Annotating Geographical Entities

Alexandru Sălăvăstru, Daniela Gîfu

Faculty of Computer Science, "Alexandru Ioan Cuza" University of Iaşi
{alexandru.salavastru,daniela.gifu}@info.uaic.ro

Abstract. This paper describes a study based on exploration of relations between geographical entities. We suggested a new tool for training and evaluation required by related annotation experiments. It relates to an annotator used for semi-automatic annotation, starting with the geography manual. We define fifteen types of entities: *location*, *geo_position, geology, landform, clime, water, dimension, person*, *organization*, *URL, Timex, resource*, *industry, cultural, unknown* with their specific subtypes. Moreover, we present the annotation conventions for three semantic relations: *referential*, *structural* and *spatial*, considered to be optimal operators in understanding a geographical manual. A part of the annotation is done manually, while the other part is done automatically, such as the token, lemma, part-of-speech. The study is intended to create a tool for the automatic detection of semantic relations in texts on geographic issues such as geography manuals, travel guides, geography atlases, etc., in order to help children, professors, guides, PR specialists and to be useful for tourists, generally to discover the complexity and the beauty of the nature.

Keywords: geography manual, entities, annotation conventions, semantic relation, annotator

1 Introduction

Starting with the NAACL 2003 Workshop on the Analysis of Geographic References[1], a new community of NLP researchers and engineers focused on different aspects of the geographic text analysis task. The motivation for our study relies on the need for objectivity in the interpretation of semantic relations between geographical entities. We present a new tool, called RelAnn (*Relations Annotator*) used for annotation and semi-automatic extraction of entities and for manual annotation of semantic relations on corpora on geographical topics. We establish annotation conventions for the semantic relations that will be the ground of defining patterns which contain information at lexical and syntactic level for the automatic recognition of those relations from all kinds of geographical texts (geography manuals, tourist guides, atlases) and to extract their functionality (e.g. natural resources from Apuseni Mountains).

Our corpus is a Romanian Geography manual containing about 160 pages of geographical entities and semantic relations between two arguments (entities). Moreover,

---

[1] http://www.kornai.com/NAACL/WS9/orig.html

we create a Gold Corpus for Romanian language based on geographical issues. The annotation process was a long one, preceded by the following modules: POS tagger [17], NP-Chunking [18], NER (*Name Entity Recognizer*) [5] and RARE (*Robust Anaphora Resolution Engine*) [6].

The paper is structured as follows: Section 2 briefly describes the background related to annotating entities and semantic relations, Section 3 discusses the annotation process using the RelAnn tool, Section 4 presents some results and statistics interpretation and finally, Section 5 depicts some conclusions and directions for future work.

2   Background

RelAnn is an annotator inspired by RECON [9] for automatic entity recognition. ANNIE tool, included in GATE [4] is well-known for this task. Until now, we used PALinkA[2] for annotating corpora in several similar projects, for purposes including: anaphoric and co-referential links in a parallel French-English corpus, summarization, different versions of the Centering Theory, co-references in email messages and web pages, or for Romanian entities. A relevant example is the annotation of the "Quo Vadis" novel [3].

The most common annotation tools are the web-based ones. One of them, called BRAT, is an annotation tool used in many research papers, which aims at the extraction of biomedical events [7], co-references in scientific papers [14], entity annotation on medical corpus [15], etc.

Another tool, WORDFREAK [11], is used in active-learning for human correction of automatically annotated data. Others like CAT, CELCT Annotation Tool[3] [1], suggest a multi-layer annotation concept, and give annotators the ability to create labels and relations with multiple properties such as font size, color, type of relation, and so on.

In order to avoid issues such as local storage, cross-platform deployment, we found Anaphora tool [2], where annotations are divided into *Entity* and *Relation* types, both stored in XML files. This methodology is similar to what we suggest in our study. Similar to Anaphora is eHost/ChartReader [19] made for multiple annotators sharing the same text, but also relies on a remote installation of ChartReader. It offers only basic relations between annotations (primarily for co-reference) and stores them in XML files.

3   Annotation process. Tools and conventions

In this section we will describe the entire annotation process including the annotation tool corpora and the annotation methodology, with many similarities found in SpaceML[10].

---

[2] PALinkA was created by Constantin Orăşan in the Research Group in Computational Linguistics, at the School of Law, Social Sciences and Communications, Wolverhampton.

[3] CELCT Annotation Tool developed by Centre for the Evaluation of Language and Communication Technologies (CELCT) - http://www.celct.it/.

3.1 RelAnn Tool

As mentioned before, RelAnn (Fig. 1.) was inspired by the RECON tool, but there are major differences between them. RECON gives the possibility to create `n-ary` relations and marks-up text spans without any relation definition. It allows annotators to create long-chain co-reference with import / export for XML standoff format. RelAnn is a new semi-automatic tool that creates and annotates entities and semantic relations. The main course was to create a user-friendly, easy-to-use application, but also with fewer restrictions as possible and so far it seems very efficient in doing our tasks.

First, we check the recognized entities. Then we define relations with argument range and add types of entities. Our tool is not restrictive to predefined features, but gives the possibility to create any kind of relations and add different types of entities with their particular color. At each step, we identify in the text entities that have semantic relations between them, a trigger that signals the relation, a direction from one argument to another and save them as a RELATION entry in the XML file. This tool can be used to annotate entities from different domains, such as: *biology*, *computer science*, *literature*, *astronomy*, *physics* and so on. Another important feature is that each text has relations and entities stored separately, and after upload, it shows your progress on that particular file.
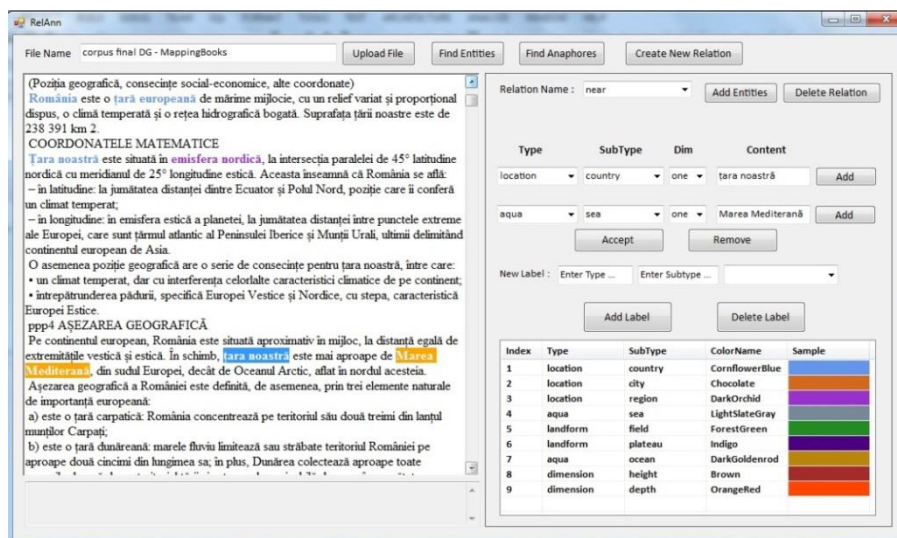

Fig. 1. RelAnn interface working session

3.2 Corpora

While preparing the preliminary conclusions in the configuration model, we decided to include in our corpora a geography manual [12] containing 160 pages and almost 37.000 tokens. At first, the text was edited in PDF, so we applied the boiling-plate technology to obtain raw text in txt format and then we made corrections to the raw

text. We intend to enlarge our corpora with numerous texts from the geographical field and thus improving discovering rates and information extraction patterns.

3.3 Methodology

This work is based on a set of principles for relation inventories found in Vivi Năstase's book [12], which cites Levi [8] and Ó Séaghdha and Copestake [16]. We followed some guidelines such as: inventory relations should give a good coverage, provide useful semantic information, and entity classes should be well defined, with no overlapping.

The research consists in: pre-processing the Corpus; annotating entities; annotating semantic relations; and evaluation.

3.3.1 Pre-processing the Corpus

The Romanian automatic pre-processing chain applied on raw texts of the book consists of the following tasks, executed in sequence:
- Segmentation (splitting the text in sentences).
- Tokenization (demarcates words or word compounds, but also numbers, punctuation marks, abbreviations, etc.).
- Lemmatization (determines lemmas of words).
- Part-of-speech tagging (identifies POS categories and morpho-syntactic information of tokens).

This is a part from the sentence segmentation annotation in XML standoff format:

```
....
<S ID="s9" offsetStart="1441" offsetStop="1492"/>
....
```

The tokenization, lemmatization and POS tagging are realized in one step.

For instance: *România ... o țară europeană* / (EN) *Romania ... an European country*.

```
<W  Case="direct"  Definiteness="yes"  Gender="feminine"
ID="w10.134"     LEMMA="românie"     MSD="Ncfsry"     Num-
ber="singular"     POS="NOUN"     Type="common"     off-
setStart="2256" offsetStop="2263" text="România"/>
...
<W    Case="direct"    Gender="feminine"    ID="w10.136"
LEMMA="un"  MSD="Tifsr"  Number="singular"  POS="ARTICLE"
Type="indefinite"   offsetStart="2269"   offsetStop="2270"
text="o"/>
<W  Case="direct"  Definiteness="no"  Gender="feminine"
ID="w10.137" LEMMA="țară" MSD="Ncfsrn" Number="singular"
POS="NOUN"     Type="common"     offsetStart="2271"     off-
setStop="2275" text="țară"/>
```

```
<W  Case="direct"  Definiteness="no"  Gender="feminine"
ID="w10.138"  LEMMA="european"  MSD="Afpfsrn"  Num-
ber="singular"  POS="ADJECTIVE"  offsetStart="2276"  off-
setStop="2285" text="europeană"/>
```

  – Noun phrase chunking [16] (recognizing the chunks that consist of noun phrases (NPs)). We provide the word POS tagging information to observe the id references between NP chunks and words that form that chunk.

  Ex: *Suprafaţa ţării noastre.* / (EN) *Our country area.*

```
<W  Case="direct"  Definiteness="yes"  Gender="feminine"
ID="w11.1"  LEMMA="suprafaţă"  MSD="Ncfsry"  Num-
ber="singular"  POS="NOUN"  Type="common"  off-
setStart="2400" offsetStop="2409" text="Suprafaţa"/>
<W  Case="oblique"  Definiteness="yes"  Gender="feminine"
ID="w11.2"  LEMMA="ţară"  MSD="Ncfsoy"  Number="singular"
POS="NOUN"  Type="common"  offsetStart="2410"  off-
setStop="2415" text="ţării"/>
<W  Case="oblique"  Gender="feminine"  ID="w11.3"
LEMMA="meu"  MSD="Ds1fsop"  Number="singular"
POS="DETERMINER" Person="first" Possessor_number="plural"
Type="possessive"  offsetStart="2416"  offsetStop="2423"
text="noastre"/>
<NP HEADID="11.1" ID="n106" WORDSID="w11.1, w11.2" off-
setStart="2400" offsetStop="2415"/>
<NP  HEADID="11.2"  ID="n107"  WORDSID="w11.2"  off-
setStart="2410" offsetStop="2415"/>
```

  – Named Entity Recognizer (NER – identifies and classifies the entities):

```
<S ID="s10" offsetStart="1493" offsetStop="2399"/>
<W  Case="direct"  Definiteness="yes"  Gender="feminine"
ID="w10.134"  LEMMA="românie"  MSD="Ncfsry"  Num-
ber="singular"  POS="NOUN"  Type="common"  off-
setStart="2256" offsetStop="2263" text="România"/>
<W  EXTRA="intranzitiv"  ID="w10.135"  LEMMA="fi"
MSD="Vmip3s"  Mood="indicative"  Number="singular"
POS="VERB"  Person="third"  Tense="present"
Type="predicative" offsetStart="2264" offsetStop="2268"
text="este"/>
<W  Case="direct"  Gender="feminine"  ID="w10.136"
LEMMA="un" MSD="Tifsr" Number="singular" POS="ARTICLE"
Type="indefinite"  offsetStart="2269"  offsetStop="2270"
text="o"/>
<W  Case="direct"  Definiteness="no"  Gender="feminine"
ID="w10.137" LEMMA="ţară" MSD="Ncfsrn" Number="singular"
```
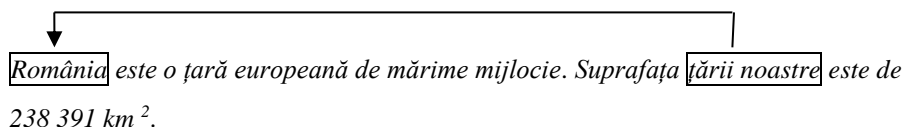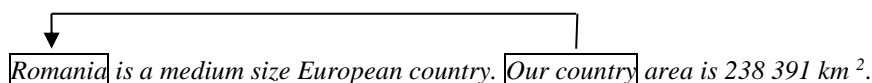
```
POS="NOUN"    Type="common"    offsetStart="2271"    off-
setStop="2275" text="ţ ară"/>
<W   Case="direct"   Definiteness="no"   Gender="feminine"
ID="w10.138"    LEMMA="european"    MSD="Afpfsrn"    Num-
ber="singular"   POS="ADJECTIVE"   offsetStart="2276"   off-
setStop="2285" text="europeană"/>
<ENTITY    ID="e1"    TYPE="location"    SUBTYPE="country"
DIMENSION="one"    WORDSID="w10.134,"    offsetStart="2256"
offsetStop="2263" Color="Chocolate" />
<ENTITY    ID="e2"    TYPE="location"    SUBTYPE="country"
DIMENSION="one"    WORDSID="w10.137,    w10.138"    off-
setStart="2271" offsetStop="2285" Color="Chocolate"/>
```

   − Anaphora Resolution (RARE - extract co-reference chains). For instance:

România este o ţară europeană de mărime mijlocie. Suprafaţa *ţării noastre* este de

238 391 km [2].

Romania is a medium size European country. *Our country* area is 238 391 km [2].

3.3.2 Annotating entities

Our intention was to markup entities (15 types and 89 subtypes)[4] as mentioned in a geography manual (see Table 1). Below we mention all entity types with one subtype examples for each. Let's note that we have too many entity subtypes in order to illustrate all.

Table 1. Entity classification with examples

Type	Subtype	Example (RO/EN)
location	county seat	Iași
geo_position	cardinal direction	Est / East
geology	rock	Granite / granit
landform	hill	Dealurile de Vest / Western Hills
clime	climate type	temperate
water	river	Olt / Olt
dimension	latitude	45º N

[4] Let's note that at this moment we have a too large range of entities, the statistical data clearly showing that we have to rethink the classification of entity subtypes.

person	person	Simion Mehedinți
organization	organization	O.N.U
url	-	www.geografie.ro
timex	-	Iulie 2004 / July 2004
resource	coal basin	bazinul Petroşani / basin Petroşani
industry	thermopower	Termocentrala Borzeşti / Thermal power plant Borzesti
cultural	palace	Peleş
unknown	-	-

3.3.3 Annotating semantic relations

In this paper we introduce three types of semantic relations with their particular sub-types. Each relation holds between two arguments, called *poles*. The part that signals the type of relation, which can be one word or expression, is called *trigger*. Our notations are expressed in XML. We use <S></S> for marking sentences with attributes id and start/stop offset, and <W></W> for words with attributes like ids, lemmas, morphosyntatic analysis, start/stop offset and text. Also, the <NP></NP> annotation for NP chunks with specific ids and references to words that form them and start/stop offset, and <ENTITY></ENTITY> to mark entities with attributes like id, type, subtype, reference to word, colour and start/stop offset. Finally we added to the file the <RELATION></RELATION> with id, type, subtype, from, to and trigger attributes. The XML standoff format is easier to make any modifications to the file without changing the structure.

In the following we give examples for each type of relation, and an XML for one of them.

I. *Referential relations* with two subtypes are listed and illustrated below:

- *coref* (anaphora).
1:[Romania]... 2:[ţării noastre]... / (EN) (1:[Romania]... 2:[our country]...) ⇒ [2] coref [1];
- *isa* (a relation from element to class (concept)).
1:[România] este o 2:[ţară europeană] / (EN) (1:[Romania] is a 2:[European country]). ⇒ [1] isa [2];
```
<S ID="s10" offsetStart="1493" offsetStop="2399"/>
<W  Case="direct"  Definiteness="yes"  Gender="feminine"
ID="w10.134"    LEMMA="românie"    MSD="Ncfsry"    Num-
ber="singular"    POS="NOUN"    Type="common"    off-
setStart="2256" offsetStop="2263" text="România"/>
<W    EXTRA="intranzitiv"    ID="w10.135"    LEMMA="fi"
MSD="Vmip3s"    Mood="indicative"    Number="singular"
POS="VERB"    Person="third"    Tense="present"
```

```
Type="predicative"  offsetStart="2264"  offsetStop="2268"
text="este"/>
<W    Case="direct"   Gender="feminine"   ID="w10.136"
LEMMA="un"  MSD="Tifsr"  Number="singular"  POS="ARTICLE"
Type="indefinite"  offsetStart="2269"  offsetStop="2270"
text="o"/>
<W   Case="direct"   Definiteness="no"   Gender="feminine"
ID="w10.137" LEMMA="ţară" MSD="Ncfsrn" Number="singular"
POS="NOUN"    Type="common"    offsetStart="2271"    off-
setStop="2275" text="ţară"/>
<W   Case="direct"   Definiteness="no"   Gender="feminine"
ID="w10.138"    LEMMA="european"    MSD="Afpfsrn"    Num-
ber="singular"  POS="ADJECTIVE"  offsetStart="2276"  off-
setStop="2285" text="europeană"/>
<ENTITY   ID="e1"    TYPE="location"    SUBTYPE="country"
DIMENSION="one"   WORDSID="w10.134,"   offsetStart="2256"
offsetStop="2263" Color="Chocolate" />
<ENTITY   ID="e2"    TYPE="location"    SUBTYPE="country"
DIMENSION="one"    WORDSID="w10.137,    w10.138"    off-
setStart="2271" offsetStop="2285" Color="Chocolate"/>
<RELATION   ID="r1"   TYPE="referential"   SUBTYPE="isa"
TRIGGER="w10.135, w10.136" FROM="e1" TO="e2"/>
```

II. *Spatial relations* with three subtypes are listed and illustrated below:

- *near* (express closeness between entities);
Lângă 1:[Moineşti] se află localitatea 2:[Comăneşti]. / (EN) (Near 1:[Moinesti] is situated 2:[Comanesti].) ⇒ [1] near [2];
- *far* (express distance between entities);
1:[Ţara noastră] … decât de 2:[Oceanul Arctic] / (EN) (1:[Our country] ... than the 2:[Artic Ocean]) ⇒ [1] far [2];
- *position* (position between entities involving cardinal point).
1:[Munţii Parâng]… se află la estul 2:[Munţilor Retezat] / (EN) (1:[Parâng Mountains] … are located east of 2:[Retezat Mountains]) ⇒ [1] position [2];

```
<W   Case="direct"   Definiteness="yes"   Gender="masculine"
ID="w158.5"  LEMMA="munte"  MSD="Ncmpry"  Number="plural"
POS="NOUN"    Type="common"    offsetStart="21084"    off-
setStop="21090" text="Munţii"/>
    <W   ID="w158.6"   LEMMA="parîng"   MSD="Np"   POS="NOUN"
Type="proper"    offsetStart="21091"    offsetStop="21097"
text="Parâng"/>
    ...
<W   Case="direct"   ID="w158.12"   LEMMA="sine"   MSD="Px3--r"
POS="PRONOUN"    Person="third"    Type="reflexive"    off-
setStart="21148" offsetStop="21150" text="se">
```

```
<W     EXTRA="tranzitiv"     ID="w158.13"     LEMMA="afla"
MSD="Vmis3s"     Mood="indicative"     Number="singular"
POS="VERB" Person="third" Tense="past" Type="predicative"
offsetStart="2151" offsetStop="21155" text="află">
<W  Case="direct"  Definiteness="yes"  Gender="masculine"
ID="w158.16"  LEMMA="munte"  MSD="Ncmpry"  Number="plural"
POS="NOUN"    Type="common"    offsetStart="21165"    off-
setStop="21172" text="Munții" />
<W              Case="direct"              Definiteness="no"
EXTRA="ParticipleLemma:reteza(tranzitiv)"          Gen-
der="masculine"      ID="w158.17"      LEMMA="retezat"
MSD="Afpmsrn"   Number="singular"   POS="ADJECTIVE"   off-
setStart="21173" offsetStop="21180" text="Retezat"/>
<ENTITY   ID="e60"   TYPE="landform"   SUBTYPE="mountain"
DIMENSION="many"      WORDSID="w158.5,      156.6"      off-
setStart="21084" offsetStop="21097" color="DimGray"/>
<ENTITY   ID="e61"   TYPE="landform"   SUBTYPE="mountain"
DIMENSION="many"      WORDSID="w158.16,      156.17"      off-
setStart="21145" offsetStop="21160" color="DimGray"/>
<RELATION   ID="r43"   TYPE="spatial"   SUBTYPE="position"
TRIGGER="w158.12,  w158.13"  CARDINAL="EST"  FROM="e60"
TO="e61"/>
```

III. *Structural relations* with four subtypes are listed and illustrated below:

- *vicinity* (neighbors).
1:[România] are o deschidere de 234 km la 2:[Marea Neagră] / (EN) (1:[Romania] has an opening of 234 km to 2:[Black Sea]) ⇒ [1] vicinity [2];
- *part-of* (one entity is part of another).
1:[România] concentrează pe teritoriul său două treimi din lanțul 2:[Munților Carpați] / (EN) (1:[Romania] concentrates on its territory two thirds of the 2:[Carpathian Mountains]) ⇒ [2] part-of [1];
- *confluent-of* (branch for rivers).
Bazinul 1:[Mureșului]… 2:[Târnava], cel mai important affluent. / (EN) (1:[Mures] Bay ... 2:[Tarnava], most important confluent.") ⇒ [2] confluent-of [1];
- *source* (root of rivers)
1:[Târnava Mică], cu izvoare în 2:[Munții Gurghiu] / (EN) ([Târnava Mică], with source in 2: [Gurghiu Mountains].) ⇒ [1] source [2];

```
<W   ID="w668.15"   LEMMA="târnava"   MSD="Np"   POS="NOUN"
Type="proper"    offsetStart="84349"    offsetStop="84356"
text="Târnava"/>
<W   Case="direct"   Definiteness="no"   Gender="feminine"
ID="w668.16" LEMMA="mic" MSD="Afpfsrn" Number="singular"
POS="ADJECTIVE"    offsetStart="84357"    offsetStop="84361"
text="Mică"/>
```

```
<W   ID="w668.17"   LEMMA=","   MSD="COMMA"   POS=""   off-
setStart="84361" offsetStop="84362" text=","/>
<W ID="w668.18" LEMMA="cu" MSD="Sp" POS="ADPOSITION" off-
setStart="84363" offsetStop="84365" text="cu"/>
<W   Case="direct"   Definiteness="no"   Gender="feminine"
ID="w668.19"  LEMMA="izvor"  MSD="Ncfprn"  Number="plural"
POS="NOUN"    Type="common"    offsetStart="84366"    off-
setStop="84373" text="izvoare"/>
<W ID="w668.20" LEMMA="în" MSD="Sp" POS="ADPOSITION" off-
setStart="84374" offsetStop="84376" text="în"/>
<W  Case="direct"  Definiteness="yes"  Gender="masculine"
ID="w668.21"  LEMMA="munte"  MSD="Ncmpry"  Number="plural"
POS="NOUN"    Type="common"    offsetStart="84377"    off-
setStop="84383" text="Munţii"/>
<W   ID="w668.22"   LEMMA="gurghiu"   MSD="Np"   POS="NOUN"
Type="proper"    offsetStart="84384"    offsetStop="84391"
text="Gurghiu"/>
<ENTITY     ID="e135"     TYPE="water"     SUBTYPE="river"
DIMENSION="one"     WORDSID="w668.15,     668.16"     off-
setStart="84349" offsetStop="84361" color="DarkGreen"/>
<ENTITY  ID="e136"  TYPE="landform"  SUBTYPE="mountain"
DIMENSION="many"     WORDSID="w668.21,     668.22"     off-
setStart="84377" offsetStop="84391" color="DimGray"/>
<RELATION   ID="r90"   TYPE="structural"   SUBTYPE="source"
TRIGGER="w668.18, w668.19" FROM="e135" TO="e136"/>.
```

4   Statistics and interpretation

From the entire set of entities presented above, our Corpus highlights the values for the 20 entity types. Because it doesn't have yet an explicit notation entirely accepted by the annotators' group, we will restrict to these partial results (see table 2), for two reasons: first, because we relied on the high precision of accepted types of entity (person, location, organization) and on the high precision of resulted tags from the process with POS-Tagger [20] and the second reason is that we wanted to accelerate the manual annotation process, which certainly resulted in many errors as well (e.g. insignificant difference between *cultural* and *tourism* entities, etc.).

Table 2 compares the automatic detection of the most general types of entities which the annotator easily discovers in the manually annotated corpus. With these results, Precision, Recall and F-Measure, we could calculate the corresponding values, after a new check of annotation for the other types. At this moment, the empirical results for automatic detection of the mentioned types of entity are significant. This shows that a classification of geographical entities not resulting in subjective interpretations will be necessary.

Table 2. Automatic and manual entities annotation results

Type	Precision	Recall	F-measure
location	81.25%	83.25%	82.24%
geo_position	83.50%	85.00%	84.24%
person	82.50%	84.00%	83.24%
organization	87.00%	90.25%	88.60%

5   Conclusions and Future Work

This research is a preliminary study in semi-automatic recognition of geographical entities representing the basis for defining the semantic relationships between them. Classification and annotation of entities has meant a long process which is leading to promising results.

The described analysis is a statistic method which proved efficient, and combines the empirical results obtained from manual and automatic annotation. Our current corpus is now Gold Corpus, this kind of study being intended for the automatic recognition of entities for any kind of text with geographical topic. Moreover, the NER was extended in respect of entity diversity, as well as database in existing gazetteers. So far we identified three types of semantic relations for a geographical text, which had been manually annotated. In the future, we aim at defining patterns containing information at lexical and syntactic level to discover these semantic relations and why not, define other types of semantic relation. Besides the presented semantic annotation, the following remain to be discussed: negative relations (e.g. România…, nu face parte din Peninsula Balcanică / (EN) Romania..., is not part of the Balkan Peninsula) or position in the past of some toponyms (e.g. Generația orașelor antice… Histria, Tomis, Callatis, Apulum, Ampelum, Napoca, Potaissa, Sucidava / (EN) Antic city generation … Histria, Tomis, Callatis, Apulum, Ampelum, Napoca, Potaissa, Sucidava), etc.

References

1. Bartalesi Lenzi, V., Moretti, G., Sprugnoli, R.: "CAT: the CELCT Annotation Tool", LREC (2012)
2. Chen, W.-T., Styler, W.: "Anafora: A Web-based General Purpose Annotation Tool", ACL, paper N13-3004 (2013)
3. Cristea, D., Gîfu, D., Colhon, M., Diac, P., Bibiri, A., Mărănduc, C., and Scutelnicu, A.-L.: "Quo Vadis: A Corpus of Entities and Relations" in Language, Production, Cognition, and the Lexicon. Text, Speech and Language Technology, Part VI - Language Resouces and Langauge Engineering, Nuria Gala, Reinhard

Rapp and Gemma Bel-Enguix (eds.), vol. 48/2015, New York, USA, pp. 505-543 (2015)

4. Cunningharn, H., Maynard, D., Bontcheva, K., Tablan, V., Aswani, N, Roberts, I., Gorell, G., Funk, A., Roberts, A. et al.: "Developing Language Processing Components with Gate Version 8" (2014)

5. Gîfu, D. and Vasilache, G.: "A language independent named entity recognition system" in M. Colhon, A. Iftene, V. Barbu Mititelu, D. Cristea, and D Tufiş (eds.), "Alexandru Ioan Cuza" University Publishing House, Iaşi, pp. 181-188 (2014)

6. Ignat, E.: "RARE-UAIC (Robust Anaphora Resolution Engine)", open-resource on META-SHARE, "Alexandru Ioan Cuza" Univrsity of Iaşi (2011)

7. Kim, J.-D., Ohta, T., Pyysalo, S., Nguyen, N., Bossy, R., Tsujii, J.: "Overview of BioNLP Shared Task 2011", ACL (2011)

8. Levi, J. N.: The Syntax and Semantics of Complex Nominals. Academic Press, New York (1978)

9. Li, H., Krause, S., Xu, F., Uszkoreit, H., Hummel, R., Mironova, V.: "Annotating Relation Mentions in Tabloid Press", LREC (2014)

10. Mani, I., Harris, D., Hitzeman, J., Quimby, R., Richer, J., Weller, B.: "SpatialML: Annotation Scheme, Resources and Evaluation", LREC (2008)

11. Morton, T., and LaCivita, J.: "WordFreak: An Open tool for linguistic Annotation", ACL (2003)

12. Năstase, V., Nakov, P., Séaghdha, D., Szpakowicz, S.: "Semantic Relations Between Nominals". Morgan & Claypool Publishers, California (USA) (2013).

13. Neguţ, S., Apostol, G., Ielenicz, M.: "Geografie", Humanitas Educaţional, Bucureşti (2008)

14. Nguyen, N., Tsujii, J., Kim, J.-D.: "Overview of the Protein Coreference task in BioNLP Shared Task 2011", ACL (2011)

15. Ohta, T., Pyysalo, S., Tsujii, J., Ananidou, S.: "Open-domain Anatomical Entity Mention Detection", NACTEM (2012)

16. Séaghdha, D.Ó , Copestake, A.: "Semantic classification with distributional kernels". In: Proceedings of the 22nd International Conference on Computational Linguistics (COLING-08), Manchester, UK (2008)

17. Simionescu, R.: "UAIC Romanian Part of Speech Tagger", resource on nlp-tools.info.uaic.ro, "Alexandru Ioan Cuza" University of Iaşi (2011)

18. Simionescu, R.: Romanian deep noun phrase chunking using graphical grammar studio. In: Moruz, M. A., Cristea, D., Tufiş, D., Iftene, A., Teodorescu, H. N. (eds.) Proceedings of the 8th International Conference "Linguistic Resources And Tools For Processing Of The Romanian Language", 135–143 (2012)

19. South, B.R, Shen, S., Leng, J., B.Forbush, T., DuVall, S.L., and Chapman, W.W.: "A prototype tool set to support machine-assisted annotation". In Proceedings of the 2012 Workshop on Biomedical Natural Language Processing, BioNLP 12, pp. 130–139, Stroudsburg, PA, USA. Association for Computational Linguistics (2012)

20. Tufiş, D. and Dragomirescu, L.: "Tiered Tagging Revisited" in Proceedings of the 4th LREC Conference, Lisabona, pp. 39-42 (2004)