Annotation of specialized corpora using a comprehensive entity and relation scheme

Louise Deléger* **Anne-Laure Ligozat*** ** **Cyril Grouin***
Pierre Zweigenbaum* **Aurélie Névéol***
*CNRS, UPR 3251, LIMSI, 91403 Orsay, France
** ENSIIE, 91000 Evry, France
firstname.lastname@limsi.fr

Abstract

Annotated corpora are essential resources for many applications in Natural Language Processing. They provide insight on the linguistic and semantic characteristics of the genre and domain covered, and can be used for the training and evaluation of automatic tools. In the biomedical domain, annotated corpora of English texts have become available for several genres and subfields. However, very few similar resources are available for languages other than English. In this paper we present an effort to produce a high-quality corpus of clinical documents in French, annotated with a comprehensive scheme of entities and relations. We present the annotation scheme as well as the results of a pilot annotation study covering 35 clinical documents in a variety of subfields and genres. We show that high inter-annotator agreement can be achieved using a complex annotation scheme.

Keywords: Annotation; Clinical Texts; Natural Language Processing

1. Introduction

Annotated corpora are essential resources for many applications in Natural Language Processing. They provide insight on the linguistic and semantic characteristics of the genre and domain covered, and can be used for the training and evaluation of automatic tools. In the biomedical domain, annotated corpora have become available for several genres and subfields. Several efforts addressed the development of annotated corpora for English free-text, covering both the biomedical literature (Kim et al., 2003; Bada et al., 2012; Doğan and Lu, 2012; Ohta et al., 2012) and clinical narrative (South et al., 2009; Uzuner et al., 2010; Deléger et al., 2012). Annotation tasks sought to cover grammatical characteristics (Smith et al., 2004), complex linguistic phenomena (Chapman et al., 2012) as well as biological or clinical phenomena (Doğan and Lu, 2012; Ohta et al., 2012) that may be described in domain knowledge bases (Bada et al., 2012) as they occurred in natural language text. However, very few similar resources are available for languages other than English. Furthermore, annotation efforts are often focused on one particular linguistic phenomenon or biological entity of interest.

In this paper we present an effort to produce a high-quality corpus of clinical documents in French, annotated with a comprehensive scheme of entities and relations. We present the annotation scheme as well as the results of a pilot annotation study covering 35 clinical documents in a variety of subfields. We show that high inter-annotator agreement can be achieved using a complex annotation scheme.

2. Material and Methods

2.1. Two clinical corpora

Presentation. The clinical documents used in this study were drawn from two sources: foetopathology case reports (referred to as *Foetopath* thereafter) from a large French city hospital and electronic health records from a different set of French hospitals (referred to as *EHR* thereafter). While documents in the first corpus are similar in terms of structure and medical content, documents in the second corpus exhibit a large variety of documents types and cover several medical specialties.

Preprocessing. All documents have been de-identified and manually reviewed to ensure that all protected information was removed. Surrogate information was generated to replace the de-identified elements. The annotation strategy built on previous findings that pre-annotations can increase inter-annotator agreement and reduce annotation time (Névéol et al., 2011). Therefore, we produced a pre-annotated version of the corpora for entities only. Automatic entity pre-annotations were supplied to the annotators using an exact match strategy based on a French UMLS dictionary and a lexicon derived from a small set of documents annotated in the preliminary stage of the project (5 documents, selected from both corpora).

2.2. Methods

2.2.1. Annotation Protocol

The annotation scheme used in this work was designed to provide a broad coverage of the clinical domain, in order to allow for the annotation of medical events of interest mentioned in the clinical documents. We used the open source Brat Rapid Annotation Tool (BRAT) (Stenetorp et al., 2012), which supports complex annotation schemes for entities and relations and allows the use of pre-annotations. Four annotators—the authors of this paper (ALL, AN, CG, LD)—participated in the annotation task. All of them had previous annotation experience.

Figure 1 shows the overall annotation process. It was conducted in two phases: (1) a preliminary annotation phase during guideline design; (2) a pilot annotation phase once the guidelines were stabilized. During the preliminary phase, a small sample of 5 documents (3 Foetopath, 2 EHR) was selected to be annotated by all four annotators. This sample was annotated for both entities and relations after a first draft of the annotation guidelines was written (first

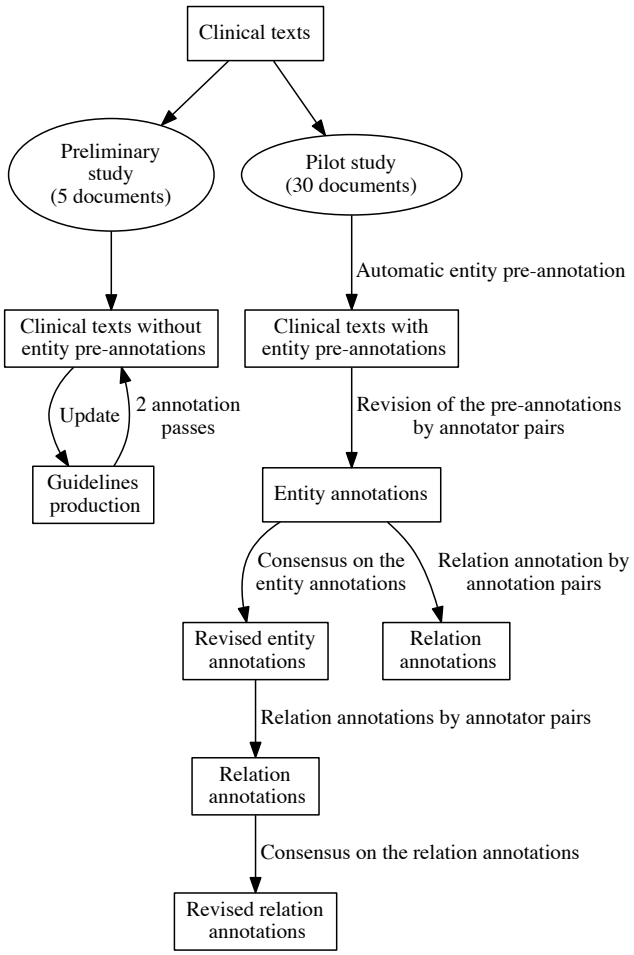

Figure 1: Overall annotation process

pass). Then annotators met to discuss issues and problems they encountered during the first annotation pass. The guidelines were extensively clarified and modified accordingly. Subsequently, annotators individually revised their annotations according to the modified guidelines (second pass). A consensus session was then held to resolve annotation disagreements and further clarify the guidelines where necessary. Finally, the annotators used the entity annotations resulting from the consensus to annotate relations. This allowed us to measure Inter-Annotator Agreement (IAA) on relations, without the influence of disagreements on entity annotations. Agreement was measured between each annotator pair for entities and relations after the first and second pass. Additionally, for relations, IAA was also measured after the consensus on entities.

After the preliminary phase, we considered the guidelines to be stable enough to conduct a pilot annotation phase with 15 documents from each source (Foetopath and EHR). One lead annotator (LD) annotated the entire pilot sample while the other three (ALL, AN, CG) each worked on one third of the documents. As a result, each document in the corpus was independently annotated by two annotators. As shown on Figure 1, the pilot annotation was conducted as follows: documents were pre-annotated automatically for entities. Annotators revised the pre-annotations to create annotations for entities and relations. Consensus sessions were held between annotators to resolve disagreements on entities, and a second pass of relation annotations was carried out using the consensus. The lead annotator ensured consistency on decisions that were made for cases that were not initially covered in the guidelines. The guidelines were updated accordingly. Inter-annotator agreement was measured between each annotator pair once for entities and twice for relations, that is (a) before resolving disagreements on entities and (b) after resolving disagreements on entities.

2.2.2. Annotation Scheme

Entities. The annotation scheme for entities was derived in part from the Unified Medical Language System® (UMLS®) Semantic Groups, described in (McCray et al., 2001) and (Bodenreider and McCray, 2003), but additional categories were created to address the need for fine-grained annotation of elements of clinical interest, such as the details of medications prescribed to a patient (Savova et al., 2012). For entity annotations, the annotators revised the pre-annotation using tools providing access to the UMLS in French[1] and in English[2]. When available, the UMLS Semantic Type of a concept was used to determine which entity type to assign to an annotated mention.

Relations. The annotation scheme for relations was derived in part from the UMLS Semantic Network. It also drew on previous annotation work for clinical documents, including (Savova et al., 2012).

3. Results

3.1. Description of the annotation scheme and annotated corpora

The final annotation scheme used comprises 19 entities (listed and defined in Table 9) and 18 relations (listed and defined in Table 10). Table 1 provides an overview of the number of consensus annotations in each corpus. It shows that the density of annotations is quite high in both corpora (on average, 60 annotations per 100 tokens), but that the distribution of entity types differs between the Foetopath vs. EHR corpus. For instance, Anatomy and Measurement entities are more frequent in the Foetopath corpus, while drug entities are more prevalent in the EHR corpus (see Table 1). Figure 2 presents a snippet of annotated text from the Foetopath corpus.

3.2. Inter-annotator agreement

Inter-annotator agreement was assessed using F-measure (Hripcsak and Rothschild, 2005), computed with a tool developed by the National Information and Communication Technology Research Center of Australia (Verspoor et al., 2013). Agreement scores during the preliminary phase are displayed in Table 2 for entities and Table 3 for relations. Agreement is low on the first pass for entities (mean of 0.502) and very low for relations (mean of 0.153), although it should be noted that all disagreements on entities have an impact on relations. IAA is higher on the second pass

[1]Portail Terminologique de Santé (PTS) `http://pts.chu-rouen.fr/`

[2]UMLS Terminology Services (UTS Metathesurus Browser) `https://uts.nlm.nih.gov/`

		Preliminary sample (N=5)	Foetopath (N=15)	EHR (N=15)	All (N=35)
General	Tokens	1605	4240	3976	9821
	Annotated tokens	814	2961	2052	5827
	Annotated entities	454	1924	1168	3546
	Annotated relations	270	1031	495	1796
Entities	Anatomy	104	787	138	1029
	Measurement	103	486	116	705
	Disorder	77	199	106	382
	Concept_Idea	29	157	20	206
	MedicalProcedure	34	117	169	320
	BiologicalProcessOrFunction	13	44	18	75
	ModalityAnchor	16	37	34	87
	LivingBeings	31	35	182	248
	Duration	13	30	16	59
	Chemicals_Drugs	8	12	71	91
	SignOrSymptom	8	7	43	58
	Genes_Proteins	2	7	4	13
	Date	13	6	87	106
	Frequency	2	0	36	38
	Devices	1	0	44	45
	Dosage	0	0	52	52
	Strength	0	0	19	19
	DrugForm	0	0	7	7
	AdministrationRoute	0	0	6	6
Relations	Measure_of	102	526	92	720
	Location_of	70	376	93	539
	Co-occurs_with	29	20	14	63
	Time_of	17	29	42	88
	Experiences	18	16	33	67
	Reveals	9	23	18	50
	Negation	7	21	15	43
	History	6	9	5	20
	Hypothetical	3	5	9	17
	Treats	4	4	26	34
	Complicates	2	1	2	5
	Precedes	2	1	20	23
	Causes	1	0	2	3
	HasAdministrationRoute	0	0	6	6
	HasDosage	0	0	53	53
	HasDrugForm	0	0	7	7
	HasDuration	0	0	4	4
	HasFrequence	0	0	35	35
	HasStrength	0	0	19	19

Table 1: Descriptive statistics (count of tokens and consensus annotations (overall and per type) in each corpus

(revision according to improved guidelines), but can be improved. Agreement on relations after revising the guidelines and reaching a consensus on entities is very good (mean of 0.817).

Table 4 and Table 5 show IAA on the Foetopath and EHR corpora during the pilot phase after a stable version of the guidelines had been produced. On the Foetopath corpus, agreement is very good for entities (mean of 0.817) and substantially higher than on the preliminary corpus (at most 0.604, see Table 2). Agreement on relations is also higher than on the preliminary corpus (0.599 vs. at most 0.299 before entity consensus, see Table 3 and 0.890 vs. 0.817 after entity consensus). On the EHR corpus, agreement on entities is fair (mean of 0.679) and also higher than during the preliminary phase. It is good on relations after entity consensus (mean of 0.773) but slightly lower than during the preliminary phase, unlike on the Foetopath corpus. However, agreement remains low when relations are annotated before reaching a consensus on entities (mean of 0.599 and 0.413). Agreement is high when relations are annotated after resolving disagreements (0.890 and 0.779).

Table 6 details inter-annotator agreement (mean) results for each entity type for the Foetopath and EHR corpora (pilot annotation phase). Agreement is very high (above 0.85) on

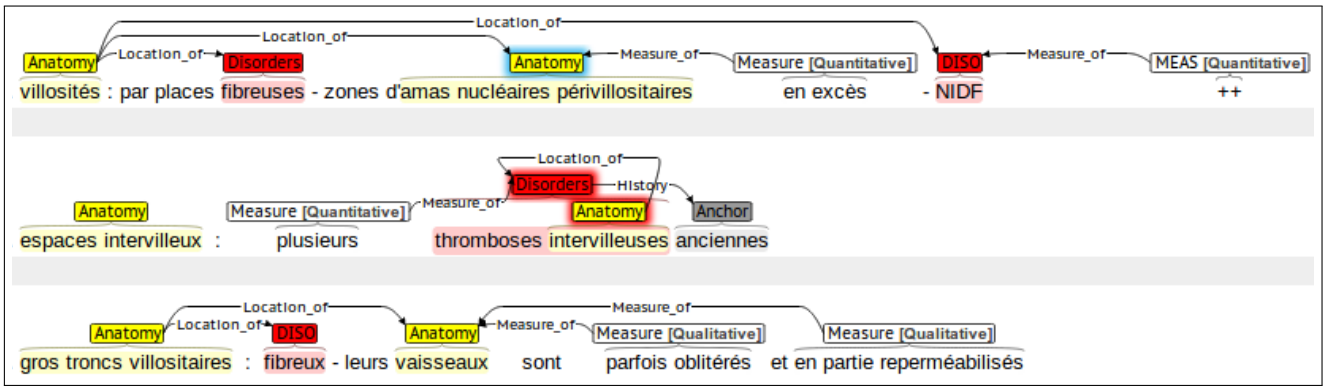

Figure 2: Snippet of annotated text (Foetopath)

	1st pass	2nd pass
AL/AN	0.546	0.670
AL/CG	0.376	0.574
AL/LD	0.496	0.589
AN/CG	0.496	0.589
AN/LD	0.627	0.657
CG/LD	0.468	0.545
Mean	**0.502**	**0.604**

Table 2: Overall inter-annotator agreement for entities during the preliminary phase (1st/2nd annotation passes)

	No consensus		Consensus
	1st pass	2nd pass	
AL/AN	0.420	0.344	0.860
AL/CG	0.048	0.178	0.783
AL/LD	0.207	0.255	0.852
AN/CG	0.062	0.285	0.743
AN/LD	0.255	0.396	0.868
CG/LD	0	0.338	0.796
Mean	**0.153**	**0.299**	**0.817**

Table 3: Overall inter-annotator agreement for relations during the preliminary phase (1st and 2nd annotation passes, and after the consensus on entities)

Anatomy, Procedure, Duration and Chemicals_Drugs entities in the Foetopath corpus. Annotators had more trouble with Disorder and BiologicalProcessOrFunction entities (agreement around 0.65, the lowest for this corpus). We do not take into account the zero agreement values for the SignOrSymptom and Devices entities. Because of the

	Foetopath	EHR
AL/LD	0.840	0.697
AN/LD	0.833	0.700
CG/LD	0.777	0.631
Mean	**0.817**	**0.676**

Table 4: Overall inter-annotator agreement for entities during the pilot annotation phase

	Foetopath		EHR	
	No consens.	Consensus	No consens.	Consensus
AL/LD	0.632	0.931	0.384	0.843
AN/LD	0.577	0.917	0.590	0.830
CG/LD	0.589	0.822	0.266	0.645
Mean	**0.599**	**0.890**	**0.413**	**0.773**

Table 5: Overall inter-annotator agreement for relations during the pilot annotation phase, annotated without resolving disagreements on entities (no consensus) or after resolving disagreements (consensus)

very small number of these entities (respectively 7 and 0), we cannot draw any significant conclusion for these two entity types. As mentioned before, agreement is generally lower on the EHR corpus. The highest agreement values (above 0.75) were on Chemicals_Drugs, Date and Medical-Procedure entities. The lowest values (below 0.45) were on Concept_Idea, BiologicalProcessOrFunction, Duration, Frequency, and Strength entities (although most of these have less than 20 occurrences). Similarly to the Foetopath corpus, agreement is only fairly good on Disorder (0.68).
Table 7 details inter-annotator agreement (mean) results for each relation type for the Foetopath and EHR corpora (pilot annotation phase, after entity consensus). Agreement is high (above 0.85) on the three most common relations (Measure_of, Location_of, Time_of), but very low on the co-occurs_with and precedes relations, in both corpora.

3.3. Pre-annotation performance

We evaluated the performance of the automatic pre-annotation against the consensus annotations. Results (Table 8) show good precision (0.844 overall), but rather low recall (0.567) for the Foetopath corpus, and average precision (0.678) and low recall (0.406) for the EHR corpus.

4. Discussion

Tables 2 to 5 show that the agreement can vary significantly between annotator pairs. This was also observed for an opinion categorization task (Osman et al., 2010).
Agreement also varies according to the type of corpus. We observed higher inter-annotator agreement on the Foetopath corpus than on the EHR corpus. This difference is

	Foetopath			EHR		
	Precision	**Recall**	**F-measure**	**Precision**	**Recall**	**F-measure**
Anatomy	0.936	0.750	0.833	0.724	0.399	0.514
Measurement	0.833	0.440	0.576	0.652	0.259	0.370
Disorder	0.560	0.283	0.376	0.551	0.462	0.503
Concept_Idea	0.943	0.529	0.678	0.500	0.050	0.091
MedicalProcedure	0.835	0.650	0.731	0.660	0.379	0.481
BiologicalProcess	0.317	0.296	0.306	0.625	0.556	0.588
ModalityAnchor	1	0.189	0.318	0.875	0.412	0.560
LivingBeings	0.750	0.257	0.383	0.556	0.517	0.536
Duration	0.727	0.800	0.762	0.429	0.188	0.261
Chemicals_Drugs	0.833	0.833	0.833	0.905	0.803	0.851
Genes_Proteins	0.667	0.571	0.615	0.167	0.250	0.200
SignOrSymptom	0.357	0.714	0.476	0.567	0.395	0.466
Date	0	0	0	0.973	0.828	0.894
Devices	0	0	0	0.539	0.159	0.246
AdministrationRoute	N/A	N/A	N/A	0	0	0
Dosage	N/A	N/A	N/A	0	0	0
DrugForm	N/A	N/A	N/A	0	0	0
Frequency	N/A	N/A	N/A	0	0	0
Strength	N/A	N/A	N/A	0	0	0
Overall	**0.844**	**0.567**	**0.678**	**0.673**	**0.406**	**0.506**

Table 8: Pre-annotation performance

	Foetopath	**EHR**
Anatomy	0.899	0.646
Measurement	0.779	0.680
Disorder	0.664	0.686
Concept_Idea	0.767	0.350
MedicalProcedure	0.862	0.762
BiologicalProcessOrFunction	0.631	0.321
ModalityAnchor	0.767	0.671
LivingBeings	0.712	0.727
Duration	0.867	0.429
Chemicals_Drugs	0.952	0.884
SignOrSymptom	0	0.582
Genes_Proteins	0.667	0.333
Date	0.778	0.935
Frequency	N/A	0.353
Devices	0	0.544
Dosage	N/A	0.532
Strength	N/A	0.256
DrugForm	N/A	0.333
AdministrationRoute	N/A	0.667

Table 6: Mean inter-annotator agreement for each entity type during the pilot annotation phase

	Foetopath	**EHR**
Measure_of	0.942	0.871
Location_of	0.861	0.926
Time_of	1.000	0.838
Reveals	0.628	0.842
Negation	0.941	0.914
co-occurs_with	0.334	0.217
Experiences	0.737	0.777
History	1.000	0.667
Hypothetical	1.000	0.886
Treats	1.000	0.507
precedes	0.333	0.280
Complicates	0.500	0.000
Causes	N/A	0.000
HasAdministrationRoute	N/A	0.650
HasDosage	N/A	0.667
HasDrugForm	N/A	0.667
HasDuration	N/A	0.500
HasFrequence	N/A	0.667
HasStrength	N/A	0.667

Table 7: Mean inter-annotator agreement for each relation type during the pilot annotation phase

due to the fact that the Foetopath corpus is composed of documents from a specific domain with very similar structure and content. The EHR corpus on the other hand includes several medical specialties and document types, and thus documents from this corpus exhibit more variation.

The level of agreement on relations varies according to the annotation strategy. Agreement is lower when relation annotation is performed at the same time as entity annotation, because disagreements on entities impact the selection of relations; specifically, annotators can select the same relation between two entities only if those entities have previously been annotated by both annotators. Agreement is much higher when relation annotation is performed separately from entity annotation, viz. on consensus entities ob-

tained after resolving entity disagreements (Tables 3 and 5). Agreement between annotators was substantially higher during the pilot annotation phase than during the preliminary annotation phase, for both entities and relations (with the exception of relations on the EHR corpus). This demonstrates that with sufficient training and adequately defined guidelines, high inter-annotator agreement can be achieved using a complex annotation scheme. However, inter-annotator agreement should be improved on the EHR corpus. Because of the higher variability of this corpus, a larger sample of documents need to be annotated before reaching a truly high agreement. All annotators found the entity pre-annotation useful for annotating the Foetopath corpus. They felt that existing annotations were often correct. While a number of additional entity annotations had to be created, few erroneous annotations had to be removed so that pre-annotations contributed to increase annotation speed. This is consistent with the performance evaluation which showed high precision of 0.844 (i.e., few spurious annotations) and low recall of 0.567 (i.e., some missing annotations). The benefit of the pre-annotation is more difficult to demontrate for the EHR corpus, due to the lower performance. It was most useful in the top performing categories, such as Dates and Chemical_Drugs. In future work, we will improve the pre-annotation system by using the annotated documents to train machine-learning algorithms.

5. Conclusion

The annotation results over the two study corpora showed that annotation with a complex entity and relation scheme is feasible. However, the annotation task is more successful (i.e. results in more consistent and higher quality annotations) if (a) relation annotations are created based on a consensus of entity annotations and (b) the corpus of documents used is focused on a limited number of genres/specialties. We plan to share the guidelines we defined for this study. Future work will address the annotation of additional documents and public release of the corpus.

6. Acknowledgements

This work was supported by the French National Agency for Research under grants CABeRneT[3] (ANR-13-JCJC) and Accordys[4] (ANR-12-CORD-007-03).

7. References

Bada, M., Eckert, M., Evans, D., Garcia, K., Shipley, K., Sitnikov, D., Baumgartner, W. A., Cohen, K. B., Verspoor, K., Blake, J. A., and Hunter, L. E. (2012). Concept annotation in the CRAFT corpus. *BMC Bioinformatics*, 13(161).

Bodenreider, O. and McCray, A. T. (2003). Exploring semantic groups through visual approaches. *J Biomed Inform*, 36(6):414–32.

Chapman, W. W., Savova, G. K., Zheng, J., Tharp, M., and Crowley, R. (2012). Anaphoric reference in clinical reports: characteristics of an annotated corpus. *J Biomed Inform*, 45(3):507–21.

Deléger, L., Li, Q., Lingren, T., Kaiser, M., Molnar, K., Stoutenborough, L., Kouril, M., Marsolo, K., and Solti, I. (2012). Building gold standard corpora for medical natural language processing tasks. In *AMIA Annu Symp Proc*, pages 144–53.

Doğan, R. I. and Lu, Z. (2012). An improved corpus of disease mentions in PubMed citations. In *Proc of BioNLP*, pages 91–9.

Hripcsak, G. and Rothschild, A. S. (2005). Agreement, the f-measure, and reliability in information retrieval. *J Am Med Inform Assoc*, 12(3):296–8.

Kim, J.-D., Ohta, T., Tateisi, Y., and Tsujii, J. (2003). GENIA corpus—a semantically annotated corpus for bio-textmining. *BMC Bioinformatics*, 19(Suppl 1):i180–2.

McCray, A. T., Burgun, A., and Bodenreider, O. (2001). Aggregating UMLS semantic types for reducing conceptual complexity. In *Proc of MedInfo*, volume 10, pages 216–20.

Névéol, A., Doğan, R. I., and Lu, Z. (2011). Semi-automatic semantic annotation of PubMed queries: a study on quality, efficiency, satisfaction. *J Biomed Inform*, 44(2):310–8.

Ohta, T., Pyysalo, S., Tsujii, J., and Ananiadou, S. (2012). Open-domain anatomical entity mention detection. In *Proc of Workshop on Detecting Structure in Scholarly Discourse*, pages 27–36.

Osman, D., Yearwood, J., and Vamplew, P. (2010). Automated opinion detection: Implications of the level of agreement between human raters. *Inf Process Manage*, 46(3):331–42.

Savova, G., Styler, W., Albright, D., Palmer, M., Harris, D., Zaramba, G., Haug, P., Clark, C., Wu, S., and Ihrke, D. (2012). SHARP template annotations: Guidelines. Technical report, Mayo Clinic.

Smith, L. H., Rindflesch, T. C., and Wilbur, W. J. (2004). MedPost: a part of speech tagger for biomedical text. *BMC Bioinformatics*, 20(14):2320–21.

South, B. R., Shen, S., Jones, M., Garvin, J., Samore, M. H., Chapman, W. W., and Gundlapalli, A. V. (2009). Developing a manually annotated clinical document corpus to identify phenotypic information for inflammatory bowel disease. *BMC Bioinformatics*, 10(Suppl 9):S12.

Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., and Tsujii, J. (2012). Brat: a web-based tool for NLP-assisted text annotation. In *Proc of EACL Demo Session*, pages 102–7.

Uzuner, O., Solti, I., and Cadag, E. (2010). Extracting medication information from clinical text. *J Am Med Inform Assoc*, 17(5):514–8.

Verspoor, K., Yepes, A. J., Cavedon, L., McIntosh, T., Herten-Crabb, A., Thomas, Z., and Plazzer, J.-P. (2013). Annotating the biomedical literature for the human variome. *Database*.

[3]CABeRneT: Compréhension Automatique de Textes Biomédicaux pour la Recherche Translationnelle

[4]Accordys: Agrégation de Contenus et de COnnaissances pour Raisonner à partir de cas de DYSmorphologie fœtale

Entity type	UMLS Semantic Type (or definition if none)	Examples
Anatomy	Anatomical Structure, Body Location or Region, Body Part Organ or Organ Component, Body Space or Junction, Body Substance, Body System, Cell, Cell Component, Embryonic Structure, Fully Formed Anatomical Structure, Tissue	*foot; right femoral artery*
Disorder	Acquired Abnormality; Anatomical Abnormality; Cell or Molecular Dysfunction; Congenital Abnormality; Disease or Syndrome; Experimental Model of Disease; Injury or Poisoning; Mental or Behavioral Dysfunction; Pathologic Function; Neoplastic Process	*diabetes; myocardial infarction*
SignOrSymptom	Sign or Symptom	*pain; cough*
Devices	Devices	*insulin pomp; pacemaker*
Concept_Idea	Classification, Conceptual Entity, Functional Concept, Group Attribute, Idea or Concept, Intellectual Product, Language, Qualitative Concept, Quantitative Concept, Regulation or Law, Spatial Concept, Temporal Concept	*weight; length*
MedicalProcedure	Diagnostic Procedures; Health Care Activity; Laboratory Procedure; Therapeutic or Preventive Procedure	*angiography, psychiatric consult*
BiologicalProcessOrFunction	Biologic Function; Cell Function; Genetic Function; Molecular Function; Natural Phenomenon or Process; Organ or Tissue Function; Organism Function; Physiologic Function	*transit*
LivingBeings	Alga; Amphibian; Animal; Archeon; Bacterium; Bird; Family Group; Fish; Fungus; Human; Invertebrate; Mammal; Organism; Patient or Disabled Group; Plant; Population Group; Professional or Occupation Group; Reptile; Rickettsia or Chlamydia;Vertebrate; Virus	*patient; salmonella*
Chemicals_Drugs	Antibiotic; Biomedical or Dental Material; Carbohydrates; Chemical; Chemical Viewed Functionally; Chemical Viewed Structurally; Clinical Drug; Hazardous or Poisonous Substance; Inorganic Chemical; Pharmacological Substance; Vitamin	*insulin; steroids; Percocet*
Genes_Proteins	Amino Acid, Peptide or Protein; Enzyme, Lipid; Immunologic Factor; Indicator, Reagent, or Diagnostic Aid; Gene or Genome; Nucleic Acid, Nucleoside or Nucleotide; Receptor	*PTX1; fibrin*
Measurement	A figure, extent, or amount obtained by measuring or observing. Measurement also include subjective qualifications of the shape, color, or other attributes of measured entities	*3 cm; normal*
Date	The time at which an event occurs	*in 1981; 02/01/2013; today*
Duration	The time during which something exists or lasts	*for two weeks*
Frequency	The number of repetitions of a periodic process in a unit of time	*twice a day, every morning*
AdministrationRoute	Route or method of administering the medication	*oral; IV*
Dosage	How many of each drug the patient is taking	*3 tablets; two puffs*
DrugForm	Form of the medication	*tablet; cream*
Strength	Strength number and unit of the prescribed drug	*10 mg; 5 mg/ml*
ModalityAnchor	A phrase or text span that provides motivation for assigning a given modality (either negation, hypothetical, or history) to an entity	*no; suspected; history of*

Table 9: Annotation scheme for entities

Relation	Definition	Involved entities
Location_of	The position, site, or region of an entity or the site of a process	*Anatomy* Location_of *Anatomy* *Anatomy* Location_of *Disorder* *Anatomy* Location_of *MedicalProcedure*
Time_of	The moment a phenomenon or procedure occurred; the length of time a phenomenon or procedure lasted.	*Date*|*Duration* Time_of *Concept_Idea* *Date*|*Duration* Time_of *Disorder* *Date*|*Duration* Time_of *SignOrSymptom* *Date*|*Duration* Time_of *MedicalProcedure* *Date*|*Duration* Time_of *Chemicals_Drugs*
Treats	Applies a remedy with the object of effecting a cure or managing a condition	*Chemicals_Drugs* Treats *Disorder* *MedicalProcedure* Treats *Disorder*
Complicates	Causes to become more severe or complex or results in adverse effects	*Disorder* Complicates *Disorder* *Chemicals_Drugs* Complicates *Disorder* *MedicalProcedure* Complicates *Disorder*
Measure_of	The quantitative or qualitative result of a medical procedure such as lab test or physical examination	*Measurement* Measure_of *Concept_Idea* *Measurement* Measure_of *Anatomy* *Measurement* Measure_of *Process* *Measurement* Measure_of *Disorder* *Measurement* Measure_of *SignSymptom*
Interacts_with	Acts, functions, or operates together with	*Chemicals_Drugs* Interacts_with *Chemicals_Drugs*
Co-occurs_with	Occurs at the same time as, together with, or jointly. This includes is co- incident with, is concurrent with, is contemporaneous with, accompanies, coexists with, and is concomitant with	*Disorder*|*SignSymptom* Co-occurs_with *Disorder*|*SignSymptom*
Precedes	Occurs earlier in time. This includes antedates, comes before, is in advance of, predates, and is prior to	*Disorder*|*SignSymptom* Precedes *Disorder*|*SignSymptom* *MedicalProcedure* Precedes *MedicalProcedure*
Reveals	When a test is conducted and the outcome is known/leads to a diagnosis	*MedicalProcedure*|*SignSymptom* Reveals *Disorder*
Conducted	When a test is conducted to investigate a Disorder and the outcome is unknown/does not result in a diagnosis	*MedicalProcedure* Conducted *Disorder*
Causes	Brings about a condition or an effect. Implied here is that an agent, such as for example, a pharmacologic substance or an organism, has brought about the effect. This includes induces, effects, evokes, and etiology	*LivingBeings* Causes *Disorder* *Chemicals_Drugs* Causes *Disorder*
Experiences	When a Living Being (e.g. patient) is affected by a Disorder, Sign or Symptom; when a Living Being (e.g. patient) is subjected to a Medical Procedure	*LivingBeings* Experiences *Disorder* *LivingBeings* Experiences *SignSymptom* *LivingBeings* Experiences *MedicalProcedure*
HasAdministrationRoute	links a medication to its administration route	*Chemicals_Drugs* HasAdministrationRoute *AdministrationRoute*
HasDosage	links a medication to its dosage	*Chemicals_Drugs* HasDosage *Dosage*
HasStrength	links a medication to its strength	*Chemicals_Drugs* HasStrength *Strength*
HasFrequence	links a medication to its frequency	*Chemicals_Drugs* HasFrequence *Frequency*
HasDuration	links a medication to its duration	*Chemicals_Drugs* HasDuration *Duration*
HasDrugForm	links a medication to its form	*Chemicals_Drugs* HasDrugForm *DrugForm*

Table 10: Annotation scheme for relations