Proposal for an Extension of Traditional Named Entities: From Guidelines to Evaluation, an Overview

Cyril Grouin$^{\alpha}$**, Sophie Rosset**$^{\alpha}$**, Pierre Zweigenbaum**$^{\alpha}$
Karën Fort$^{\beta,\gamma}$**, Olivier Galibert**$^{\delta}$**, Ludovic Quintard**$^{\delta}$
$^{\alpha}$LIMSI–CNRS, France $^{\beta}$INIST–CNRS, France $^{\gamma}$LIPN, France $^{\delta}$LNE, France
{cyril.grouin,sophie.rosset,pierre.zweigenbaum}@limsi.fr
karen.fort@inist.fr, {olivier.galibert,ludovic.quintard}@lne.fr

Abstract

Within the framework of the construction of a fact database, we defined guidelines to extract named entities, using a taxonomy based on an extension of the usual named entities definition. We thus defined new types of entities with broader coverage including substantive-based expressions. These extended named entities are hierarchical (with types and components) and compositional (with recursive type inclusion and metonymy annotation). Human annotators used these guidelines to annotate a 1.3M word broadcast news corpus in French. This article presents the definition and novelty of extended named entity annotation guidelines, the human annotation of a global corpus and of a mini reference corpus, and the evaluation of annotations through the computation of inter-annotator agreements. Finally, we discuss our approach and the computed results, and outline further work.

1 Introduction

Within the framework of the Quaero project—a multimedia indexing project—we organized an evaluation campaign on named entity extraction aiming at building a fact database in the news domain, the first step being to define what kind of entities are needed. This campaign focused on broadcast news corpora in French. While traditional named entities include three major classes (persons, locations and organizations), we decided to extend the coverage of our campaign to new types of entities and to broaden their main parts-of-speech from proper names to substantives, this extension being necessary for ever-increasing knowledge extraction from documents. We thus produced guidelines to specify the way corpora had to be annotated, and launched the annotation process.

In this paper, after covering related work (Section 2), we describe the taxonomy we created (Section 3) and the annotation process and results (Section 4), including the corpora we gathered and the tools we developed to facilitate annotation. We then present inter-annotator agreement measures (Section 5), outline limitations (Section 6) and conclude on perspectives for further work (Section 7).

2 Related work

2.1 Named entity definitions

Named Entity recognition was first defined as recognizing proper names (Coates-Stephens, 1992). Since MUC-6 (Grishman and Sundheim, 1996; SAIC, 1998), named entities have been proper names falling into three major classes: persons, locations and organizations.

Proposals were made to sub-divide these entities into finer-grained classes. The "politicians" subclass was proposed for the "person" class by (Fleischman and Hovy, 2002) while the "cities" subclass was added to the "location" class by (Fleischman, 2001; Lee and Lee, 2005).

The CONLL conference added a miscellaneous type that includes proper names falling outside the previous classes. Some classes have thus sometimes been added, e.g. the "product" class by (Bick, 2004; Galliano et al., 2009).

Proceedings of the Fifth Law Workshop (LAW V), pages 92–100,
Portland, Oregon, 23-24 June 2011. ©2011 Association for Computational Linguistics

Specific entities are proposed and handled in some tasks: “language” or “shape” for question-answering systems in specific domains (Rosset et al., 2007), “email address” or “phone number” to process electronic messages (Maynard et al., 2001). Numeric types are also often described and used. They include “date”, “time”, and “amount” types (“amount” generally covers money and percentage). In specific domains, entities such as gene, protein, are also handled (Ohta, 2002), and campaigns are organized for gene detection (Yeh et al., 2005). At the same time, extensions of named entities have been proposed: (Sekine, 2004) defined a complete hierarchy of named entities containing about 200 types.

2.2 Named Entities and Annotation

As for any other kind of annotation, some aspects are known to lead to difficulties in obtaining coherence in the manual annotation process (Ehrmann, 2008; Fort et al., 2009). Three different classes of problems are distinguished: (1) selecting the correct category in cases of ambiguity, where one entity can fall into several classes, depending on the context (“*Paris*” can be a town or a person name); (2) detecting the boundaries (in a person designation, is only the proper name to be annotated or the trigger “*Mr*” too?) and (3) annotating metonymies (“*France*” can be a sports team, a country, etc.).

In the ACE Named Entity task (Doddington et al., 2004), a complex task, the obtained inter-annotator agreement was 0.86 in 2002 and 0.88 in 2003. Some tasks obtain better agreement. Desmet and Hoste (2010) described the Named Entity annotation realized within the Sonar project, where Named Entity are clearly simpler. They follow the MUC Named Entity definition with the subtypes as proposed by ACE. The agreement computed over the Sonar Dutch corpus ranges from 0.91 to 0.97 (kappa values) depending of the emphasized elements (span, main type, subtype, etc.).

3 Taxonomy

3.1 Guidelines production

Having in mind the objective of building a fact database through the extraction of named entities from texts, we defined a richer taxonomy than those used in other information extraction works.

Following (Bonneau-Maynard et al., 2005; Alex et al., 2010), the annotation guidelines were first written from December 2009 to May 2010 by three researchers managing the manual annotation campaign. During guidelines production, we evaluated the feasibility of this specific annotation task and the usefulness of the guidelines by annotating a small part of the target corpus. Then, these guidelines were delivered to the annotators. They consist of a description of the objects to annotate, general annotation rules and principles, and more than 250 prototypical and real examples extracted from the corpus (Rosset et al., 2010). Rules are important to set the general way annotations must be produced. Additionally, examples are essential for human annotators to grasp the annotation rationale more easily.

Indeed, while producing the guidelines, we knew that the given examples would never cover all possible cases because of the specificity of language and of the ambiguity of formulations and situations described in corpora, as shown in (Fort et al., 2009). Nevertheless, guidelines examples must be considered as a way to understand the final objective of the annotation work. Thanks to numerous meetings from May to November 2010, we gathered feedback from the annotators (four annotators plus one annotation manager). This feedback allowed us to clarify and extend the guidelines in several directions. The guidelines are 72 pages long and consist of 3 major parts: general description of the task and the principles (25% of the overall document), presentation of each type of named entity (57%), and a simpler “cheat sheet” (18%).

3.2 Definition

We decided to use the three general types of named entities: *name* (person, location, organization) as described in (Grishman and Sundheim, 1996; SAIC, 1998), *time* (date and duration), and *quantity* (amount). We then included named entities extensions proposed by (Sekine, 2004; Galliano et al., 2009) (respectively products and functions) and we extended the definition of named entities to expressions which are not composed of proper names (e.g., phrases built around substantives). The extended named entities we defined are both hierarchical and compositional. For example, type *pers* (person) is split into two subtypes, *pers.ind* (indi-

Person			**Function**		
pers.ind (individual person)	*pers.coll* (group of persons)		*func.ind* (individual function)	*func.coll* (collectivity of functions)	
Location			**Product**		
administrative (*loc.adm.town, loc.adm.reg, loc.adm.nat, loc.adm.sup*)	physical (*loc.phys.geo, loc.phys.hydro, loc.phys.astro*)	facilities (*loc.fac*), oronyms (*loc.oro*), address (*loc.add.phys, loc.add.elec*)	*prod.object* (manufactured object)	*prod.serv* (transportation route)	*prod.fin* (financial products)
			prod.doctr (doctrine)	*prod.rule* (law)	*prod.soft* (software)
			prod.art	*prod.media*	*prod.award*
Organization			**Time**		
org.adm (administration)	*org.ent* (services)		*time.date.abs* (absolute date), *time.date.rel* (relative date)	*time.hour.abs* (absolute hour), *time.hour.rel* (relative hour)	
Amount					
amount (with unit or general object), including duration					

Table 1: Types (in bold) and subtypes (in italic)

vidual person) and *pers.coll* (collective person), and *pers* entities are composed of several components, among which are *name.first* and *name.last*.

3.3 Hierarchy

We used two kinds of elements: types and components. The types with their subtypes categorize a named entity. While types and subtypes were used before (ACE, 2000; Sekine, 2004; ACE, 2005; Galliano et al., 2009), we consider that structuring the contents of an entity (its components) is important too. Components categorize the elements inside a named entity.

Our taxonomy is composed of 7 main types (*person, function, location, product, organization, amount* and *time*) and 32 subtypes (Table 1). Types and subtypes refer to the general category of a named entity. They give general information about the annotated expression. Almost each type is then specified using subtypes that either mark an opposition between two major subtypes (individual person vs. collective person), or add precisions (for example for locations: administrative location, physical location, etc.).

This two-level representation of named entities, with types and components, constitutes a novel approach.

Types and subtypes To deal with the intrinsic ambiguity of named entities, we defined two specific transverse subtypes: 1. *other* for entities with a different subtype than those proposed in the taxonomy (for example, *prod.other* for games), and 2. *unknown* when the annotator does not know which subtype to use.

Types and subtypes constitute the first level of annotation. They refer to a general segmentation of the world into major categories. Within these categories, we defined a second level of annotation we call *components*.

Components Components can be considered as clues that help the annotator to produce an annotation: either to determine the named entity type (e.g. a first name is a clue for the *pers.ind* named entity subtype), or to set the named entity boundaries (e.g. a given token is a clue for the named entity, and is within its scope, while the next token is not a clue and is outside its scope). Components are second-level elements, and can never be used outside the scope of a type or subtype element. An entity is thus composed of components that are of two kinds: transverse components and specific components (Table 2). Transverse components can be used in several types of entities, whereas specific components can only be used in one type of entity.

Transverse components			
name (name of the entity), *kind* (hyperonym of the entity), *qualifier* (qualifying adjective), *demonym* (inhabitant or ethnic group name), *demonym.nickname* (inhabitant or ethnic group nickname), *val* (a number), *unit* (a unit), *extractor* (an element in a series), *range-mark* (range between two values), *time-modifier* (a time modifier).			
pers.ind	**loc.add.phys**	**time.date.abs/rel**	**amount**
name.last, name.first,	*address-number, po-box,*	*week, day, month, year,*	*object*
name.middle, pseudonym,	*zip-code,*	*century, millennium,*	**prod.award**
name.nickname, title	*other-address-component*	*reference-era*	*award-cat*

Table 2: Transverse and specific components

3.4 Composition

Another original point in this work is the compositional nature of the annotations. Entities can be compositional for three reasons: (i) a type contains a component; (ii) a type includes another type, used as a component; and (iii) in cases of metonymy. During the Ester II evaluation campaign, there was an attempt to use compositionality in named entities for two categories: persons and functions, where a person entity could contain a function entity.

> *<pers.hum> <func.pol> président </func.pol>*
> *<pers.hum> Chirac </pers.hum> </pers.hum>*

Nevertheless, the Ester II evaluation did not take this inclusion into account and only focused on the encompassing annotation (*<pers.hum> président Chirac </pers.hum>*). We drew our inspiration from this experience, and allowed the annotators and the systems to use compositionality in the annotations.

Cases of inclusion can be found in the *function* type (Figure 1), where type *func.ind*, which spans the whole expression, includes type *org.adm*, which spans the single word "*budget*". In this case, we consider that the designation of this function ("*ministre du budget*") includes both the kind ("*ministre*") and the name ("*budget*") of the ministry, which itself is typed as is relevant (*org.adm*). Recursive cases of embedding can be found when a subtype includes another named entity annotated with the same subtype (*org.ent* in Figure 2).

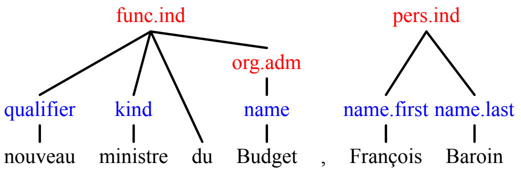

Figure 1: Multi-level annotation of entity types (red tags) and components (blue tags): *new minister of budget , François Baroin*.

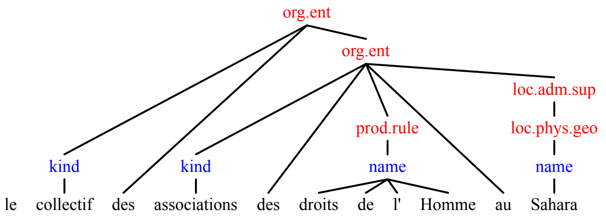

Figure 2: Recursive embedding of the same subtype: *Collective of the Human Rights Organizations in Sahara*.

Cases of metonymy include strict metonymy (a term is substituted with another one in a relation of contiguity) and antonomasia (a proper name is used as a substantive or vice versa). In such cases, the entity must be annotated with both types, first (inside) with the intrinsic type of the entity, then (outside) with the type that corresponds to the result of the metonymy. Basically, country names correspond to "national administrative" locations (*loc.adm.nat*) but they can also designate the administration (*org.adm*) of the country (Figure 3).

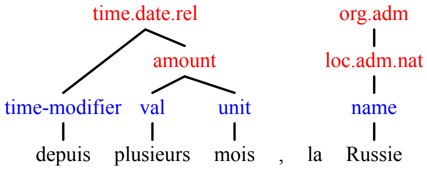

Figure 3: Annotation with a metonymic use of country "Russia" as its government: *for several months , Russia...*

3.5 Boundaries

Our definition of the scope of entities excludes relative clauses, subordinate clauses, and interpolated clauses: the annotation of an entity must end before these clauses. If an interpolated clause occurs inside an entity, its annotation must be split. Moreover, two distinct persons sharing the same last name must be annotated as two separate entities (Figure 4); we intend to use relations between entities to gather these segments in the next step of the project.

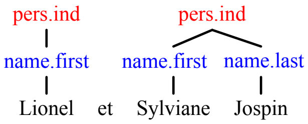

Figure 4: Separate (coordinated) named entities.

4 Annotation process

4.1 Corpus

We managed the annotation of a corpus of about one hundred hours of transcribed speech from several French-speaking radio stations in France and Morocco. Both news and entertainment shows were transcribed, including dialogs, with speaker turns.[1]

Once annotated, the corpus was split into a development corpus: one file from a French radio station;[2] a training corpus: 188 files from five French stations[3] and one Moroccan station;[4] and a test corpus: 18 files from two French stations already studied in the training corpus[5] and from unseen sources, both radio[6] and television,[7] in order to evaluate the robustness of systems. These data have been used in the 2011 Quaero named entity evaluation campaign.

This corpus allows us to perform different evaluations, depending of the knowledge the systems have of the source (source seen in the training corpus vs. unseen source), the kind of show (news vs. entertainment), the language style (popular vs. refined), and the type of media (radio vs. television).

4.2 Tools for annotators

To perform our test annotations (see Section 2.2), we developed a very simple annotation tool as an interface based on XEmacs. We provided the human annotators with this tool and they decided to use it for the campaign, despite the fact that it is very simple and that we told them about other, more generic, annotation tools such as GATE[8] or Glozz.[9] This is probably due to the fact that apart from being very simple to install and use, it has interesting features.

The first feature is the insertion of annotations using combinations of keyboard shortcuts based on the initial of each type, subtype and component name. For example, combination F2 key + initial keys is used to annotate a subtype (*pers.ind*, etc.), F3 + keys for a transverse component (*name, kind*, etc.), F4 + keys for a specific component (*name.first*, etc.), and F5 to delete the annotation selected with the cursor (both opening and closing tags).

The second feature is boundary management: if the annotator puts the cursor over the token to annotate, the annotation tool will handle the boundaries of this token; opening and closing tags will be inserted around the token.

However, it presents some limitations: tags are inserted in the text (which makes visualization more complex, especially for long sentences or in cases of multiple annotations on the same entity), no personalization is offered (tags are of only one color), and there is no function to express annotator uncertainty (the user must choose among several possible tags the one that fits the best;[10] while producing the guidelines, we did not consider it could be of interest: as a consequence, no uncertainty management was implemented). Therefore, this tool allows users to insert tags rapidly into a text, but it offers no external resources, as real annotation tools (e.g. GATE) often do.

[1] Potential named entities may be split across several segments or turns.

[2] News from France Culture.

[3] News from France Culture (refined language), France Info (news with short news headlines), France Inter (generalist radio station), Radio Classique (classical music and economic news), RFI (international radio broadcast out of France).

[4] News from RTM (generalist French speaking radio).

[5] News from France Culture, news and entertainment from France Inter.

[6] A popular entertainment show from Europe 1.

[7] News from Arte (public channel with art and culture), France 2 (public generalist channel), and TF1 (private generalist popular channel).

[8] `http://gate.ac.uk/`

[9] `http://www.glozz.org/`

[10] Uncertainty can be found in cases of lack of context.

These simplistic characteristics combined with a fast learning curve allow the annotators to rapidly annotate the corpora. Annotators were allowed not to annotate the transverse component *name* (only if it was the only component in the annotated phrase, e.g. “Russia” in Figure 3, blue tag) and to annotate events, even though we do not focus on this type of entity as of yet. We therefore also provided a normalization tool which adds the transverse component *name* in these instances, and which removes event annotations.

4.3 Corpus annotation

Global annotation It took four human annotators two months and a half to annotate the entire corpus (10 man-month). These annotators were hired graduate students (MS in linguistics). The overall corpus was annotated in duplicate. Regular comparisons of annotations were performed and allowed the annotators to develop a methodology, which was subsequently used to annotate the remaining documents.

Mini reference corpus To evaluate the global annotation, we built a mini reference corpus by randomly selecting 400 sentences from the training corpus and distributing them into four files. These files were annotated by four graduate human annotators from two research institutes (Figure 5) with two humans per institute, in about 10 hours per annotator.

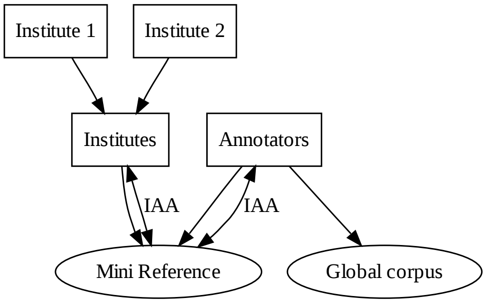

Figure 5: Creation of mini reference corpus and computation of inter-annotator agreement. Institute 1 = LIMSI–CNRS, Institute 2 = INIST–CNRS

First, we merged the annotations of each file within a given institute (1.5h per pair of annotators), then merged the results across the two institutes (2h). Finally, we merged the results with the annotations of the hired annotators (8h). We thus spent about 90 hours to annotate and merge annotations in this mini reference corpus (0.75 man-month).

4.4 Annotation results

Our broadcast news corpus includes 1,291,225 tokens, among which there are 954,049 non-punctuation tokens. Its annotation contains 113,885 named entities and 146,405 components (Table 3), i.e. one entity per 8.4 non-punctuation tokens, and one component per 6.5 non-punctuation tokens. There is an average of 6 annotations per line.

Data Inf.	Training	Test
# shows	188	18
# lines	43,289	5,637
# words	1,291,225	108,010
# entity types	113,885	5,523
# distinct types	41	32
# components	146,405	8,902
# distinct comp.	29	22

Table 3: Statistics on annotated corpora.

5 Inter-Annotator Agreement

5.1 Procedure

During the annotation campaign, we measured several criteria on a regular basis: inter-annotator agreement and disagreement. We used them to correct erroneous annotations, and mapped these corrections to the original annotations. We also used these measures to give the annotators feedback on the encountered problems, discrepancies, and residual errors. Whereas we performed these measurements all along the annotation campaign, this paper focuses on the final evaluation on the mini reference corpus.

5.2 Metrics

Because human annotation is an interpretation process (Leech, 1997), there is no “truth” to rely on. It is therefore impossible to really evaluate the validity of an annotation. All we can and should do is to evaluate its reliability, i.e. the consistency of the annotation across annotators, which is achieved through computation of the inter-annotator agreement (IAA).

The best way to compute it is to use one of the Kappa family coefficients, namely Cohen's Kappa (Cohen, 1960) or Scott's Pi (Scott, 1955), also known as Carletta's Kappa (Carletta, 1996),[11] as they take chance into account (Artstein and Poesio, 2008). However, these coefficients imply a comparison with a "random baseline" to establish whether the correlation between annotations is statistically significant. This baseline depends on the number of "markables", i.e. all the units that *could* be annotated.

In the case of named entities, as in many others, this "random baseline" is known to be difficult—if not impossible—to identify (Alex et al., 2010). We wish to analyze this in more detail, to see how we could actually compute these coefficients and what information it would give us about the annotation.

Markables	Annotators	Both institutes
	F = 0.84522	F = 0.91123
U1: n-grams	κ = 0.84522 π = 0.81687	κ = 0.91123 π = 0.90258
U2: n-grams ≤ 6	κ = 0.84519 π = 0.81685	κ = 0.91121 π = 0.90257
U3: NPs	κ = 0.84458 π = 0.81628	κ = 0.91084 π = 0.90219
U4: Ester entities	κ = 0.71300 π = 0.71210	κ = 0.82607 π = 0.82598
U5: Pooling	κ = 0.71300 π = 0.71210	κ = 0.82607 π = 0.82598

Table 4: Inter-Annotator Agreements (κ stands for Cohen's Kappa, π for Scott's Pi, and F for F-measure). IAA values were computed by taking as the reference the hired annotators' annotation or that obtained by merging from both institutes (see Figure 5).

In the present case, we could consider that, potentially, all the noun phrases can be annotated (row U3 in Table 4, based on the PASSAGE campaign (Vilnat et al., 2010)). Of course, this is a wrong approximation as named entities are not necessarily noun phrases (e.g., "à partir de l'automne prochain", *from next autumn*).

We could also consider all n-grams of tokens in the corpus (row U1). However, it would be more relevant to limit their size. For a maximum size of six, we get the results shown in row U2. All this, of course, is artificial, as the named entity annotation process is not random.

To obtain results that are closer to reality, we could use numbers of named entities from previous named entity annotation campaigns (row U4 based on the Ester II campaign (Galliano et al., 2009)), but as we consider here a largely extended version of those, the results would again be far from reality.

Another solution is to consider as "markables" all the units annotated by at least one of the annotators (row U5). In this particular case, units not annotated by any of the annotators (i.e. silence) are overlooked.

The lowest IAA will be the one computed with this last solution, while the highest IAA will be equal to the F-measure (i.e. the measure computed with all the markables as shown in row U1 in Table 4). We notice that the first two solutions (U1 and U2 with n-grams) are not acceptable because they are far from reality; even extended named entities are sparse annotations, and just considering all tokens as 'markables' is not suitable. The last three ones seem to be more relevant because they are based on an observed segmentation on similar data. Still, the U3 solution (NPs) overrates the number of markables because not all noun phrases are extended named entities. Although the U4 solution (Ester entities) is based on the same corpus used for a related task, it underrates the number of markables because that task produced 16.3 times less annotations. Finally the U5 solution (pooling) gives the lower bound for the κ estimation which is an interesting information but may easily undervalue the quality of the annotation.

As (Hripcsak and Rothschild, 2005) showed, in our case κ tends towards the F-measure when the number of negative cases tends towards infinity. Our results show that it is hard to build a justifiable hypothesis on the number of markables which is larger than the number of actually annotated entities while keeping κ significantly under the F-measure. But building no hypothesis leads to underestimating the κ value.

This reinforces the idea of using the F-measure as the main inter-annotator agreement measure for named entity annotation tasks.

[11] For more details on terminology issues, we refer to the introduction of (Artstein and Poesio, 2008).

6 Limitations

We used syntax to define some components (e.g. a *qualifier* is an adjective) and to set the scope of entities (e.g. stop at relative clauses). Nevertheless, this syntactic definition cannot fit all named entities, which are mainly defined according to semantics: the phrase "*dans les mois qui viennent*" ("*in the coming months*") expresses an entity of type *time.date.rel* where the relative clause "*qui viennent*" is part of the entity and contributes the *time-modifier* component.

The distinction between some types of entities may be fuzzy, especially for the organizations (is the Social Security an administrative organization or a company?) and for context-dependent annotations (is *lemonde.fr* a URL, a media, or a company?). As a consequence, some entity types might be converted into specific components in a future revision, e.g. the *func* type could become a component of the *pers* type, where it would become a description of the function itself instead of the person who performs this function (Figure 6).

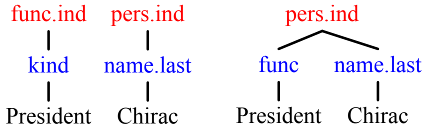

Figure 6: Possible revision: current annotation (left), transformation of *func* from entity to component (right).

7 Conclusion and perspectives

In this paper, we presented an extension of the traditional named entity categories to new types (functions, civilizations) and new coverage (expressions built over a substantive). We created guidelines that were used by graduate annotators to annotate a broadcast news corpus.

The organizers also annotated a small part of the corpus to build a mini reference corpus. We evaluated the human annotations with our mini-reference corpus: the actual computed κ is between 0.71 et 0.85 which, given the complexity of the task, seems to indicate a good annotation quality. Our results are consistent with other studies (Dandapat et al., 2009) in demonstrating that human annotators' training is a key asset to produce quality annotations.

We also saw that guidelines are never fixed, but evolve all along the annotation process due to feedback between annotators and organizers; the relationship between guidelines producers and human annotators evolved from "parent" to "peer" (Akrich and Boullier, 1991). This evolution was observed during the annotation development, beyond our expectations. These data have been used for the 2011 Quaero Named Entity evaluation campaign.

Extensions and revisions are planned. Our first goal is to add a new type of named entity for all kinds of events; guidelines are being written and human annotation tests are ongoing. We noticed that some subtypes are more difficult to disambiguate than others, especially *org.adm* and *org.ent* (definition and examples in the guidelines are not clear enough). We shall make decisions about this kind of ambiguity, either by merging these subtypes or by reorganizing the distinctions within the *organization* type. We also plan to link the annotated entities using relations; further work is needed to define more precisely the way we will perform these annotations. Moreover, the taxonomy we defined was applied to a broadcast news corpus, but we intend to use it in other corpora. The annotation of an old press corpus was performed according to the same process. Its evaluation will start in the coming months.

Acknowledgments

We thank all the annotators who did such a great work on this project, as well as Sabine Barreaux (INIST–CNRS) for her work on the reference corpus.

This work was partly realized as part of the Quaero Programme, funded by Oseo, French State agency for innovation and by the French ANR Etape project.

References

ACE. 2000. Entity detection and tracking, phase 1, ACE pilot study. Task definition. http://www.nist.gov/speech/tests/ace/phase1/doc/summary-v01.htm.

ACE. 2005. ACE (Automatic Content Extraction) English annotation guidelines for entities version 5.6.1 2005.05.23. http://www.ldc.upenn.edu/Projects/ACE/docs/English-Entities-Guidelines_v5.6.1.pdf.

Madeleine Akrich and Dominique Boullier. 1991. Le mode d'emploi, genèse, forme et usage. In Denis Chevallier, editor, *Savoir faire et pouvoir transmettre*, pages 113–131. éd. de la MSH (collection Ethnologie de la France, Cahier 6).

Beatrice Alex, Claire Grover, Rongzhou Shen, and Mijail Kabadjov. 2010. Agile Corpus Annotation in Practice: An Overview of Manual and Automatic Annotation of CVs. In *Proc. of the Fourth Linguistic Annotation Workshop*, pages 29–37, Uppsala, Sweden. ACL.

Ron Artstein and Massimo Poesio. 2008. Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics*, 34(4):555–596.

Eckhard Bick. 2004. A named entity recognizer for danish. In *LREC'04*.

Hélène Bonneau-Maynard, Sophie Rosset, Christelle Ayache, Anne Kuhn, and Djamel Mostefa. 2005. Semantic Annotation of the French Media Dialog Corpus. In *InterSpeech*, Lisbon.

Jean Carletta. 1996. Assessing Agreement on Classification Tasks: the Kappa Statistic. *Computational Linguistics*, 22:249–254.

Sam Coates-Stephens. 1992. The analysis and acquisition of proper names for the understanding of free text. *Computers and the Humanities*, 26:441–456.

Jacob Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46.

Sandipan Dandapat, Priyanka Biswas, Monojit Choudhury, and Kalika Bali. 2009. Complex Linguistic Annotation - No Easy Way Out! A Case from Bangla and Hindi POS Labeling Tasks. In *Proc. of the Third Linguistic Annotation Workshop*, Singapour. ACL.

Bart Desmet and Véronique Hoste. 2010. Towards a balanced named entity corpus for dutch. In *LREC*.

George Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, and Ralph Weischedel. 2004. The automatic content extraction (ACE) program tasks, data, and evaluation. In *Proc. of LREC*.

Maud Ehrmann. 2008. *Les entités nommées, de la linguistique au TAL : statut théorique et méthodes de désambiguïsation*. Ph.D. thesis, Univ. Paris 7 Diderot.

Michael Fleischman and Eduard Hovy. 2002. Fine grained classification of named entities. In *Proc. of COLING*, volume 1, pages 1–7. ACL.

Michael Fleischman. 2001. Automated subcategorization of named entities. In *Proc. of the ACL 2001 Student Research Workshop*, pages 25–30.

Karën Fort, Maud Ehrmann, and Adeline Nazarenko. 2009. Towards a Methodology for Named Entities Annotation. In *Proceeding of the 3rd ACL Linguistic Annotation Workshop (LAW III)*, Singapore.

Sylvain Galliano, Guillaume Gravier, and Laura Chaubard. 2009. The ESTER 2 evaluation campaign for the rich transcription of French radio broadcasts. In *Proc of Interspeech 2009*.

Ralph Grishman and Beth Sundheim. 1996. Message Understanding Conference - 6: A brief history. In *Proc. of COLING*, pages 466–471.

George Hripcsak and Adam S. Rothschild. 2005. Technical brief: Agreement, the f-measure, and reliability in information retrieval. *JAMIA*, 12(3):296–298.

Seungwoo Lee and Gary Geunbae Lee. 2005. Heuristic methods for reducing errors of geographic named entities learned by bootstrapping. In *IJCNLP*, pages 658–669.

Geoffrey Leech. 1997. Introducing corpus annotation. In Geoffrey Leech Roger Garside and Tony McEnery, editors, *Corpus annotation: Linguistic information from computer text corpora*, pages 1–18. Longman, London.

Diana Maynard, Valentin Tablan, Cristian Ursu, Hamish Cunningham, and Yorick Wilks. 2001. Named entity recognition from diverse text types. In *Recent Advances in NLP 2001 Conference, Tzigov Chark*.

Tomoko Ohta. 2002. The genia corpus: An annotated research abstract corpus in molecular biology domain. In *Proc. of HLTC*, pages 73–77.

Sophie Rosset, Olivier Galibert, Gilles Adda, and Eric Bilinski. 2007. The LIMSI participation to the QAst track. In *Working Notes for the CLEF 2007 Workshop*, Budapest, Hungary.

Sophie Rosset, Cyril Grouin, and Pierre Zweigenbaum. 2010. Entités nommées : guide d'annotation Quaero, November. T3.2, presse écrite et orale.

SAIC. 1998. Proceedings of the seventh message understanding conference (MUC-7).

William A Scott. 1955. Reliability of Content Analysis: The Case of Nominal Scale Coding. *Public Opinion Quaterly*, 19(3):321–325.

Satoshi Sekine. 2004. Definition, dictionaries and tagger of extended named entity hierarchy. In *Proc. of LREC*.

Anne Vilnat, Patrick Paroubek, Eric Villemonte de la Clergerie, Gil Francopoulo, and Marie-Laure Guénot. 2010. Passage syntactic representation: a minimal common ground for evaluation. In *Proc. of LREC*.

Alex Yeh, Alex Morgan, Marc Colosimo, and Lynette Hirschman. 2005. BioCreAtIvE task 1A: gene mention finding evaluation. *BMC Bioinformatics*, 6(1).