Named Entity Recognition in Speech Transcripts following an Extended Taxonomy

Mohamed Hatmi[1], *Christine Jacquin*[1], *Emmanuel Morin*[1], *Sylvain Meignier*[2]

[1] LINA, University of Nantes, France,
[2]LIUM, University of Le Mans, France
{mohamed.hatmi, christine.jacquin, emmanuel.morin}@univ-nantes.fr
sylvain.meignier@lium.univ-lemans.fr

Abstract

In this paper, we present a French named entity recognition (NER) system that was first developed as part of our participation in the ETAPE 2012 evaluation campaign and then extended to cover more entity types. The ETAPE 2012 evaluation campaign considers an hierarchical and compositional taxonomy that makes the NER task more complex. We present a multi-level methodology based on conditional random fields (CRFs). With respect to existing systems, our methodology allows a fine-grained annotation. Experiments were conducted using the manually annotated training and evaluation corpora provided by the organizers of the campaign. The obtained results are presented and discussed.

Index Terms: Named Entity Recognition, Structured Named Entities, CRF model.

1. Introduction

Named Entities (NEs) are defined as autonomous monoreferential linguistic expressions. They cover traditionally the names of all the person, organization and location. There are two most widespread approaches for the Named Entity Recognition (NER): symbolic approaches which rely on hand-coded grammar and gazetteers, and learning-based approaches which require large quantities of manually-annotated corpus [1].

NER from speech is mainly performed by transcribing speech and then applying NER approaches to transcripts. NER systems are adapted to fit in with the characteristics of automatic speech transcripts such as speech disfluencies, automatic speech recognition errors and out-of-vocabulary (OOV) problems. To that is added the problem of lack of some important NER features such as capitalization and punctuation. In order to improve speech NER, previous work has included restoring punctuation and capitalization in transcripts [2], using the Part-of-speech (POS) tags as features [3], incorporating indicative OOV words and ASR confidence features [4, 5, 6]. The ESTER 2 evaluation campaign [7, 8] has shown that the symbolic systems produce best results on manual transcripts whereas the learning-based systems show best results on automatic transcripts [9, 3].

NER systems require manually transcribed and annotated data, whether for performance evaluation or learning an annotation model. The adopted annotation schema has a direct impact on NER performance. For example, flat and relatively small entity types and granularity can achieve good results. The problem becomes more complex by using a fine-grained hierarchical taxonomy. As in [3], we propose a CRF-based approach that integrates the POS tags as features. However, the fundamental difference in our approach is that the adopted taxonomy is hierarchical and compositional.

In this paper, we present a French NER system that was first developed as part of our participation in the ETAPE 2012 evaluation campaign and then extended to cover more entity types. We propose a multi-level methodology which allows NER annotation following a fine-grained taxonomy. Three levels of annotation are defined : the first level consists of annotating the main categories, the second level has to do with the annotation of components and the last level deals with the problem of nested named entities.

This paper is organized as follows: Section 2 briefly presents the ETAPE evaluation campaign. Section 3 describes the Quaero extended taxonomy adopted in this campaign. Section 4 presents the corpora and the metrics used for evaluation. Section 5 presents the method used. Section 6 reports experimental results, while Section 7 concludes and presents future work.

2. The ETAPE evaluation campaign

The ETAPE evaluation campaign aimed to measure the performance of speech technologies for the French language [10]. Three main tasks were considered in this campaign: segmentation, transcription and information extraction. The evaluation concerned a variety of TV materials with various level of spontaneous speech and overlapping speech from multiple speakers. We are interested in the information extraction task that consists of detecting and categorizing all direct mentions of named entities following the Quaero named entity taxonomy.

3. Quaero named entity taxonomy

The Quaero annotation schema [11, 12] adopts a fine-grained hierarchical taxonomy. Named entity tagset is composed of 7 main categories and 32 sub-categories:

- **person**: invidual person (pers.ind), collectivity of persons (pers.coll),
- **location**: administrative location (loc.adm.town, loc.adm.reg, loc.adm.nat, loc.adm.sup), physical location (loc.phys.geo, loc.phys.hydro, loc.phys.astro),
- **organization**: services (org.ent), administration (org.adm),
- **function**: individual function (func.ind), collectivity of functions (func.coll),
- **human production**: manufactury object (prod.object), art products (prod.art), media products (prod.media), financial products (prod.fin), software (prod.soft), award

Proceedings of the First Workshop on Speech, Language and Audio in Multimedia (SLAM), Marseille, France, August 22-23, 2013.

(prod.award), transportation route (prod.serv), doctrine (prod.doctr), law (prod.rule),

- **time**: absolute date (time.date.abs), date relative to the discourse (time.date.rel), absolute hour (time.hour.abs), hour relative to the discourse (time.hour.rel),
- **amount**

Entity tags are organized in a structured way so that a named entity can include another one. For example, in the named entity "<func.ind>Minister of <org.adm>Education</org.adm></func.ind>", the *func.ind* type includes the *org.adm* type.

In addition, the elements inside the named entities are categorized using components. A named entity includes at least one component. For example, a street name can be composed of a kind and a name : <loc.oro> <kind> rue </kind> de <name> Vaugirard </name> </loc.oro> (<loc.oro> <kind> *street* </kind> *of* <name> *Vaugirard* </name> </loc.oro>). There are two kinds of components:

- Transverse components that can fit each type of entity: name, kind, qualifier, demonym, val, unit, object, range-mark,
- Specific components which are only used for a reduced set of components: pers.ind (name.last, name.first, name.middle, title), loc.add.phys (address.number, po-box, zip-code, other-address-component), and time.date (week, day, month, year, century, millenium, reference-era, time-modifier)

In cases of metonymy, the named entity is double annotated with the type to which the entity intrinsically belongs and with the type to which the entity belongs according to the context. For example, the named entity "Roland Garros" is annotated as *loc.fac* and *pers.ind* in the sentence "We are in Roland Garros"

4. Corpora and metrics description

We first present the corpora used in this work, then we present the different metrics used for evaluation.

4.1. Corpora

To carry out the experiments, we used the ETAPE and ESTER 2 data which have been made available to the participants in the ETAPE evaluation campaign. The ETAPE corpus consists of 42.5 hours of data recorded from different French speaking radio and TV stations which are BFM TV, La Chaîne Parlementaire and TV8. The ESTER 2 corpus comprises about 100 hours of radio broadcast from various French speaking radios which are France Inter, Radio France International, France Culture, Radio Classique, Africa number one, Radio Congo and Radio Television du Maroc. These corpora have been manually transcribed and annotated following the Quaero named entity taxonomy.

The ETAPE and ESTER data jointly are divided into three parts. The first part (1,761,677 words) is used to train various CRF models. The second part (108,340 words) is used as development corpus to experiment with and adjust some parameters. The remaining (106,803 words) is used in the final evaluation.

4.2. Evaluation metrics

The evaluation of the NER performance is performed using the SER and the F-measure.

The SER [13] (cf. equation 1) combines different types of error: insertions (I), deletions (D) and substitutions (errors both in span and in type (S_{ST}), errors in span (S_S), errors in type (S_T)). The corresponding equation is given by:

$$SER = \frac{D + I + S_{ST} + 0.5 \times (S_S + S_T)}{\#\ of\ entities\ in\ the\ reference} \quad (1)$$

The F-measure (cf. equation 4) combines precision and recall. Precision (cf. equation 3) represents the percentage of annotated entities that are correct. Recall (cf. equation 2) represents the percentage of correct entities that are annotated. The corresponding equations are given by:

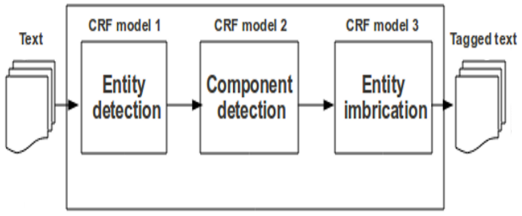
$$Recall = \frac{\#\ of\ correct\ annotated\ entities}{\#\ of\ annotated\ entities\ in\ the\ reference} \quad (2)$$

$$Precision = \frac{\#\ of\ correct\ annotated\ entities}{\#\ of\ annotated\ entities\ in\ the\ hypothesis} \quad (3)$$

$$F{-}measure = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (4)$$

5. Method used

Figure 1: *Architecture of the NER system*



Several machine learning methods have been used for annotating named entities in text. Annotation is considered as sequence labeling task. Each word in the sequence is labeled with its appropriate tag. Tags include the category of the named entity and the location of the word within the named entities (BIO annotation). The first word in a named entity is tagged with "entity-tag-B", and further named entity words are tagged with "entity-tag-I". Words outside named entities are tagged with "O" (Other). Several studies [14] have shown that discriminant methods like Maximum Entropy Markov Model (MEMM) [15] or CRF [16] overcome the difficulties encountered in generative methods like Hidden Markov Model (HMM) [17]. Discriminative models allow to relax the independence assumptions needed by generative models and to include much more features in the model. The machine learning method employed in this work is CRF which is a discriminative undirected graphical model.

Named entities in the training data are organized in a structured way as shown in section 3. Named entities contain nested tagging of other named entities and components. Therefore, words constituting the named entities can belong at

the same time to one or more categories. This is a problem for the preparation of the training data for classification because each word must be assigned to just one category. Here is an example of a sentence in training corpus:

Vous êtes <func.ind> <kind> directeur </kind> de l' <org.ent> école nationale d' assurance </org.ent> </func.ind>
(*You are the* <func.ind> <kind> *director* </kind> *of the* <org.ent> *national insurance school* </org.ent> </func.ind>)

In order to handle structured tagging, we defined three levels of annotation. The first level consists of annotating the 32 categories in a flat way. The second level has to do with the annotation of components. The last level allows overlapping annotation when a category includes another category. We trained a CRF model for each level of annotation. Figure 1 shows the architecture of the NER system.

We used the open source implementation of CRF CRF++ toolkit[1] to implement the different models.

5.1. Entity detection

The first CRF model aims to annotate a text with the 32 categories. To achieve this, we presented the training data in a flat way by separating nested annotations and eliminating component tags. Here is how the sample sentence given in section 5 is presented in training corpus:

Vous êtes <func.ind> directeur de l' </func.ind> <org.ent> école nationale d' assurance </org.ent>
(*You are the* <func.ind> *director of the* </func.ind> <org.ent> *national insurance school* </org.ent>)

We then encoded the obtained corpus in BIO notation (Begin, Inside, Outside) and train the CRF model. Two types of features are used to predict if a word is part of a named entity or not:

- Contextual information: we took a context window of $[-2, +2]$ and consider unigram, bigram and trigram combinations.
- Semantic and syntactic information: we used the French POS tagger LIA_Tagg[2] to assign a POS tag to each word. LIA_Tagg is a free tool based on HMM. The POS tags are enriched with four semantic labels for proper names: person, organization, location and product. We augmented the lexicon of LIA_Tagg with 111,600 new named entities extracted form the Web (30,300 persons, 18,700 organizations et 62,600 locations). This allows a first lexicon-based level of annotation.

Figure 2 shows an example of the CRF training corpus.

5.2. Component detection

The second CRF model is applied on the output of the first CRF model. Here, the goal is to predict a component label to each word of a text. Two kinds of features are used to train the CRF model:

Figure 2: Example of the CRF training corpus for entity annotation

Vous	PPERP	O
êtes	VEP	O
directeur	NMS	func.ind-b
de	DETFS	func.ind-i
l'	DETFS	func.ind-i
école	NFS	org.ent-b
nationale	AFS	org.ent-i
d'	PREPADE	org.ent-i
assurance	NFS	org.ent-i

Figure 3: Example of the CRF training corpus for component annotation

Vous	O	O
êtes	O	O
directeur	func.ind-b	kind-b
de	func.ind-i	O
l'	func.ind-i	O
école	org.ent-b	O
nationale	org.ent-i	O
d'	org.ent-i	O
assurance	org.ent-i	O

- Contextual information: we took a context window of $[-2, +2]$.
- Semantic information: we used the first CRF model to assign a named entity tag to each word of a text.

Figure 3 shows an example of the CRF training corpus.

5.3. Entity imbrication

The third CRF model is applied on the output of the second CRF model in order to deal with the problem of nested named entities. In fact, named entities are annotated in a flat way in the second level. Therefore, we need to change the boundaries of certain named entities in order to overlap other ones. For example, in the sentence given in section 5.1, we need to move the closing tag of the function entity to overlap the organization entity. We used two types of features to train the CRF model in order to learn the imbrication rules:

- Contextual information: we took a context window of $[-2, +2]$.
- Semantic information: we used the first and the second CRF models to assign a named entity and a component tag to each word of a text.

Figure 4 shows an example of the CRF training corpus.

[1] http://crf.sourceforge.net

[2] http://lia.univ-avignon.fr/fileadmin/documents/Users/Intranet/chercheurs/bechet/download_fred.html

Figure 4: Example of the CRF training corpus for entity imbrication

Vous	O	O	O
êtes	O	O	O
directeur	func.ind-b	kind-b	func.ind-b
de	func.ind-i	O	func.ind-i
l'	func.ind-i	O	func.ind-i
école	org.ent-b	O	func.ind-i
nationale	org.ent-i	O	func.ind-i
d'	org.ent-i	O	func.ind-i
assurance	org.ent-i	O	func.ind-i

Table 1: *Global NER results for the 32 categories computed on the manually-transcribed test corpus and ASR output. (F: F-measure, P: precision, R: recall)*

	Manual transcriptions	**ASR output (WER=23)**	**ASR output (WER=30)**
	F (R/P) (%)	**F (R/P) (%)**	**F (R/P) (%)**
S1	71.1 (62.5/18.7)	52.3 (43.7/65.1)	55.4 (46/69.6)
LL	**68** (64.5/71.7)	**49.9** (44.7/56.3)	**53.1** (46.59/61.91)
S2	66.4 (65.7/67.3)	40 (33.8/49.1)	41.4 (34.3/52.1)
S3	66.2 (65.6/77)	43.4 (41/46.1)	46.2 (44/48.6)
S4	66.2 (64.4/68.1)	43.7 (37.6/52.2)	48.8 (41.2/59.8)
S5	46.1 (67/61.4)	41.5 (40.7/42.4)	45 (43.3/46.8)
S6	55.5 (61/50.8)	23 (21.6/24.5)	27.7 (26.6/28.8)
S7	34.8 (28.3/45.1)	6.1 (19.1/9.2)	7.8 (23/11.6)

6. Results

We used the ETAPE test corpus to evaluate the performance of our system. This corpus contains 5,705 named entity occurrences and 7,174 component occurrences.

The NER system we used to participate in the ETAPE 2012 evaluation campaign annotates only the categories without components. It uses the first and the third CRF models. Table 1 shows the results obtained by our system, named *LL*, and the results of other participating NER systems for the 32 categories without components. The evaluation is performed on the manual transcriptions and ASR output with different Word Error Rate (WER). The ASR output is obtained from different ASR systems. The proposed approach achieves the second best F-measure on manual and automatic transcriptions for the 32 categories. Obviously, the F-measure decreases when dealing with automatic transcriptions. The NER features used for the well-written text appear insufficient to deal with noisy text and new specific ASR features are needed to be added.

After the ETAPE evaluation campaign, we extended our NER system to annotate also the components using the second CRF model. The system shows 37.5 % of SER on the manually-transcribed test corpus and 62.2 % of SER on the ASR output (WER=23). Table 2 shows the NER results by category. The results show good performance for some standard categories such as *pers.ind* and *loc.nat*, and poorer performances for others such as *loc.fac* and *prod.object*. These are characterized by a poor recall. This is mainly due to a low frequency in the training corpus. In addition, we observe some categorization errors particularly for the entities with metonymic sense (Paris as a town or as an organization) and between certain sub-categories of location and product. There is also some annotation ambiguity problems

Table 2: *NER results by category and component computed on the manually-transcribed test corpus and ASR output. (F: F-measure, P: precision, R: recall).*

	Manual transcriptions	**ASR output (WER=23)**	**Entities in reference**
	F (%)	**F (%)**	
amount	65.5	49.6	705
pers.ind	84.8	54.7	1,398
pers.coll	49.7	40.3	177
pers.other	0	0	1
time.date.abs	42.8	33	192
time.date.rel	74.7	65.5	348
time.hour.abs	58.1	32.8	46
time.hour.rel	74.2	62	84
loc.oro	0	50	2
loc.fac	13.7	11.3	81
loc.add.phys	0	0	4
loc.add.elec	83.3	46.15	18
loc.adm.town	71.4	44.5	279
loc.adm.reg	56.5	51.8	47
loc.adm.nat	86.9	77.9	276
loc.adm.sup	76.4	53.3	33
loc.phys.geo	23.5	0	30
loc.phys.hydro	0	0	5
loc.phys.astro	0	0	0
prod.object	6.2	0	58
prod.art	16.4	0	87
prod.media	68.1	56.8	164
prod.fin	19.6	11.6	84
prod.soft	0	0	0
prod.award	44.4	37.5	13
prod.serv	0	0	0
prod.doctr	0	33.3	3
prod.rule	54.5	0	5
prod.other	0	0	7
prod.unk	0	0	2
func.ind	59.8	51.6	383
func.coll	47.6	30.8	243
org.ent	45.7	38.3	307
org.adm	54.9	45.9	286
kind	50.4	42.3	1,163
extractor	40	33.3	4
qualifier	31.7	28.2	250
title	38.1	31.5	75
val	85	70.1	808
unit	87.6	75.4	463
object	61.7	49	225
range-mark	77	59.3	71
day	85.7	71.8	36
week	84.9	77.7	39
month	69	59.8	74
year	88.8	72.4	95
century	100	80	3
reference-era	25	33.3	2
time-modifier	64.3	57.2	337
award-cat	0	0	2
demonym	57.9	49	206
name	68.1	57.3	1,593
name.last	82.9	41.2	928
name.first	86.4	53.5	1,032
name.nickname	30.6	28.2	71
all	69.7	52.5	12,879

which concerns particularly some named entities composed of common nouns such as for *pers.coll* (e.g. classes populaires (working classes)), *func.coll* (e.g. sentinelles citoyennes (sentinel citizens)) and *prod.art* (e.g. devons-nous payer 100 % des tudes des futurs traders ? (should we pay 100 % of the studies of future traders)).

7. Conclusions

In this paper, we have presented a French NER system using CRF. We have proposed a multi-level method that annotates named entities following a fine-grained hierarchical taxonomy. The evaluation has shown good results on manual and automatic transcriptions. Future work will concentrate on improving the annotation of some categories and components that shows a weak performance. This is due to their limited appearance in the training corpus. We also intend to explore new features gathered from the ASR process to improve NER in automatic transcriptions.

8. References

[1] D. Nadeau and S. Sekine, "A survey of named entity recognition and classification," *Linguisticae Investigationes*, vol. 30, pp. 3–26, January 2007.

[2] A. Gravano, M. Jansche, and M. Bacchiani, "Restoring punctuation and capitalization in transcribed speech," in *Proceedings of ICASSP '09*, Taipei, Taiwan, 2009, pp. 4741–4744.

[3] F. Béchet and E. Charton, "Unsupervised knowledge acquisition for Extracting Named Entities from speech," in *Proceedings of ICASSP '10*, Dallas, Texas, USA, 2010, pp. 5338–5341.

[4] D. D. Palmer and M. Ostendorf, "Improving information extraction by modeling errors in speech recognizer output," in *Proceedings of HLT '01*, San Diego, California, USA, 2001, pp. 1–5.

[5] K. Sudoh, H. Tsukada, and H. Isozaki, "Incorporating speech recognition confidence into discriminative named entity recognition of speech data," in *Proceedings of ACL '06*, Sydney, Australia, 2006, pp. 617–624.

[6] C. Parada, M. Dredze, and F. Jelinek, "OOV Sensitive Named-Entity Recognition in Speech." in *Proceedings of INTERSPEECH '11*, Florence, Italy, 2011, pp. 2085–2088.

[7] S. Galliano, G. Gravier, and L. Chaubard, "The ester 2 evaluation campaign for the rich transcription of French radio broadcasts," in *Proceedings of INTERSPEECH '09*, Brighton, UK, 2009, pp. 2583–2586.

[8] A. Zidouni, S. Rosset, and H. Glotin, "Efficient combined approach for named entity recognition in spoken language," in *Proceedings of INTERSPEECH '10*, Makuhari, Japan, 2010, pp. 1293–1296.

[9] J.-h. Kim and P. Woodland, "A Rule-Based Named Entity Recognition System for Speech Input," in *Proceedings of ICSLP '00*, Beijing, China, 2000, pp. 521–524.

[10] G. Gravier, G. Adda, N. Paulson, M. Carré, A. Giraudel, and O. Galibert, "The etape corpus for the evaluation of speech-based tv content processing in the french language," in *Proceedings of LREC '12*, Istanbul, Turkey, 2012, pp. 114–118.

[11] S. Rosset, C. Grouin, and P. Zweigenbaum, "Entités nommées structurées : guide d'annotation quaero," in *Technical Report*, LIMSI-CNRS, France, 2011.

[12] O. Galibert, S. Rosset, C. Grouin, P. Zweigenbaum, and L. Quintard, "Structured and extended named entity evaluation in automatic speech transcriptions," in *Proceedings of IJCNLP '11*, Chiang Mai, Thailand, 2011, pp. 518–526.

[13] J. Makhoul, F. Kubala, R. Schwartz, and R. Weischedel, "Performance measures for information extraction," in *Proceedings of DARPA Broadcast News Workshop*, Herndon, Virginia, USA, 1999, pp. 249–252.

[14] C. Raymond and G. Riccardi, "Generative and discriminative algorithms for spoken language understanding." in *Proceedings of INTERSPEECH '07*, Antwerp, Belgium, 2007, pp. 1605–1608.

[15] A. L. Berger, V. J. D. Pietra, and S. A. D. Pietra, "A maximum entropy approach to natural language processing," *Comput. Linguist.*, vol. 22, pp. 39–71, 1996.

[16] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of ICML'01*, Williamstown, MA, USA, 2001, pp. 282–289.

[17] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," in *Proceedings of the IEEE '89*, 1989, pp. 257–286.