Le benchmarking de la reconnaissance d'entités nommées pour le français

Jungyeul Park
CONJECTO. 74 rue de Paris, 35000 Rennes, France
http://www.conjecto.com

RÉSUMÉ

Cet article présente une tâche du benchmarking de la reconnaissance de l'entité nommée (REN) pour le français. Nous entrainons et évaluons plusieurs algorithmes d'étiquetage de séquence, et nous améliorons les résultats de REN avec une approche fondée sur l'utilisation de l'apprentissage semi-supervisé et du reclassement. Nous obtenons jusqu'à 77.95%, améliorant ainsi le résultat de plus de 34 points par rapport du résultat de base du modèle.

ABSTRACT

Benchmarking for French NER.

This paper presents a benchmarking task of named-entity recognition for French. We train and evaluate several sequence labeling algorithms, and we improve named-entity recognition results using semi-supervised learning and reranking. We obtain up to 77.95%, in which we improve the result by over 34 points compared to the baseline results.

MOTS-CLÉS : Reconnaissance d'entités nommées, REN, benchmarking, évaluation, français.

KEYWORDS: Named-entity recognition, NER, benchmarking, evaluation, French.

1 Named Entity Recognition

Named entities are phrases that contain the names of persons, organizations and locations (Tjong Kim Sang & De Meulder, 2003). The task of named-entity recognition (NER) seeks to identify elements into predefined categories such as the names of persons (PER), locations (LOC), organizations (ORG), etc. The following example[1] is from CoNLL 2003 English NER data :

U.N.	NNP	I-ORG
official	NN	O
Ekeus	NNP	I-PER
heads	VBZ	O
for	IN	O
Baghdad	NNP	I-LOC
.	.	O

In this example, entities such as PER, LOC and ORG are tagged using the BIO format alongside their

[1] The example excerpted from https://www.clips.uantwerpen.be/conll2003/ner

original :	preprocessed :
`Emmanuel I-PER`	`Emmanuel NAM I-PER`
`DESOLES I-PER`	`DESOLES NAM I-PER`
`de O`	`de PRP O`
`LOU O`	`LOU NAM O`
`Directeur O`	`Directeur NAM O`
`politique O`	`politique ADJ O`
`BÊ>ÀCTION O`	`BÊ>ÀCTION NAM O`
`ET O`	`ET NAM O`
`ADMINISTRATION O`	`ADMINISTRATION NAM O`
`9& O`	`9& ADJ O`
`, O`	`, PUN O`
`Rue I-LOC`	`Rue NOM I-LOC`
`du I-LOC`	`du PRP:det I-LOC`
`Pré-Botté I-LOC`	`Pré-Botté NAM I-LOC`
`, O`	`, PUN O`
`aS O`	`aS VER:simp O`
`RENNES I-LOC`	`RENNES NAM I-LOC`
`ABONNEMENTS O`	`ABONNEMENTS NAM O`
`Dép O`	`Dép NAM O`
`. O`	`. SENT O`

FIGURE 1 – Original and preprocessed NER data for French

words and Penn tagset part-of-speech (POS) labels. B-I-O stands for beginning-inside-outside of each entity.

This paper presents a benchmarking task for French NER. We train and evaluate several sequence labeling algorithms such as a Hidden Markov model (HMM) (Rabiner, 1989), conditional random fields (CRF) (Lafferty *et al.*, 2001), and bi-directional long-short-term-memory recurrent neural network (bi-LSTM RNN) (Graves & Schmidhuber, 2005) for French NER. We also improve NER results by introducing semi-supervised learning in which we use a large monolingual corpus to augment the training data, and reranking which adjusts the results based on several sequence labeling algorithms.

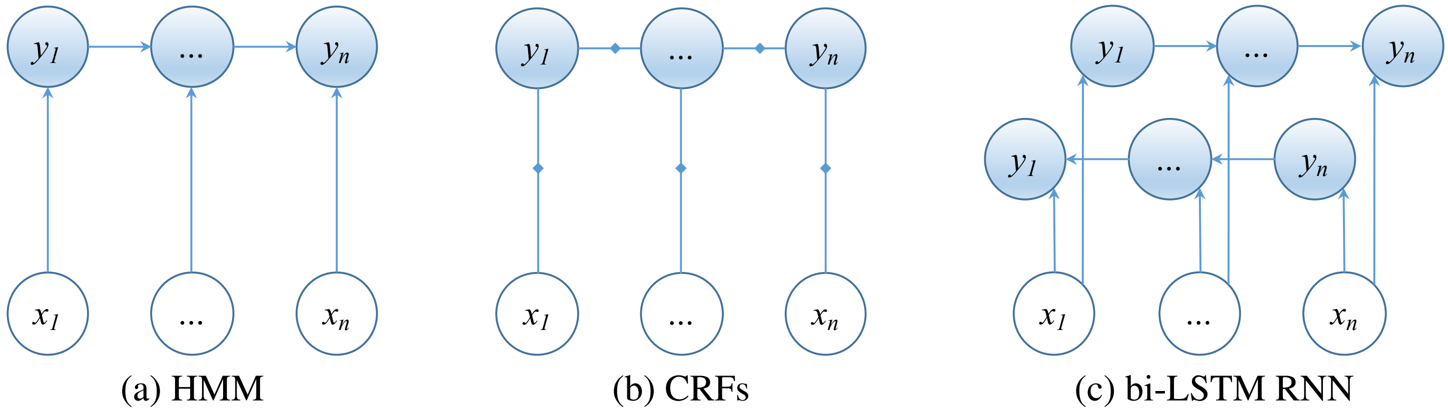
2 Experiments and Results

2.1 Data

We use the French NER data provided by Europeana Newspapers[2]. They are OCRed newspaper from 1870 to 1939 taken from the National Library of France. The original data only provides automatically tokenized text and named entity label for each token. There are no sentence boundaries. For training and evaluation, we add "rough" sentence boundaries and POS labels by `TreeTagger` (Schmid, 1994)[3]. To the best of author's knowledge, there are no previous results on this corpus. We explicitly introduce sentence boundaries that machine learning algorithms are trained sentence by sentence based on the `TreeTagger` sentence segmentation. We then split the corpus 80/10/10 ratio as training/development/test data sets, and it gives 10,041/1,255/1,255 sentences, respectively. Figure 1 shows the original data and preprocessed NER data for French. Note that the present corpus is “original”. If there may be errors, it is not corrected in this paper.

[2] Available at `https://github.com/EuropeanaNewspapers/ner-corpora`

[3] Available at `http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/`



(a) HMM (b) CRFs (c) bi-LSTM RNN

FIGURE 2 – Learning models for NER : figures for HMM and CRFs are inspired by Sutton & McCallum (2012).

2.2 Learning models

We use following learning models to train and evaluate NER for French :

- HMM using `TnT` (Brants, 2000) [4]
- CRFs using `CRF++` [5]
- CRFs using `Wapiti` (Lavergne *et al.*, 2010) [6]
- bi-LSTM RNN using `NeuroNER` (Dernoncourt *et al.*, 2017) [7] with a pre-trained embedding vector for French (Bojanowski *et al.*, 2017) [8].

Figure 2 summarizes the learning models of HMM, CRFs and bi-LSTM RNN where x_i is a word and y_i is a label $(1 \leq i \leq n)$. While an HMM uses only the token's observation probability and the transition probability of states (label) for learning features, CRFs can use their features as we define. We use ±2 word and POS window context information and a bi-gram word and POS model are used as a feature set for CRFs. The neural network will learn the optimal features during training for the bi-LSTN RNN. We run the experiment with 50 epochs with stochastic gradient descent (SGD), 0.005 learning rate, and 0.5 dropout rate . A pre-trained embedding vector for French (Bojanowski *et al.*, 2017) is in 300 dimensional space, and it enriches word vector results with subword information.

2.3 Results

We evaluate NER results with the standard F_1 metric using `conlleval` [9]. Table 1 shows the overall baseline results on NER for French using several sequence labeling algorithms. Note that we use train+dev data for training for `TnT` and `CRF++` because they cannot have development data during training. Training without dev data can obtain 45.36% and 63.16% for `TnT` and `CRF++`, respectively, which are being outperformed by training with train+dev data as we present in Table 1. Otherwise, we use train/dev/evaluation data as described in §2.1. crf (w) can improve up to 65.95% if the L2 penalty parameter for ridge regression is set λ to 0.01, which can penalize the high-value weights to

[4] Available at `http://www.coli.uni-saarland.de/~thorsten/tnt/`
[5] Available at `https://taku910.github.io/crfpp/`
[6] Available at `https://wapiti.limsi.fr/`
[7] Available at `http://neuroner.com/`
[8] Available at `https://fasttext.cc/docs/en/pretrained-vectors.html`
[9] Available at `https://www.clips.uantwerpen.be/conll2003/ner/`

	hmm (t)	crf (+)	crf (w)	bi-lstm
precision	38.99	58.49	60.13	73.71
recall	55.37	72.06	73.01	78.99
F_1	45.76	64.57	65.38	76.26

TABLE 1 – Overall baseline results on NER for French : crf (+) and crf (w) represent CRFs using CRF++ and Wapiti, respectively.

avoid overfitting. We also note that results on CRFs can be improved using the different feature set. Even though CRF++ and Wapiti implement the same algorithm, Wapiti gives the better results. We assume that this is because stop criteria of implementations and default values that we use for learning. [10] While bi-LSTM RNN improves up to 77.76% during training epochs, we present the best result based on dev data.

3 Improving NER Models Using Semi-supervised Learning

We employ the NER model described in the previous section (§2.3) to improve NER results using semi-supervised learning, in which we automatically annotate a large monolingual corpus. This kind of practice is often called self-training (McClosky *et al.*, 2006a), self-taught learning (Raina *et al.*, 2007), and lightly-supervised training (Schwenk, 2008). For semi-supervised learning we introduce the consensus method $\hat{\mathcal{D}}$ (Brodley & Friedl, 1999). We use it by intersection between entity-annotated results using

$$\hat{\mathcal{D}} = \mathcal{D}(\mathcal{M}_1) \cap \cdots \cap \mathcal{D}(\mathcal{M}_n) \tag{1}$$

where $\mathcal{D}$ is raw text data, $\mathcal{M}_i$ is a learning model to annotate raw text data ($1 \leq i \leq n$), and $\hat{\mathcal{D}}$ is filtered annotated data. For raw text data for French, we use the monolingual corpus from the French treebank (Abeillé *et al.*, 2003) [11] (sentences only), and the French News Commentary v10 corpus [12]. We directly use morphologically segmented tokens in the treebank, and the preprocessing tools of Moses (Koehn *et al.*, 2007) for the new commentary corpus : normalizing punctuations and tokenization. [13] Table 2 summarizes the size of the monolingual corpus. To present the characteristics of the monolingual corpus, we provide the ratio of entity labels comparing to per in $\hat{\mathcal{D}}$, in which per is the most frequent entity in the original corpus. For example, the original NER training data set (train) contains 4,977 per and 4,432 loc entities, in which we represent 0.89 for loc. Note that the number and the ratio of entities in the French treebank and the New Commentary corpora are based on the automatically labeled entities ($\hat{\mathcal{D}}$).

Table 3 shows the overall results on NER using semi-supervised learning. Since hmm (t) gives the weakest results in the previous section, we exclude it for data intersection. Therefore, we obtain $\hat{\mathcal{D}}$ only from $\mathcal{D}(\mathcal{M}_{crf(+)}) \cap \mathcal{D}(\mathcal{M}_{crf(w)}) \cap \mathcal{D}(\mathcal{M}_{bilstm})$ for the current semi-supervised learning task. All learning algorithms can improve the NER results using semi-supervised learning by benefiting from the larger training data, even though they are automatically labeled. Such improvements using "self-

[10] We would like to thank reviewer #3 for indicating this problem.

[11] Available at http://www.llf.cnrs.fr/Gens/Abeille/French-Treebank-fr.php

[12] Available at http://www.statmt.org/wmt15/training-parallel-nc-v10.tgz

[13] Available at http://www.statmt.org/moses

	size ($\mathcal{D}$)	size ($\hat{\mathcal{D}}$)	`per`	`loc`	`org`
(original train)	0.16 M	-	1	0.89	0.42
French treebank	0.62 M	0.38 M	1	0.74	0.18
News Commentary	6.09 M	3.65 M	1	3.00	0.23

TABLE 2 – Size of the monolingual corpus and the ratio of entity labels

	hmm (t)	crf (+)	crf (w)	bi-lstm
French treebank	50.34	65.94	66.63	77.49
News Commentary	49.69	66.18	68.28	76.65

TABLE 3 – Overall results (F_1) on NER for French using semi-supervised learning described in §3

training" have already been shown in many NLP tasks, for example in syntactic parsing (McClosky *et al.*, 2006a).

4 Improving Results Using Reranking

We also propose a reranking algorithm using $\hat{\mathcal{L}} = \mathrm{rerank}(\mathcal{L}_1, ..., \mathcal{L}_n)$ where $\mathcal{L}_i$ is an assigned label by a learning algorithm, and $\hat{\mathcal{L}}$ is a reranked label by the rerank function. We exclude $\mathcal{L}_{hmm(t)}$, and we then obtain $\hat{\mathcal{L}}$ from $\mathrm{rerank}(\mathcal{L}_{crf(+)}, \mathcal{L}_{crf(w)}, \mathcal{L}_{bilstm})$ for reranking labels. We calculate the rerank function as follows :

$$\begin{array}{rcll} rerank(\cdot) & = & \mathrm{argmax}(\texttt{per}, \texttt{loc}, \texttt{org}) & \text{if there is } \textit{any} \text{ entity label} \\ & & \texttt{O} & \text{otherwise} \end{array}$$

For each entity score (`per`, `loc` or `org`), we calculate

$$\alpha_1 \mathcal{L}_{crf(+)} + \alpha_2 \mathcal{L}_{crf(w)} + \alpha_3 \mathcal{L}_{bilstm} \tag{2}$$

where α_i is a normalized weight. For example, α_1 for $\mathcal{L}_{crf(+)}$ is calculated by its baseline result in F_1 being normalized by the sum of all F_1 scores by learning algorithms : $\frac{64.57}{64.57+65.38+76.26}$. We use $\alpha_1 = 0.3131$, $\alpha_2 = 0.3171$ and $\alpha_3 = 0.3698$. For example, if a word *Loiret* (a department name in north-central France) is annotated as `I-LOC`, `I-ORG` and `I-LOC` by CRFs and bi-lstm, $\hat{\mathcal{L}}$ is `I-LOC` by the rerank calculation described in Figure 3. Finally, Table 4 shows the reranking results on NER for French.

$$\begin{array}{rll} \texttt{per} = & \alpha_1 \times 0 + \alpha_2 \times 0 + \alpha_3 \times 0 & = 0 \\ \texttt{loc} = & \alpha_1 \times 1 + \alpha_2 \times 0 + \alpha_3 \times 1 & = 0.6829 \\ \texttt{org} = & \alpha_1 \times 0 + \alpha_2 \times 1 + \alpha_3 \times 0 & = 0.3171 \end{array}$$

FIGURE 3 – An example for the rerank function to calculate $\mathrm{argmax}(\texttt{per}, \texttt{loc}, \texttt{org})$ for *Loiret*. We use $\alpha_1 = 0.3131$, $\alpha_2 = 0.3171$ and $\alpha_3 = 0.3698$.

		monolingual corpus
reranking based on Table 1	76.41	baseline
reranking + semi-supervised based on Table 3	77.95	French treebank
reranking + semi-supervised based on Table 3	77.03	News Commentary

TABLE 4 – Reranking results (F_1) on NER for French described in §4

5 Previous Work

Ollagnier *et al.* (2014) used the Open Edition corpus the Quaero Broadcast News Extended Named Entity corpus [14], which contains over 1.2M tokens. They evaluated NER results with LIA_NE (HMM-CRFs) [15], OpenNLP (Maximum entropy) [16] and Standford NER (CRFs) [17] with different sizes of training data. They obtained up to 57,9 F_1 score with LIA_NE. Partalas *et al.* (2016) compared NER systems in the e-Commerce domain for the cosmetics products by using handcrafted rules and machine learning techniques. They used two 50K tokens data sets (cosmetics magazines and blog articles). They presented only entity level results and a system of lexical combined syntactic rules with a domain-specific dictionary usually outperformed CRFs. Their rule-based systems yielded between 60.00 and 90.68 F_1 scores based on different entities.

There were efforts to create corpora annotated in named entities for French. Sagot *et al.* (2012) and Dutrey *et al.* (2012) manually annotated named entities in the French treebank, and in restricted domain such as oral dialogs recored by the EDF call center for information extraction, respectively. Okinina *et al.* (2013) enriched proper nouns by mining Wikipedia with the combination of DBpedia rules and a support vector machine classification. Hatmi (2012) used a cross-lingual approach by converting a rule-based English NER system into French by using lexical and grammar adaptations.

Fraisse *et al.* (2013) employed NER for better classification results on opinion mining and sentiment analysis. Sagot & Gábor (2014) corrected OCRed named entities errors by using a rule-based NER system. Brando *et al.* (2016) used NER for recognizing geographical references. These are applications, in which NER results improved other natural language processing tasks. Otherwise, Dupont & Tellier (2014) proposed a pipeline for French NER based on `Wapiti`.

6 Conclusion

In this paper, we trained and evaluated several sequence labeling algorithms to perform benchmarking for French named-entity recognition data. We then improved NER results using semi-supervised learning and reranking. We obtained up to 77.95%, in which we improved the result by over 34 points compared to the baseline results of the HMM.

While incorporating unlabeled data into a new model is a simple method, it would not be surprising that self-training is not normally effective because errors in the original model can be amplified in

[14] Available at `http://catalog.elra.info/product_info.php?products_id=1195`

[15] Available at `http://pageperso.lif.univ-mrs.fr/~frederic.bechet`

[16] Available at `https://opennlp.apache.org`

[17] Available at `https://nlp.stanford.edu/software/CRF-NER.shtml`

the new model (McClosky *et al.*, 2006a). We discard the weakest learner's results for the consensus method. This decision actually improves the NER results. For example, while hmm (t) obtains only 47.35% with intersection of all data for the French treebank, it achieve 50.34% by excluding $\mathcal{D}(\mathcal{M}_{hmm(t)})$ for data intersection. This semi-supervised process can be iterated, and it can be performed over other sets of unlabeled data for French. We assume that iterating the semi-supervised process and using a larger unlabeled data can improve NER results. We leave this to future work. However, while learning models for HMM and CRFs are relatively quick, we note that training bi-lstm using a large annotated corpus (*e.g.* over 3.65M tokens in News Commentary) takes several days even on a GPU for a single iteration.

Reranking basically selects the best result from the set of NER results for each sentence to have constructed high-performance NLP systems such as parsing (Charniak & Johnson, 2005). Combining reranking and self-training is not new, which has been, for example, already proposed for syntactic parsing (McClosky *et al.*, 2006b). While reported results show a minor improvement (*e.g.* we obtain 76.41% using ranking baseline, compared to 76.26% in the best baseline result), it is cheap and easy to implement for immediate improvements.

Comparison of results using previously proposed NER systems for French, and benchmark learning using other previously proposed NER data would be an interesting task, and we leave this to future work. All trained models and data will be publicly available at `https://github.com/jungyeul/taln2018`.

Remerciements

We would like to thank the anonymous reviewers for their suggestions and comments.

Références

ABEILLÉ A., CLÉMENT L. & TOUSSENEL F. (2003). Building a Treebank for French. In *Treebanks*, p. 165–188. Kluwer.

BOJANOWSKI P., GRAVE E., JOULIN A. & MIKOLOV T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, **5**, 135–146.

BRANDO C., DOMINGUÈS C. & CAPEYRON M. (2016). Evaluation of NER Systems for the Recognition of Place Mentions in French Thematic Corpora. In *Proceedings of the 10th Workshop on Geographic Information Retrieval*, GIR '16, p. 7 :1–7 :10, New York, NY, USA : ACM.

BRANTS T. (2000). TnT – A Statistical Part-of-Speech Tagger. In *Proceedings of the Sixth Conference on Applied Natural Language Processing*, p. 224–231, Seattle, Washington, USA : Association for Computational Linguistics.

BRODLEY C. E. & FRIEDL M. A. (1999). Identifying Mislabeled Training Data. *Journal of Artificial Intelligence Research*, **11**, 131–167.

CHARNIAK E. & JOHNSON M. (2005). Coarse-to-Fine n-Best Parsing and MaxEnt Discriminative Reranking. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, p. 173–180, Ann Arbor, Michigan : Association for Computational Linguistics.

DERNONCOURT F., LEE J. Y. & SZOLOVITS P. (2017). NeuroNER : an easy-to-use program for named-entity recognition based on neural networks. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing : System Demonstrations*, p. 97–102, Copenhagen, Denmark : Association for Computational Linguistics.

DUPONT Y. & TELLIER I. (2014). Un reconnaisseur d'entités nommées du Français. In *Proceedings of TALN 2014 (Volume 3 : System Demonstrations)*, p. 40–41, Marseille, France : Association pour le Traitement Automatique des Langues.

DUTREY C., CLAVEL C., ROSSET S., VASILESCU I. & ADDA-DECKER M. (2012). Quel est l'apport de la détection d'entités nommées pour l'extraction d'information en domaine restreint ? (What is the contribution of named entities detection for information extraction in restric- ted domain ?) [in French]. In *Proceedings of the Joint Conference JEP-TALN-RECITAL 2012, volume 2 : TALN*, p. 359–366, Grenoble, France : ATALA/AFCP.

FRAISSE A., PAROUBEK P. & FRANCOPOULO G. (2013). L'apport des Entités Nommées pour la classification des opinions minoritaires. In *Proceedings of TALN 2013 (Volume 2 : Short Papers)*, p. 588–595, Les Sables d'Olonne, France : ATALA.

GRAVES A. & SCHMIDHUBER J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, **18**(5-6), 602–610.

HATMI M. (2012). Adaptation d'un système de reconnaissance d'entités nommées pour le français à l'anglais à moindre coût (Adapting a French Named Entity Recognition System to English with Minimal Costs) [in French]. In *Proceedings of the Joint Conference JEP-TALN-RECITAL 2012, volume 3 : RECITAL*, p. 151–161, Grenoble, France : ATALA/AFCP.

KOEHN P., HOANG H., BIRCH A., CALLISON-BURCH C., FEDERICO M., BERTOLDI N., COWAN B., SHEN W., MORAN C., ZENS R., DYER C., BOJAR O., CONSTANTIN A. & HERBST E. (2007). Moses : Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, p. 177–180, Prague, Czech Republic : Association for Computational Linguistics.

LAFFERTY J. D., MCCALLUM A. & PEREIRA F. C. N. (2001). Conditional Random Fields : Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, p. 282–289, San Francisco, CA, USA : Morgan Kaufmann Publishers Inc.

LAVERGNE T., CAPPÉ O. & YVON F. (2010). Practical Very Large Scale CRFs. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, p. 504–513, Uppsala, Sweden : Association for Computational Linguistics.

MCCLOSKY D., CHARNIAK E. & JOHNSON M. (2006a). Effective Self-Training for Parsing. In *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference*, p. 152–159, New York City, USA : Association for Computational Linguistics.

MCCLOSKY D., CHARNIAK E. & JOHNSON M. (2006b). Reranking and Self-Training for Parser Adaptation. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, p. 337–344, Sydney, Australia : Association for Computational Linguistics.

OKININA N., NOUVEL D., FRIBURGER N. & ANTOINE J.-Y. (2013). Supervised learning on encyclopaedic resources for the extension of a lexicon of proper names dedicated to the recognition of named entities (Apprentissage supervisé sur ressources encyclopédiques pour l'enrichissement d'un lexique de noms propres destiné. In *Proceedings of TALN 2013 (Volume 2 : Short Papers)*, p. 667–674, Les Sables d'Olonne, France : ATALA.

OLLAGNIER A., FOURNIER S., BELLOT P. & BÉCHET F. (2014). Impact of the nature and size of the training set on performance in the automatic detection of named entities (Impact de la nature et de la taille des corpus d'apprentissage sur les performances dans la détection automatique des entités nommées) [in Frenc. In *Proceedings of TALN 2014 (Volume 2 : Short Papers)*, p. 511–516, Marseille, France : Association pour le Traitement Automatique des Langues.

PARTALAS I., LOPEZ C. & SEGOND F. (2016). Comparing Named-Entity Recognizers in a Targeted Domain : Handcrafted Rules vs. Machine Learning. In *Actes de la conférence conjointe JEP-TALN-RECITAL 2016, volume 2 : TALN*, p. 389–395, Paris, France : Association pour le Traitement Automatique des Langues.

RABINER L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, **77**(2), 257–286.

RAINA R., BATTLE A., LEE H., PACKER B. & NG A. Y. (2007). Self-taught Learning : Transfer Learning from Unlabeled Data. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, p. 759–766, New York, NY, USA : ACM.

SAGOT B. & GÁBOR K. (2014). Détection et correction automatique d'entités nommées dans des corpus OCRisés. In *Proceedings of TALN 2014 (Volume 2 : Short Papers)*, p. 437–442, Marseille, France : Association pour le Traitement Automatique des Langues.

SAGOT B., RICHARD M. & STERN R. (2012). Annotation référentielle du Corpus Arboré de Paris 7 en entités nommées (Referential named entity annotation of the Paris 7 French TreeBank) [in French]. In *Proceedings of the Joint Conference JEP-TALN-RECITAL 2012, volume 2 : TALN*, p. 535–542, Grenoble, France : ATALA/AFCP.

SCHMID H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.

SCHWENK H. (2008). Investigations on Large-Scale Lightly-Supervised Training for Statistical Machine Translation. In *Proceedings of the International Workshop on Spoken Language Translation*, p. 182–189, Hawaii, USA.

SUTTON C. & MCCALLUM A. (2012). An Introduction to Conditional Random Fields. *Foundations and Trends® in Machine Learning*, **4**(4), 267–373.

TJONG KIM SANG E. F. & DE MEULDER F. (2003). Introduction to the CoNLL-2003 Shared Task : Language-Independent Named Entity Recognition. In W. DAELEMANS & M. OSBORNE, Eds., *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, p. 142–147.