégoût, dépit, dépité, déplaisir, désagrément, désappointement, désolé, détresse, douloureux, ennui, ennuyé, gémissant, grimaud, grincheux, grognon, hypocondriaque, hypocondrie, inconsolable, inquiet, larme, larmoyant, lugubre, mélancolie, malheur, maussade, misanthrope, misère, morne, morose, morosité, mortifié, navré, neurasthénique, pataquès, peau, peiné, pelage, pissevinaigre, plaintif, pleurard, pleureur, rabat-joie, rechigné, regret, revêche, sinistre, sombre, tégument, tourment, tracasserie, tribulation, triste, tristesse, anorexie, antipathie, aversion, satiété, blasement, démoralisation, déplaisance, exécration, hautle-coeur, horreur, humiliation, inappétence, indigestion, mépris, mortification, nausée, répugnance, répulsion, ressentiment, affolement, affres, alerte, émotion, angoisse, anxiété, cauchemar,

couardise, dédain, effroi, foire, frayeur, frisson, frousse, intimidation, lâcheté, malepeur, panique, pétoche, phobie, pleutrerie, poltronnerie, pusillanimité, saisissement, souleur, spectre, stupéfaction, terreur, timidité, trouille, venette, vertige, vigilance, préoccupation, ébahissement, abasourdissement, admiration, ahurissement, émerveillement, épatement, épouvante, étonnure, effarement, impression, pétrification, scandale, stupeur, surprise **foi** assurance, autorité, évangile, catéchisme, certitude, confession, confiance, conviction, créance, credo, croyance, déloyauté, dogme, droiture, espérance, exactitude, fausseté, fidélité, fourberie, franchise, hommage, honnêteté, honneur, loyauté, mystique, opinion, perfidie, promesse, religion, serment, sincérité, spiritualité, témoignage, vérité **tromper** éblouir, abuser, échauder, égarer, amuser, apaiser, arnaquer, étriller, éviter, bafouer, baiser, balancer, bercer, berner, blouser, bluffer, capter, carotter, circonvenir, cocufier, coiffer, confondre, consoler, couillonner, croire, décevoir, déguiser, déjouer, dépiter, désappointer, dissimuler, duper, embabouiner, emberlificoter, embobeliner, embobiner, empaumer, empiler, encorner, endormir, enfiler, engeigner, enjôler, enquinauder, entôler, entortiller, entourlouper, entuber, envelopper, errer, escroquer, estamper, exploiter, cocu, cornard, marcher, miroiter, feindre, feinter, ficher, flatter, flouer, foutre, frauder, frustrer, gonfler, gourer, imposer, induire, jobarder, jouer, leurrer, méprendre, mentir, moquer, mystifier, pigeonner, piper, promener, refaire, renarder, repasser, resquiller, ruser, séduire, surprendre, trahir, tricher, truquer **mort** tuer, abattre, achever, échiner, écraser, égorger, allonger, anéantir, épuiser, éreinter, assassiner, assommer, étendre, étouffer, étrangler, étriper, bousiller, buter, casser, chagriner, chouriner, crever, décapiter, décimer, dégeler, dégringoler, démolir, dépêcher, détruire, descendre, empoisonner, emporter, estourbir, exécuter, excéder, exténuer, exterminer, cesser, mourir, fatiguer, faucher, flinguer, foudroyer, fusiller, gâcher, guillotiner, harasser, immoler, juguler, lapider, lasser, liquider, lyncher, massacrer, meurtrir, miner, moissonner, navrer, nettoyer, noyer, occire, peiner, pendre, percer, poignarder, pourfendre, ratiboiser, refroidir, sacrifier, saigner, défaire, servir, stériliser, suicider, supplicier, supprimer, suriner, trucider, user, vanner, mort, écroulement, agonie, anéantissement, ankylosé, apathique, épuisé, éreinté, assassinat, éteint, évanoui, brisé, cadavre, camarde, chute, claqué, condamné, corps, crève, crevé, croupissant, décédé, décès, décomposition, défunt, délavé, dépouille, désert, détruit, cujus, destruction, disparition, disparu, dormant, effondrement, enterrement, enveloppe, esquinté, euthanasie, exécution, excédé, exténué, extinction, fade, fantôme, feu, fichu, figé, fin, fini, flapi, foutu, voyage, harassé, immobile, inanimé, inerte, inhabité, insensible, intimider, irrécupérable, languide, lessivé, macchab, macchabée, malemort, mânes, meurtre, mourant, néant, nécrosé, nuit, éternelle, ombre, parque, passé, perdu, perte, plat, plongeon, rétamé, recru, rendu, repos, éternel, restes, rompu, ruine, silencieux, sommeil, somnolent, stagnant, supplice, tombe, tombeau, torture, tranquille, trépas, trépassé, tué, usé, victime, vidé **guérir** échapper, adoucir, mieux, améliorer, calmer, châtier, cicatriser, corriger, couper, débarrasser, délivrer, désintoxiquer, estomper, tomber, pallier, réchapper, rambiner, relever, remédier, remettre, renaître, ressusciter, revivre, sauver, fermer, rétablir, soigner, soulager, traiter **dormir** coucher, dormir, dormailler, lambiner, lanterner, négliger, pioncer, ronfler, roupiller, sommeiller, somnoler **salutations** bonjour, adieu, bienvenue, bonsoir, salamalec, salut, ciao, bye, salutations **vouloir** accepter, accorder, acquiescer, affirmer, aimer, alléguer, ambitionner, appéter, aspirer, attendre, envie, briguer, commander, concéder, consentir, convoiter, craindre, décider, défendre, daigner, désirer, demander, disposition, entendre, envier, escompter, essayer, exiger, forme, gouverner, guigner, intention, interdire, lorgner, loucher, manière, nécessiter, objectif, ordonner, ordre, oser, permettre, pouvoir, prétendre, prescrire, résolution, recevoir, requérir, revendiquer, s'efforcer, s'essayer, soif, solliciter, soupirer, subir, tenir, tolérer, viser, vouloir **soupir, sanglot** soupir, plainte, sanglot, silence, remord, cri, gémissement, geignement, hoquet, jérémiade, larmoiement, pleur, pleurnichement, spasme **raison** allégation, équité, argument, argumentation, base, but, cause, cerveau, cervelle, compréhension, conception, connaissance, conscience, considérant, considération, couleur, dédommagement, démonstration, discernement, droit, entendement, esprit, excuse, explication, facultés, fondement, indice, intellect, intelligence, judiciaire, jugeote, justesse, justice, justification, lieu, logique, lucidité, lumière, méthode, mobile, modération, motif, occasion, office, origine, philosophie, pondération, pourquoi, prétexte, preuve, principe, principes, probabilité, probité, quoi, réfutation, raison, raisonnement, rapport, rationnel, rectitude, sagesse, satisfaction, tête, tact **métiers** abatteur, aboyeur, accessoiriste, accisien, acconier, accordéoniste, accordeur, accordeuse, accoucheur, accouveur, accouveuse, accréditeur, accréditeuse, accréditrice, accrocheur, acériculteur, acéricultrice, acheteur, acheteuse, aciériste, aconier, acousticienne, acousticien, acrobate, acteur, actrice, actuaire, acuponcteur, acuponctrice, acupuncteur, acupunctrice, adaptateur, adaptatrice, adjudicateur, adjudicatrice, administrateur, administratrice, aède, aérodynamicien, aéronaute, aérostier, afficheur, affichiste, affileur, affineur, affineuse, affranchisseur, affréteur, affuteur, affuteuse, agencier, agencière, agioteur, agréeur, agréeuse, agriculteur, agricultrice, agrobiologiste, agronome, aiguilleur, aiguillier, aiguiseur, aiguiseuse, ajusteur, ajusteuse, albatrier, alcade, alchimiste, alcoolier, alcoologue, alderman, alem, aleseur, aléseuse, alfatier, algébriste, algologue, alguazil, aliéniste, alisier, allergologiste, allergologue, allopathe, allumettier, allumettière, alphabétiseur, alphabétiseuse, alpiniste, altiste, amareyeur, amareyeuse, ambassadeur, ambassadrice, ambulancier, ambulancière, aménagiste, amidonnier, amidonnière, amodiateur, amphibie, amuseur, amuseuse, analyste, anatomiste, andrologue, anesthésiste, ânier, ânière, animalier, animalière, animateur, animatrice, annaliste, annonceur, annonceuse, annoncière, annoncier, annotateur, annotatrice, antenniste, anthropologue, antiquaire, aouteron, apiculteur, apicultrice, apiéceur, apiéceuse, aplatisseur, apothicaire, appareilleur, appareilleuse, appariteur, apponteur, apprentie, apprenti, apprêteur, apprêteuse, aquaculteur, aquacultrice, aquafortiste, aquarelliste, aquatintiste, aquiculteur, aquicultrice, arbalétrier, arbitragiste, arbitre, arboriculteur, arboricultrice, archéologue, archer, archetière, archetier, archiâtre, architecte, archiviste, ardoisier, ardoisière, argenteur, argenteuse, argentier, argotiste, arithméticien, armaillis, armateur, armoriste, armurier, aromaticienne, aromaticien, arpenteur, arpenteuse, arpète, arpette, arquebusier, arracheur, arrangeur, arrangeuse, arrêtiste, arrimeur, arrondisseur, arroseur, arroseuse, artificier, artilleur,

assainisseur, assemblagiste, assesseure, assesseur, assistant, assistante, assureur, assyriologue, astaciculteur, astacicultrice, astrologue, astronaute, astronauticien, astronome, astrophysicien, atomiste, attorney, aubergiste, audiencier, audit, auditeur, auditrice, aumônier, aurige, auriste, auteur, auteure, autocariste, automaticienne, automaticien, autoursier, autrice, avaleur, avaleuse, aventurier, aventurière, aviateur, aviatrice, aviculteur, avicultrice, avionneur, avitailleur, avocat, avocate, avoué, avouée, bachoteur, badigeonneur, baes, baesine, bagagiste, baigneur, baigneuse, baillie, bailli, baladin, balancier, balayeur, balayeuse, baleinier, baliseur, ballerine, bandagiste, banquier, banquière, barbier, barbière, barde, barmaid, barman, barmen, barragiste, barreur, barreuse, baryton, basculeur, basketteur, basketteuse, bassiste, bateleur, bateleuse, batelier, batelière, bâtisseur, bâtisseuse, bâtonnier, bâtonnière, batteur, batteuse, baudrier, bêcheur, bedeau, berger, bergère, bibliographe, bibliothécaire, bijoutier, bijoutière, billettiste, bimbelotière, bimbelotier, bineur, bioethicienne, bioethicien, biogéographe, biographe, biologiste, biométricienne, biométricien, biophysicienne, biophysicien, biscuitier, biscuitière, blanchisseur, blanchisseuse, bobineur, bobineuse, bobinière, bobinier, boiseur, boisselier, boitier, bombagiste, bombeur, bonne, bonnetier, bonnetière, bordier, bordière, botaniste, botteleur, botteleuse, bottier, bottière, boucanier, boucher, bouchère, boucheur, boucholeur, bouchonnier, bouchoteur, bouclier, boueur, bougnat, bouilleur, boulanger, boulangère, boulier, boulinier, bouquetier, bouquiniste, bourgmestre, bourreau, bourrelière, bourrelier, bourrier, boursier, boursière, bousilleur, bouteiller, bouteur, boutillier, boutonnier, boutonnière, bouvier, bouvière, boxeur, boxeuse, boyaudier, boyaudière, braconnier, brancardière, brancardier, brandevinière, brandevinier, brasseur, brasseuse, brillanteur, briqueteur, briquetier, brocanteur, brocanteuse, brocheur, brodeur, brodeuse, broker, bronzeur, bronzeuse, bronzière, bronzier, brossier, brossière, brouetteur, brouettier, broyeur, bruiteur, bruiteuse, brûleur, brunisseur, brunisseuse, buandier, buandière, bucheron, bucheronne, bucheur, buffetier, buffetière, bulbiculteur, bulbicultrice, buraliste, bureauticienne, bureauticien, burineur, buriniste, bustier, bustière, buvetier, cabaretière, cabaretier, cabinier, cabinotier, cableur, cableuse, cabliste, caboteur, cadreur, cadreuse, cafetier, cafetière, caid, caissier, caissière, calandreur, calandreuse, calfat, calicot, calier, calleur, calleuse, calligraphe, cambiste, cambreur, cambusier, camelot, cameraman, camerier, camerière, cameriste, camionneur, camionneuse, campaniste, canalisateur, canalisatrice, cancerologue, canetière, canissier, cannetière, canneur, canneuse, cannier, cannière, cannissier, canonnier, canotier, canotière, cantatrice, cantinier, cantinière, cantonnier, cantonnière, cantor, canuse, canut, capilliculteur, capitaine, carabinier, cardeur, cardeuse, cardiologue, caricaturiste, carillonneur, carillonneuse, cariste, carpettier, carreleur, carreleuse, carrier, carrossier, cartier, cartographe, cartonnier, cartonnière, cartons, cartooniste, cascadeur, cascadeuse, casernier, casquettier, casseur, casseuse, catcheur, catcheuse, catéchiste, caviste, cellerière, celliste, censeur, censeure, censière, censier, ceramiste, cerealier, chainetière, chainetier, chaineur, chaineuse, chainier, chainiste, chaisier, chaisière, chalcographe, chaloupier, chalutier, chambellan, chambrière, chambrier, chamelier, chamoiseur, chancelière, chandelier, changeur, changeuse, chansonnière, chansonnier,

chanteur, chanteuse, chantre, chanvrière, chanvrier, chapelier, chapelière, charbonnier, charcutier, charcutière, charpentier, charretier, charron, chartrier, chasseur, chasublier, chasublière, chaudronnier, chaudronnière, chauffagiste, chauffeur, chaufournier, chaumier, chausseur, chaussonnier, chef, cheminot, cheminote, chemisier, chemisière, chercheur, chercheuse, chevillard, chevilleur, chevrière, chevrier, chiffonnière, chiffonnier, chiffreur, chiffreuse, chimiste, chiropracteur, chiropraticien, chirurgien, chirurgienne, chocolatier, chocolatière, chorégraphe, choriste, chromeur, chromiste, chroniqueur, cicérone, cigarettière, cigarettier, cigarier, cigarière, cimentier, cineaste, cireur, cireuse, cirier, cirière, ciseleur, ciseleuse, cithariste, clapman, clarinettiste, claveciniste, claviériste, claviste, clerc, clicheur, clicheuse, climatologue, clinicienne, clinicien, cloutier, cloutière, clown, coache, coach, cocher, codeur, codeuse, codirecteur, codirectrice, coéditeur, coffreur, cogérante, cogérant, cogniticienne, coiffeur, coiffeuse, colleur, colleuse, coloriste, colporteur, coltineur, colzatier, comédien, comédienne, comique, commandant, commerçante, commerçant, commis, commise, commissaire, communarde, communard, communicateur, communicatrice, compagnon, compositeur, compositrice, comptable, compteuse, concepteur, conceptrice, concierge, conducteur, conductrice, confectionneur, conférencière, conférencier, confiseur, confiseuse, confiturière, confiturier, connecticienne, connecticien, conseiller, conseillère, conservateur, conservatrice, conserveur, conserveuse, consignataire, constable, consul, consule, consultant, consultante, contactologue, conteur, conteuse, contrebassiste, contremaître, contrôleur, contrôleuse, convoyeur, convoyeuse, coolie, coordinateur, coordinatrice, coordonnateur, coordonnatrice, copilote, copiste, coprésidente, coprésident, coqueleuse, coquetière, coquetier, corailleur, corailleuse, cordeur, cordeuse, cordier, cordière, cordonnier, cordonnière, cornac, cornemuseur, cornettiste, corneur, corneuse, corniste, coroner, correcteur, correctrice, correspondante, correspondant, corroyeur, corsaire, corsetier, corsetière, coryphée, cosmétologue, cosmologiste, cosmologue, cosmonaute, costumier, costumière, coteur, cotonnier, cotonnière, coupeur, coupeuse, courriériste, courrier, coursier, coursière, courtier, courtière, couseur, couseuse, coutelier, coutelière, couturier, couturière, couvreur, crassier, créancière, créancier, créateur, créatif, créatrice, crémier, crémière, créoliste, crêpier, crêpière, crieur, crieuse, criminologue, crinier, cristallière, crocheteur, croupier, croupière, cryptographe, cueilleur, cueilleuse, cuiseur, cuisinier, cuisinière, cuisiniste, culottière, culottier, cultivateur, cultivatrice, curateur, curatrice, cytologiste, dactylographe, dalleur, damasquineur, dameur, danseur, danseuse, débardeur, débardeuse, débatteur, débiteur, décapeur, décatisseur, décatisseuse, déchiqueteur, déchiqueteuse, décideur, décideuse, déclamateur, déclamatrice, décolleteur, décolleteuse, décorateur, décoratrice, découpeur, découpeuse, découvreur, découvreuse, décrotteur, défenseur, défibreur, défibreuse, défricheur, défricheuse, dégraisseur, dégraisseuse, dégustateur, dégustatrice, démarcheur, démarcheuse, déménageur, demandeuse, démineur, démographe, démolisseur, démonstrateur, démonstratrice, dendrologue, dentellière, dentellier, dentiste, déontologue, dépanneur, dépeceur, dépeceuse, dépollueur, dépoussiéreur, dermatologue, dermato, designer, dessinateur, dessinatrice, déstockeur, détacheur, detacheuse, détaillant, détaillante, detective,

développeur, développeuse, devideur, diablotin, diagnosticien, diagnostiqueur, dialectologue, dialoguiste, diamantaire, dictionnariste, diéséliste, diététicienne, diététicien, dinandier, dinandière, dindonnière, dindonnier, diplomate, directeur, directrice, discothécaire, discounter, discounteur, dispatcher, disquaire, distillateur, distillatrice, divette, docimologue, docker, docteur, doctoresse, documentaliste, documentariste, domestique, dominotière, dominotier, dompteur, dompteuse, donneur, doreur, doreuse, douanier, douanière, doubleuse, doublure, doucheur, doucheuse, dramaturge, drapier, drapière, draveur, draveuse, dresseur, dresseuse, droguiste, duègne, duumvir, dynamiteur, dynamiteuse, ébarbeur, ébaucheur, ébéniste, éboueur, éboueuse, écailleur, écailleuse, écangueur, écangueuse, échanson, échenilleur, échevin, échotier, échotière, éclairagiste, éclaireur, éclaireuse, éclateur, éclusier, éclusière, écogarde, écologue, économètre, économétricien, économiste, écorceur, écorcheur, écorcheuse, écrivain, écrivaine, écureur, écuyer, écuyère, édile, éditeur, éditorialiste, éditrice, éducateur, éducatrice, effileur, effilocheur, effilocheuse, égoutier, égratigneur, élagueur, élagueuse, électricien, électricienne, électronicien, éleveur, éleveuse, émailleur, émailleuse, emballeur, embaumeur, embaumeuse, embouteilleur, embouteilleuse, embryologiste, embryologue, émondeur, émondeuse, émotteur, émotteuse, émouleur, émouleuse, empailleur, empailleuse, empaqueteur, empaqueteuse, empileur, empileuse, employé, employée, encadreur, encadreuse, encaisseur, encaveur, encaveuse, encodeur, encodeuse, encolleur, encolleuse, endocrinologue, endosseur, enfileur, enfileuse, enfourneur, engraisseur, enlumineur, enlumineuse, enquêteur, enquêtrice, enrouleur, ensacheur, ensacheuse, enseignant, enseignante, ensemblier, entraineur, entraineuse, entrepreneur, envoyeur, épaviste, épicier, épicière, épigraphiste, épileur, épileuse, épinceteur, épinceteuse, épinceur, épinceuse, épinglière, épinglier, épistolier, éplucheur, éplucheuse, équilibriste, équipier, équipière, ergonome, ergothérapeute, espion, espionne, essayeur, essayeuse, essayiste, estafier, estampeur, estampeuse, esthéticienne, esthéticien, estimateur, étalagiste, étalière, étalier, étalonneur, étalonneuse, étameur, étameuse, étampeur, étampeuse, ethnobotaniste, ethnographe, ethnologue, éthologue, étiqueteur, étiqueteuse, étireur, étireuse, expéditeur, expéditrice, expert, experte, fabuliste, façadier, facilitateur, facilitatrice, façonneur, façonnier, façonnière, facteur, factrice, fagoteur, fagoteuse, fagotier, fagotière, faïencier, faïencière, faneur, faneuse, farinier, farinière, faucheur, faucheuse, fauconnier, fendeur, fendeuse, ferblantier, fermier, fermière, ferrailleur, ferrailleuse, ferreur, ferronnier, feudiste, feuillagiste, ficelier, figurant, figurante, figuriniste, figuriste, filandier, filandière, filateur, fildeferiste, fileur, fileuse, filmeur, filmeuse, financier, finisseur, finisseuse, fiscaliste, fleuriste, floriculteur, floricultrice, flotteur, flutiste, fonctionnaire, fondeur, fondeuse, fontainier, footballeur, forain, foraine, forceur, forestier, foreur, foreuse, forfaitiste, forgeron, formateur, formatrice, fossoyeur, fossoyeuse, fouacier, foudrier, fouleur, foulon, foulonnier, fourbisseur, fournier, fourreur, fourreuse, fraiseur, fraiseuse, frappeur, frappeuse, freineur, fresquiste, freteur, frigoriste, fripier, fripière, friteuse, friturier, fromager, fromagère, fructiculteur, fruitière, fruitier, fumiste, funambule, gabarier, gabarrier, gabier, gâcheur, gâcheuse, gagiste, gainier, gainière, galeniste, galeriste, galibot, galonnier, galonnière, gambiste, gantier, gantière,

garagiste, garanceur, garanceuse, garde, gardeur, gardeuse, gardien, gardienne, gargotière, gargotier, garnisseur, garnisseuse, gaucho, gaufreur, gaufreuse, gaveur, gaveuse, gazetier, gazetière, gazier, gazière, geisha, gemmeur, gemmeuse, gemmologue, gendarme, généalogiste, généraliste, généticienne, généticien, genévrier, géodésienne, géodésien, géographe, geôlier, geôlière, géologue, géomètre, géophysicienne, géophysicien, gériatre, gérontologue, gestionnaire, giletier, giletière, gitologue, glaceur, glaceuse, glacier, glacière, glaciologue, glaneuse, gobeletière, gobeletier, goémonier, goémonière, golfeur, golfeuse, gondolier, gondolière, gonfalonier, gonfanonier, goudronneur, goudronneuse, goudronnier, gouteur, gouteuse, gouvernante, grainetier, grainetière, graineur, grainier, grainière, graisseur, grammairienne, grammairien, grammatiste, granitier, graphiste, graphologue, gratteur, gratteuse, gravatier, graveur, graveuse, gravier, gréeur, greffeur, greffeuse, greffier, greffière, greneur, grenier, grignoteur, grillageur, groom, grossiste, groupeur, groupeuse, grutier, grutière, guetteur, guichetière, guichetier, guide, guillotineur, guitariste, gymnaste, gynécologue, gypsier, habilleur, habilleuse, hagiographe, haleur, haleuse, handicapeur, harengère, harmoniciste, harmoniste, harnacheur, harpiste, haubergeon, haveur, heaumier, héliciculteur, hélicicultrice, hélicier, héliograveur, héliograveuse, hépatologue, héraut, herbager, herboriste, hercheur, herscheur, hippologue, historien, historienne, hockeyeur, hockeyeuse, hongreur, hongroyeur, horloger, horlogère, horticulteur, horticultrice, hortillon, hote, hotelier, hotelière, hotesse, hotteur, hotteuse, houppier, huchet, huchier, huilier, huissier, huissière, humoriste, hybrideur, hybrideuse, hypnologue, hypnotiseur, hypnotiseuse, ichtyologiste, ichtyologue, iconographe, iconologiste, iconologue, illusionniste, illustrateur, illustratrice, ilotier, ilotière, imageur, imagier, imagière, imitateur, imitatrice, imposeur, impresario, imprimeur, imprimeuse, incrusteur, incrusteuse, indexeur, indexeuse, indianiste, industrielle, industriel, infectiologue, infirmier, infirmière, infographiste, informaticien, ingénieriste, ingénieur, inséminateur, inspecteur, inspectrice, installateur, installatrice, instituteur, institutrice, instructeur, instructrice, instrumentiste, intendante, intendant, intermittent, interprète, introducteur, ivoirier, ivoirière, jardinier, jardinière, jardiniste, jaugeur, joaillière, joaillier, jockey, jongleur, jongleuse, journalier, journaliste, juge, junior, jupier, jupière, jurat, juré, juriste, kinesiste, kinésithérapeute, kiosquière, kiosquier, laborantin, laborantine, laboureur, laceur, laceuse, lâcheur, lad, laineur, laineuse, lainier, lainière, laitier, laitière, lamaneur, lamier, lamineur, lamineuse, lampier, lampiste, lancier, lanternier, lapicide, lapidaire, laquais, laqueur, laqueuse, latiniste, lavandière, laveur, laveuse, layetier, layeur, lecteur, lectrice, legionnaire, légiste, legumier, lessiveuse, lessivier, lettreur, lettreuse, leveur, leveuse, levurie, levurier, lexicographe, lexicologue, libraire, librettiste, licier, liégeur, liégeuse, lieur, lieuse, liftier, liftière, limeur, limeuse, limonadier, limonadière, linger, lingère, linguiste, linier, linière, linotypiste, liquoriste, liseur, liseuse, lisseur, lisseuse, lissier, lithographe, livreur, livreuse, logisticien, logisticienne, logographe, logopède, logothète, lotisseur, loueur, loueuse, louvetier, ludologue, ludothécaire, lunetier, lunetière, lunettier, lunettière, lustreur, lustreuse, luthier, luthière, luthiste, lutteur, lutteuse, machiniste, maçon, maçonne, magasinier, magasinière, magicien, magicienne, magistrat, magistrate,

magnanarelle, magnanière, magnanier, magnarelle, magnétiseur, magnétiseuse, maieure, maieur, maieuticien, maître, maîtresse, malletière, malletier, malteur, maltotier, mammalogiste, manadier, manager, manageur, manageuse, manieur, manipulateur, manipulatrice, mannequin, manoeuvre, manoeuvrière, manouvrier, manucure, manufacturier, manufacturière, maquettiste, maquignon, maquilleur, maquilleuse, maraicher, maraichère, marathonienne, marathonien, marbreur, marbreuse, marbrier, marcaire, marchand, marchande, marchandiseur, marcheur, mareyeur, mareyeuse, margarinier, margeur, margeuse, margoulin, marieur, marieuse, marin, marinier, marinière, marionnettiste, marketeur, marketeuse, marneur, maroquinier, maroquinière, marqueteur, marqueur, marqueuse, marteleur, masseur, masseuse, massier, mastologue, matelassier, matelassière, matelot, mateur, mathématicien, matrone, mayeure, mayeur, mécanicien, mécanicienne, mécanographe, médailleur, médecin, médiateur, médiatrice, mégissier, mélangeur, mélangeuse, mélodiste, menager, menagère, ménagiste, ménestrel, ménétrier, menuisier, menuisière, mercaticien, mercaticienne, merchandiser, mercier, mercière, messager, mesureur, mesureuse, metallier, metallière, metallographe, metallurgiste, métayère, métayer, météorologiste, météorologue, métreur, métreuse, métrologiste, métrologue, meulier, meunier, meunière, militaire, mime, minerviste, mineur, miniaturiste, ministre, minotier, minotière, mireur, mireuse, miroitière, miroitier, missilier, mitron, mixeur, modèle, modeleur, modeleuse, modélisateur, modélisatrice, modéliste, modiste, moireur, moireuse, moissonneur, moissonneuse, moniteur, monitrice, monnayeur, monteur, monteuse, mortaiseur, morutier, morutière, mosaïste, motociste, motoriste, mouleur, mouleuse, moulineur, moulineuse, moulinier, moulinière, mouliste, moulurier, moulurière, mousse, moutardier, moutonnier, muezzin, mulassier, muletier, muletière, muséographe, musicien, musicienne, musicologue, mycologue, myrmecologue, mytiliculteur, mytilicultrice, nageur, nageuse, naisseur, nattier, nattière, naturaliste, nautonier, nautonière, navigateur, navigatrice, négociante, négociant, négociateur, neoniste, néphrologue, netsurfeur, netsurfeuse, nettoyeur, nettoyeuse, neurologue, nez, nielleur, nielleuse, nivologue, nocher, notaire, notateur, notatrice, noueur, nourrice, nourrisseur, novice, noyauteur, nuiteuse, nurse, nutritionniste, obstétricienne, obstétricien, océanaute, oculariste, oculiste, odontologiste, oenologue, officier, officière, oiseleur, oiseleuse, oiselier, oiselière, oleiculteur, oleicultrice, oliveur, oliveuse, ombudsman, omnipraticien, oncologiste, oncologue, onirologue, opérateur, opératrice, ophtalmologue, opticien, opticienne, optométriste, orchestrateur, orchestratrice, ordinaticienne, ordinaticien, orfèvre, organicien, organicienne, organier, organière, organisateur, organisatrice, organiste, orienteur, orienteuse, ornemaniste, ornithologue, orpailleur, orpailleuse, orthodontiste, orthopédiste, orthophoniste, orthoptiste, ostéopathe, ostéopraticien, ostréiculteur, ostréicultrice, otorhino, outilleur, outplacer, outplaceur, outplaceuse, ouvreur, ouvreuse, ouvrier, ouvrière, oxycoupeur, packager, packageur, pailler, pailleteur, pailleteuse, pailleur, pailleuse, palefrenière, palefrenier, paléographe, paléologue, paléontologue, palettiseur, palissonneur, paloteur, paludier, paludière, paludologue, palynologue, panetier, panseur, panseuse, pantomime, pantouflière, pantouflier, papetier, papetière, papillonneur, parachutiste,

parasitologue, pareur, pareuse, parfumeur, parfumeuse, parolier, parolière, parqueteur, parqueur, parqueuse, parquier, parurier, parurière, passementier, passementière, passeur, pastelliste, pasteur, pasteure, pasteurisateur, pastilleur, pastilleuse, pastourelle, patineur, patineuse, pâtissier, pâtissière, pâtre, patron, patronne, patronnier, patronnière, pattière, pattier, paveur, paveuse, paysagiste, paysan, paysanne, péager, péagère, péagiste, peaussier, pêcheur, pêcheuse, pédiatre, pédicure, pédologue, pédopsychiatre, peigneur, peigneuse, peignier, peintre, pelletier, pelletière, pendulier, peon, pepinieriste, percepteur, perceptrice, perceur, perceuse, perchiste, perchman, perforateur, perforatrice, perliculteur, perlicultrice, perlier, perlière, perruquière, perruquier, personnage, peseur, pétrisseur, pétrisseuse, pétrographe, pétrolier, pétrolière, pharmacienne, pharmacien, philologue, phlébologue, phoniatre, photographe, photograveur, physicienne, physicien, physiologiste, physionomiste, phytogeographe, pianiste, picador, pigiste, pileur, pileuse, pilote, pipier, pipière, piqueur, piqueuse, piroguier, pisciculteur, pisteur, pisteuse, pistoleur, pizzaiolo, placeur, placeuse, placier, placière, plagiste, planétologue, planeur, planteur, planteuse, plaquiste, plasticien, plasticienne, plasturgiste, plâtrier, plieur, plieuse, plisseur, plisseuse, plombeur, plombier, plongeur, plongeuse, plumassière, plumassier, plumeur, plumeuse, pneumologue, pocheuse, podiatre, podologue, poêlier, poêlière, poète, poétesse, poinçonneur, poinçonneuse, pointeau, pointeur, pointeuse, poissonnier, poissonnière, policeman, policier, policière, polisseur, polisseuse, pompeur, pompier, pompière, pompiste, ponceur, ponceuse, pontier, pontière, pontonnier, pope, populiculteur, populicultrice, porcelainier, porcelainière, porchère, porcher, porion, portefaix, porteur, porteuse, portier, portière, portraitiste, poseur, poseuse, posticheur, posticheuse, postier, postière, potier, potière, praticien, praticienne, précepteur, préceptrice, préfet, préfète, préparateur, préparatrice, préposé, préposée, présentateur, présentatrice, président, présidente, presseur, presseuse, pressier, prêteur, prêtre, prévisionniste, prévôt, primeur, primeuriste, principal, principale, priseur, priseuse, privée, privé, procureur, procureure, producteur, productrice, professeur, professeure, profileur, programmateur, programmatrice, programmeur, programmeuse, projeteur, projeteuse, promoteur, promotrice, prospecteur, prospectrice, prote, prothésiste, prototypiste, proviseur, psy, psychanalyste, psychiatre, psychologue, publicitaire, puériculteur, puéricultrice, puisatier, pupitreur, pupitreuse, pyrotechnicien, pyrotechnicienne, qualiticien, qualiticienne, questeur, quincaillier, quincaillière, rabatteur, rabatteuse, rabbin, raboteur, raboteuse, raccommodeur, raccommodeuse, racleur, radariste, radio, radioastronome, radiologue, radionavigant, raffineur, raffineuse, ramasseur, ramasseuse, ramendeur, ramendeuse, rameur, ramoneur, rapsode, raseur, raseuse, rateleur, rateleuse, raucheur, ravaleur, ravaudeur, ravaudeuse, ravitailleur, rayonneur, réalisateur, réalisatrice, réassureur, rebouteuse, recenseur, recenseuse, réceptionnaire, réceptionniste, receveur, receveuse, recherchiste, récolteur, récolteuse, recors, recruteur, recruteuse, recteur, rectifieur, rectifieuse, rectrice, récupérateur, récupératrice, rédacteur, rédactrice, redresseur, redresseuse, régent, régente, régisseur, régisseuse, registraire, régleur, régleuse, regrattière, regrattier, régulateur, régulatrice, rejointoyeur, rejointoyeuse, relationniste, relecteur, relectrice, releveur, relieur, relieuse, remailleur,

remailleuse, remisier, remisière, remmailleur, remmailleuse, remonteur, remouleur, remouleuse, rempailleur, rempailleuse, remplisseur, remplisseuse, rentoileur, rentoileuse, rentrayeur, rentrayeuse, renvideur, renvideuse, réparateur, réparatrice, repasseur, repasseuse, répétiteur, répétitrice, reporter, reporteur, reportrice, représentante, représentant, repriseuse, réserviste, résinière, résinier, responsable, restaurateur, restauratrice, rétameur, retordeur, retoucheur, retoucheuse, réviseur, réviseuse, rhabilleur, rhabilleuse, rhéteur, rhumatologue, rinceuse, riveur, riveuse, riziculteur, rizicultrice, robeuse, robinetier, roboticien, roboticienne, rocailleur, rôdeur, rogneur, rogneuse, romancier, romancière, romaniste, rosiériste, rôtisseur, rôtisseuse, rouennier, rouennière, rouleur, roulier, routeur, routeuse, routier, routière, rubanier, rubanière, rugbyman, sableur, sableuse, sabotier, sabotière, sabreur, sacristain, sagard, saigneur, saigneuse, saleur, saleuse, saliculteur, salicultrice, salinier, salinière, salmoniculteur, salonnier, salonnière, salpetrière, salpetrier, sandalier, sandalière, santonnier, santonnière, sapeur, sapiteur, sardinier, sardinière, sarodiste, sasseur, sasseuse, satineur, satineuse, saucier, saucière, saunier, saunière, saurisseur, saurisseuse, sauveteur, savetier, savetière, savonnier, savonnière, saxophoniste, sbire, scaphandrier, scénariste, scénographe, scieur, scieuse, scoliaste, scribe, scripte, sculpteur, secouriste, secrétaire, sellier, semeur, semeuse, séranceur, séranceuse, serriste, serrurier, serrurière, sertisseur, servant, servante, serveur, serveuse, sexologue, shampouineur, shampouineuse, sherif, sherpa, signaleur, sismologue, sitologue, sociologue, soignante, soignant, soigneur, soigneuse, solier, soliste, sommelier, sommelière, sondeur, sondeuse, sonneur, soudeur, soudeuse, souffleur, soufreur, soufreuse, souscripteur, souscriptrice, soutier, spadassin, spationaute, speakerine, speaker, spéléo, spéléologue, spéléonaute, stadier, stadière, staffeur, standardiste, stationnaire, statisticienne, statisticien, stéarinier, stenciliste, steno, sténodactylo, sténographe, sténotypiste, steward, stomatologiste, stomatologue, storiste, stripteaseur, stripteaseuse, stucateur, stylicien, stylicienne, styliste, sujet, superviseur, surveillant, surveillante, sylviculteur, sylvicultrice, tabaculteur, tabacultrice, tabletier, tabletière, tacheron, taillandier, tailleur, tailleuse, talonneur, tamiseur, tamisier, tamisière, tanneur, tanneuse, tapisseur, tapisseuse, tapissier, tapissière, taraudeur, taraudeuse, tatoueur, tatoueuse, taulier, taulière, taupier, taupière, tavernier, tavernière, taxateur, taxatrice, taxidermiste, taximan, taxiste, technicien, technicienne, teilleur, teilleuse, teinturière, teinturier, téléacteur, téléactrice, téléaste, téléconseiller, télégraphiste, télémétreur, télémétreuse, téléopérateur, téléphoniste, téléreporter, télévendeur, télévendeuse, téléxiste, tenancier, tenancière, tendeur, tendeuse, tennisman, tenniswoman, terminologue, terrassier, testeur, testeuse, thanatologue, théâtreuse, thermicienne, thermicien, thonier, tilleur, tilleuse, timbreur, timbreuse, timonier, tireur, tisserand, tisserande, tisseur, tisseuse, titreur, titreuse, toiletteur, toiletteuse, toilier, toilière, tolier, tolière, tombeur, tondeur, tondeuse, tonnelier, topographe, tordeur, tordeuse, toreador, torera, torero, torréfacteur, toucheur, toucheuse, toueur, toueuse, toupilleur, tourbière, tourbier, tourier, tourneur, tourneuse, traceur, traceuse, tractoriste, trader, traducteur, traductrice, tragédienne, tragédien, traiteur, trameur, trameuse, traminot, trancheur, trancheuse, transformiste, transitaire, transporteur, transporteuse, trapéziste, trappeur, trappeuse,

traqueur, traqueuse, trayeur, trayeuse, tréfileur, tréfileuse, trempeur, trésorier, trésorière, tresseur, tresseuse, tribun, tricoteur, tricoteuse, trieur, trieuse, tripier, tripière, trompettiste, tronçonneur, trousseur, trufficulteur, trufficultrice, truquiste, tubiste, tuilier, tulliste, tuteur, tutrice, tuyauteur, tuyauteuse, typographe, ubiquiste, ufologue, urbaniste, urgentiste, urgentologue, urologue, usineur, usineuse, usinier, usurière, usurier, vacataire, vacher, vachère, vaguemestre, vaisselier, valet, vanneur, vanneuse, vannier, vannière, veilleur, veloutier, vendangeur, vendangeuse, vendeur, vendeuse, veneur, ventriloque, vépéciste, verdier, verdurière, verdurier, vergetier, vérificateur, vérificatrice, vérifieur, vérifieuse, vernisseur, vernisseuse, verrier, verrière, vétérinaire, vibraphoniste, vicaire, vidangeur, vidangeuse, vidéaste, videur, videuse, vigie, vigile, vigil, vigneron, vigneronne, viguier, vinaigrier, vinificateur, vinificatrice, violiste, violoncelliste, violoniste, virolière, virolier, virologue, visagiste, visitateur, visitatrice, visiteuse, viticulteur, viticultrice, vitrier, vitrière, vivandière, vivandier, voilier, voiturier, voiturière, voiturin, volailler, volailleur, volcanologue, voltigeur, voltigeuse, voyagiste, voyer, vulcanologue, wagonnier, wattman, webdesigner, webmarketeur, webmarketeuse, webmaster, webmestre, webplaner, zingueur, zoologiste, zoologue.

B.2. Abstract Concepts

prier prier, adorer, conjurer, louer, supplier, implorer, solliciter, revendiquer, enjoindre **culte, religion** culte, admiration, adoration, adulation, église, amour, célébration, cérémonie, catholicisme, confession, croyance, déification, dévotion, dévouement, fétichisme, foi, hommage, idolâtrie, liturgie, messe, mystère, office, pèlerinage, passion, piété, pratique, prière, révérence, religion, respect, rite, théogonie.

B.3. Places

habitation village, agglomération, bourg, bourgade, feux, garnison, hameau, bastide, cité, citadelle, habitant, habitante, locataire, métropole, peuple, peuplement, commerce, cabaret, cambuse, guinguette, hôtel, hôtellerie, palace, pension, restau, restaurant, pharmacie, drugstore, officine, boutique, affaire, atelier, attirail, bahut, bande, banneton, bazar, bric-à-brac, camarilla, chapelle, clan, clique, crémerie, débit, maffia, magasin, outillage, réserve, vitrine, vivier, mairie, capitole, capitoul, commune, municipalité, douane, laboratoire, arrière-boutique, cabinet, importation, exportation, académie, école, assemblée, athénée, campus, centre, collège, constitution, cours, département, faculté, laboratoire, lycée, organisme, préfecture, séminaire, société, abri, édifice, église, appartement, artisanal, asile, établissement, baraque, bas-lieu, bâtiment, bâtisse, bercail, bicoque, boîte, bouge, branche, building, bungalow, cabane, cahute, campagne, cassine, chacunière, chalet, château, chaumière, chez-soi, clapier, clinique, construction, couronne, couvert, demeure, descendance, domesticité, domestique, domicile, dynastie, entreprise, ferme, feu, firme, foyer, gîte, gens, habitacle, hôpital, home, hutte, immeuble, institut, institution, intérieur, lares, lieu, lignée, loge, logement, logis, maison, maisonnée, maisonnette, ménage, manoir, manufacture, masure, monde, naissance, nid, nom, origine, palais, pénates, parents, pavillon, pigeonnier, place, prison, propriété, race, réduit, résidence, retraite, séjour, sanctuaire, serviteur,

taudis, temple, trône **chambre** alcôve, alvéole, antichambre, assemblée, association, cabine, cagibi, cambriole, carrée, case, cavité, cellule, chambre, chambrée, chambrette, compartiment, conseil, corps, cour, crèche, creux, dortoir, galetas, gourbi, habitation, local, mansarde, nursery, parlement, pièce, piaule, poêle, reposée, salle, studette, studio, taule, tourelle, tribunal, turne, salon, bal, boudoir, bringue, cercle, club, entourage, exposition, fiesta, hall, living, living-room, musée, nouba, raout, redoute, sauterie **chemin** rue, allée, artère, asphalte, avenue, boulevard, chaussée, chemin, cheminée, galerie, impasse, loggia, passage, pavé, ruelle, accès, barrage, billet, bouffée, boyau, brèche, bribe, canal, chenal, circulation, col, conduit, conjoncture, corridor, couloir, défilé, dégagement, déplacement, détroit, endroit, extrait, flânerie, foulée, fragment, franchissement, fuite, gorge, goulet, gué, issue, moment, morceau, ouverture, péage, page, pèlerinage, partie, pas, passée, passe, passerelle, pâturage, percée, piste, port, randonnée, route, sédentarisation, saut, sentier, seuil, sillage, stage, strophe, suture, trafic, transit, transition, traversée, traverse, trouée, tunnel, vadrouille, va-et-vient, venelle, venue, vestibule, voie, pont, appontement, aqueduc, bau, bordé, bordage, dunette, embelle, gaillard, passavant, plate-forme, pontil, quai, spardeck, superstructure, wharf **nature** forêt, affluence, arbre, bois, breuil, feuillage, frondaison, futaie, jungle, labyrinthe, perçoir, perchis, sylve, taillis, taraud, montagne, accident, accumulation, élévation, alpage, amas, éminence, amoncellement, assemblage, chaîne, chaînon, cime, colline, contrefort, djebel, entassement, fatras, fouillis, hauteur, haut, mamelon, monceau, mont, monticule, pic, pile, piton, sierra, sommet, tas, mer, étendue, baille, déluge, espace, essaim, flot, flots, fourmillement, large, marine, océan, onde, plaine, reflux, rivage, île, îlot, archipel, atoll, havre, oasis, affluent, bain, collier, cours, eau, fleuve, flux, gave, oued, ravine, ruisseau, torrent, marais, paysage, bergerie, bucolique, campagne, carte, décor, horizon, localisation, panorama, sous-bois, verdure **localisation** à, bord, ailleurs, auprès, autour, abords, adjacent,

touchant, limitrophe, contre, collé, avoisinant, côtés, côté, proximité, approchant, rapproché, attenant, imminent, jouxte, loin, près, presque, proche, sur, voisin, riverain, depuis, dès, par, pour, dedans, dehors, en, hors, proche, éloigné, distant, adresse, but, destination, sens, trimard, nord, arctique, boréal, borée, hyperboréen, nordique, polaire, septentrion, sud, antarctique, austral, méridional, est, aurore, orient, ouest, occident, occidental **actions** aborder, aboutir, échoir, apparaître, approcher, arriver, atteindre, avancer, débarquer, dériver, descendre, entrer, marcher, monter, partir, parvenir, rappliquer, diriger, remonter, redescendre, abouler, déplacer, rapprocher, transporter, suivre, surgir, survenir, venir, abandonner, échapper, éclipser, aller, éloigner, émigrer, appareiller, bouger, circuler, déménager, embarquer, enfuir, fuir, nager, promener, provenir, quitter, repartir, tourner, trotter, vider, voyager, traverser, pénétrer, sillonner, transir, retourner, habiter, louer, domicilié, domicilier, cabaner, camper, crécher, demeurer, exister, fréquenter, gîter, hanter, loger, nicher, obséder, occuper, percher, peupler, posséder, poursuivre, préoccuper, régner, résider, séjourner, rencontrer, siéger, travailler, visiter, vivre, rapatriement, exfiltration, réinsertion **transports** paquebot, barque, bateau, batelet, caboteur, canoë, caraque, caravelle, cargo, chalutier, dériveur, embarcation, felouque, jonque, navire, paquebot, péniche, périssoire, pirogue, radeau, rafiot, steamer, steamboat, skiff, submersible, transatlantique, trirème, vaisseau, vapeur, voilier, aérodyne, aéronef, aéroplane, aéroscaphe, airbus, appareil, avion-taxi, bac, aérien, bombardier, canadair, charter, chasseur, avion, giravion, hydravion, jet, tacot, taxi, transport, U.L.M., voiture, attelage, auto, berline, cabriolet, carriole, charrette, fardier, fourgon, train, véhicule, locomotive, automotrice, coucou, locomotrice, machine, motrice, bagnole, chignole, coche, fiacre, limousine, roadster, tire, tramway, vélo, deux-roues, bécane, bicyclette, moto, cyclomoteur, enduro, engin, motocycle, motocyclisme, scooter, solex, motocyclette, vélomoteur, embarquement.