The DDI corpus: An annotated corpus with pharmacological substances and drug–drug interactions

María Herrero-Zazo [a,*], Isabel Segura-Bedmar [a], Paloma Martínez [a], Thierry Declerck [b]

[a] Computer Science Department, Universidad Carlos III de Madrid, Leganés 28911, Madrid, Spain
[b] Language Technology Lab, DFKI GmbH, Saarbrücken D-66123, Germany

ABSTRACT

The management of drug–drug interactions (DDIs) is a critical issue resulting from the overwhelming amount of information available on them. Natural Language Processing (NLP) techniques can provide an interesting way to reduce the time spent by healthcare professionals on reviewing biomedical literature. However, NLP techniques rely mostly on the availability of the annotated corpora. While there are several annotated corpora with biological entities and their relationships, there is a lack of corpora annotated with pharmacological substances and DDIs. Moreover, other works in this field have focused in pharmacokinetic (PK) DDIs only, but not in pharmacodynamic (PD) DDIs. To address this problem, we have created a manually annotated corpus consisting of 792 texts selected from the DrugBank database and other 233 Medline abstracts. This fined-grained corpus has been annotated with a total of 18,502 pharmacological substances and 5028 DDIs, including both PK as well as PD interactions. The quality and consistency of the annotation process has been ensured through the creation of annotation guidelines and has been evaluated by the measurement of the inter-annotator agreement between two annotators. The agreement was almost perfect (Kappa up to 0.96 and generally over 0.80), except for the DDIs in the MedLine database (0.55–0.72). The DDI corpus has been used in the SemEval 2013 DDIExtraction challenge as a gold standard for the evaluation of information extraction techniques applied to the recognition of pharmacological substances and the detection of DDIs from biomedical texts. DDIExtraction 2013 has attracted wide attention with a total of 14 teams from 7 different countries. For the task of recognition and classification of pharmacological names, the best system achieved an F1 of 71.5%, while, for the detection and classification of DDIs, the best result was F1 of 65.1%. These results show that the corpus has enough quality to be used for training and testing NLP techniques applied to the field of Pharmacovigilance. The DDI corpus and the annotation guidelines are free for use for academic research and are available at http://labda.inf.uc3m.es/ddicorpus.

1. Introduction

Motivated by the information explosion in Biomedicine, over the last decade there has been a surge of interest in using Natural Language Processing (NLP) techniques to retrieve and extract information from biomedical texts [1]. Most of these techniques rely heavily on annotated corpora to learn models that can be used to extract information from raw text. Annotated corpora are valuable resources as they provide a gold standard data for the repeatable automatic training and evaluation of NLP techniques [2]. Most recent research has centered on biological entities and their relationships (such as gene and protein interactions) mainly as a result of the availability of annotated corpora in the biological domain (see Section 2).

Other biomedical domains such as Pharmacovigilance (the science concerned with the prevention of adverse drug reactions) can benefit from NLP techniques. This is the case of drug–drug interactions (DDIs), which are a common adverse drug reactions having a significant impact on patient safety and healthcare costs [3,4]. Although there is a large quantity of drug databases and semi-structured resources (such as DrugBank [5], Stockley [6] or Drug Interactions Facts [7], among others) to assist healthcare professionals avoid DDIs, the quality of these databases is very uneven and the consistency of their content is limited, so it is very difficult to assign a real clinical significance to each interaction [8,9]. On the other hand, despite the availability of these databases, a large amount of the most current and valuable information is unstructured, written in natural language and hidden in published articles, scientific journals, books and technical reports [3]. Only in the

* Corresponding author.
E-mail addresses: mhzazo@pa.uc3m.es (M. Herrero-Zazo), isegura@inf.uc3m.es (I. Segura-Bedmar), pmf@inf.uc3m.es (P. Martínez), declerck@dfki.de (T. Declerck).

bibliographic database MedLine, the number of published researches is increased between 10,000 and 20,000 articles per week [10]. Therefore, researchers and healthcare professionals must review a large number of drug safety reports as well as publications in the fields of medicine and pharmacology in order to be kept up-to-date with everything published about DDIs. Several published drug safety studies have shown that DDIs may be detected too late, when millions of patients have already been exposed [11]. Therefore, the management of DDIs is a critical issue as a result of the overwhelming amount of information available on them [12]. Thus, the development of automatic methods for collecting, maintaining and interpreting the information about drugs is crucial to achieving a real improvement in the early detection of DDIs.

In this paper, we describe the construction of a gold standard corpus, called the DDI corpus, which is annotated with pharmacological substances as well as the interactions between them. The DDI corpus is the first corpus which includes pharmacodynamic (PD) and pharmacokinetic (PK) DDIs. A PD DDI occurs when the pharmacological effects of one drug are modified by the presence of another drug, while a PK DDI is the result from the interference of drug absorption, distribution, metabolism and/or elimination of a drug by another drug. While there are several annotated corpora with biological entities and their relationships (see Section 2), the shortage of annotated corpora for DDI extraction is the main bottleneck in the development of NLP systems for this area of Pharmacovigilance.

The DDI corpus has been developed for the DDI Extraction 2013 challenge (http://www.cs.york.ac.uk/semeval-2013/task9/), whose main goal is to provide a common framework for the evaluation of information extraction techniques applied to the recognition of pharmacological substances and the detection of DDIs from biomedical texts. For this purpose, two substasks have been proposed: the recognition and classification of drug names and the extraction and classification of their interactions. A detailed description of the task, the participant teams, the results and the evaluation metrics can be found in [13].

2. Related work

The identification of drug names is a preliminary and crucial step in many text mining tasks such as the detection of the outbreak of diseases [14], the extraction of medication-related information [15], the detection of adverse-drug events [16] or the extraction of relationships such as drug-disease [17], drug-gene interactions [18], DDIs [19], among many others. In fact, several corpora have been built for these purposes in recent years. Here, we review the main corpora annotated with drug entities, giving a special focus on those corpora that also contain DDIs.

Since each corpus has been developed for a specific task, the definition of the drug entity varies significantly from corpus to corpus. Thus, for example, in Clinical E-Science Framework (CLEF) [20] and BioText [21] corpora, drug names and therapeutic devices or interventions are annotated with the same entity type. Other corpora such as ADE (Adverse Drug Effect) [22], EU-ADR (Exploring and Understanding Adverse Drug Reactions) [2] or ITI TXM (Tissue Expressions and Protein–Protein Interactions) [23] use a single entity type to annotate both drugs and chemicals, while the BioCaster [14] corpus distinguishes between substances for the treatment of diseases and chemicals not intended for therapeutic purposes. Corpora such as PK-DDI (Pharmacokinetic drug–drug interaction) [24] or that developed by Rubrichi and Quaglini [25] propose a more fine-grained classification of pharmacological substances. The annotation schema of the PK-DDI corpus describes three types of entity to annotate pharmacological substances: ACTIVE INGREDIENT, DRUG PRODUCT and METABOLITE. Similarly, Rubrichi et al., propose three different entity types: ACTIVE DRUG INGREDIENT, DRUG or DRUG CLASS. In the PK corpus [26], drug names and metabolites are automatically annotated using a dictionary from the DrugBank database, and later manually reviewed.

To the best of our knowledge, the only two works that have addressed the annotation of DDIs are the PK-DDI corpus and the PK corpus. These are the closest works to ours, but differ from it in the scope of the annotated DDIs. While these corpora only contain PK DDIs, the DDI corpus also includes PD DDIs.

As regards the type of document, almost all of the aforementioned corpora are made up of MedLine abstracts (BioText, ADE, EU-ADR, and PK), while full articles have been used only in the ITI TXM corpora. Other corpora focus on the annotation of clinical texts such as discharge summaries [27], patient records [20] or clinical notes [28]. Both the PK-DDI corpus and that developed by Rubrichi et al., consist of texts taken from drug package inserts, which are one of the most important sources of information for healthcare professionals and patients on the use of medicines.

Table 1 shows the size of the aforesaid corpora. As regards the number of drug entities, the ITI TXM corpus is the largest corpora with almost 18,000 drug compounds, followed by the PK-DDI corpus. The PK-DDI corpus was annotated with a total of 3551 active ingredients, 234 drug products and 201 metabolites. Only the PK and PK-DDI corpus contain DDIs. The size of the different corpora of MedLine abstracts never exceeds 500 abstracts.

Although almost all of the aforementioned corpora were created in accordance with a set of guidelines, only a small number of corpora have published their guidelines. Similarly, only some corpora provide inter-annotator agreement scores. In general, the consistency in the annotation of drug entities is high (greater than 75%), while IAA scores reported in the PK-DDI corpus (around 60%) suggest that annotation of DDIs is a more complex task than simply the identification of drug names.

3. Methods

A previous version of the DDI corpus was created for the DDIExtraction 2011 challenge [29]. The goal of this task was to promote research and provide a common framework for comparing the latest advances in Information Extraction techniques applied to the extraction of DDIs from biomedical texts. This earlier version of the DDI corpus consisted of 579 documents describing drug interactions that were taken from the DrugBank database. The documents were parsed using the Unified Medical Language System (UMLS) MetaMap Transfer tool (MMTx) [30] to automatically recognize drugs. Then, a pharmacist manually annotated the DDIs in texts. The main limitations of this previous version were (1) drugs were automatically annotated without any manual intervention in the process, (2) no guidelines were produced, (3) the annotation was carried out by a single annotator, and (4) the quality of the corpus was not evaluated because the inter-agreement annotator was not measured.

In the current new version, we have made several improvements. First of all, two experts and two text miners with background in pharmacovigilance participated in the annotation task

Table 1
Size of the biomedical corpora annotated with drugs.

Corpus	Type of document	Size	Drugs	DDIs
EU-ADR	MedLine abstract	300	1753	
ADE	MedLine sentences	4272	5063	
ITI TXM	Full articles	400	18,000	
PK-DDI	Drug package inserts	68	3986	592
PK	MedLine abstract	428		1333
CLEF	Patient records	150	197	

(see Section 3.3). Annotation guidelines were produced and inter-annotator agreement (IAA), a measure of the agreement between annotators annotating the same recording, was measured in order to assess the level of difficulty of the annotation task as well as the quality of the corpus. Another significant improvement was the addition of MedLine abstracts to the DDI corpus.

3.1. Building the corpus

As pointed out in Section 3, the DDI corpus is made up of texts from two different sources: documents describing DDIs from the DrugBank database (DDI-DrugBank corpus) and MedLine abstracts (DDI-MedLine corpus). The size of the previous version of the DDI corpus was increased with 213 new texts from DrugBank (which are added to the 579 existing documents) and with 233 MedLine abstracts. Therefore, the whole DDI corpus consists of 1,025 documents. A detailed description of the method used to collect the documents for the DDI-DrugBank corpus can be found in [19]. Document selection for the DDI-MedLine corpus was carried out against PubMed. An initial set of documents was selected from PubMed using a query with "drug interactions" as Medical Subject Headings (MeSH) Terms. This query returned 116,919 citations (published between 1975 and 2011) of which 233 documents were randomly selected for annotation (documents without an abstract section were ruled out).

The MMTx [30] tool was used to analyze the documents in the corpus both syntactically and semantically. The basic function of this program is to map text to concepts in the UMLS Metathesaurus. This mapping between texts and UMLS concepts enables entities to be pre-annotated automatically. A more detailed description of this process can be found in [29].

The DDI corpus is distributed in XML documents following the format proposed by Pyysalo et al. [31], in order to unify the different formats of the main corpora for protein–protein interaction extraction. Our main goal is to guarantee high corpus usage and to encourage groups studying PPI extraction to adapt their systems to the problem of DDI extraction. This shared format follows the standoff annotation principle in which the original sentence text is preserved and all entities are stored as offsets (see Fig. 1).

3.2. Annotation guidelines

Our annotation guidelines provide clear and accurate definitions for all those entities and relationships described in the annotation schema (see the Graphical Abstract). This document also contains the rules and conventions on how the annotation task should be carried out as well as providing examples clarifying their use. Moreover, during the annotation process different linguistic phenomena affecting the annotation of drug names arose. Also the annotation of relationships between these drugs was affected by syntactic aspects. All of them are collected in the annotation guidelines and have been reviewed in [32].

Four entity types were proposed to annotate pharmacological substances: drug, brand, group and drug_n. The drug type is used to annotate those human medicines known by a generic name,

whereas those drugs described by a trade or brand name are annotated as brand entities. A drug medication frequently has several brand names since different companies can market it. The use of a brand-name drug instead of its generic name may be related to a higher risk of adverse drug events [33,34]. The use of either generic or brand names depends on the drug information source. Thus, while generic names are used in medical and pharmacological textbooks as well as scientific medical journals, brand names are to be used in drug product labels.

Since the descriptions of DDIs involving groups of drugs are very common in texts, our annotators decided to include the group type to annotate groups of drugs. Extrapolating from drug interactions involving a specific compound to interactions involving its group is a common procedure in some DDI information sources. However some authors have established that this procedure is wrong because this generalization is not true for all drugs [35,36]. The last entity type, drug_n, refers to those active substances not approved for human use, such as, and among others, toxins or pesticides. This type was included because interactions between drugs and substances not approved for human use are frequently reported in Medline documents.

As regards the relationships, four different types of DDI relationships are proposed:

- mechanism: This type is used to annotate DDIs that are described by their PK mechanism (e.g. *Grepafloxacin may inhibit the metabolism of theobromine*).
- effect: This type is used to annotate DDIs describing an effect (e.g. *In uninfected volunteers, 46% developed rash while receiving SUSTIVA and clarithromycin*) or a PD mechanism (e.g. *Chlorthalidone may potentiate the action of other antihypertensive drugs*).
- advice: This type is used when a recommendation or advice regarding a drug interaction is given (e.g. *UROXATRAL should not be used in combination with other alpha-blockers*).
- int: This type is used when a DDI appears in the text without providing any additional information (e.g. *The interaction of omeprazole and ketoconazole has been established*).

Figs. 2 and 3 show sentences describing DDIs. In Fig. 2, the first sentence describes two interactions: effect and mechanism, and the last one also describes a DDI of effect type. In Fig. 3, DDIs of effect type are described between *fenfluramine* and a group of drugs, *antihypertensive drugs*, as well as with some of its members (*guanethidine, methyldopa, reserpine*). The last sentence gives an advice to avoid a DDI.

The proposed classification of DDIs is consistent with the information requirements established by pharmacology experts for an appropriate management of DDIs in the clinical setting [35,36]. For this purpose, healthcare professionals must be provided with information on how the interaction occurs (mechanism), what consequences can be expected (effect) and how it can be managed to avoid or reduce the associated risk (advice).

Furthermore, this classification is useful to reflect the type of information provided from different sources. Thus, drug product labels provide little advice on how to minimize the risk of an

```
<document id="DDI-DrugBank.d548">
 <sentence id="DDI-DrugBank.d548.s0" text="Tetracycline, a bacteriostatic antibiotic, may antagonize the
  bactercidal effect of penicillin and concurrent use of these drugs should be avoided.">
   <entity id="DDI-DrugBank.d548.s0.e0" charOffset="0-11" type="drug" text="Tetracycline"/>
   <entity id="DDI-DrugBank.d548.s0.e1" charOffset="16-40" type="group" text="bacteriostatic antibiotic"/>
   <entity id="DDI-DrugBank.d548.s0.e2" charOffset="84-93" type="drug" text="penicillin"/>
   <ddi id="DDI-DrugBank.d548.s0.d0" e1="DDI-DrugBank.d548.s0.e0" e2="DDI-DrugBank.d548.s0.e2"
    type="effect"/>
 </sentence>
</document>
```
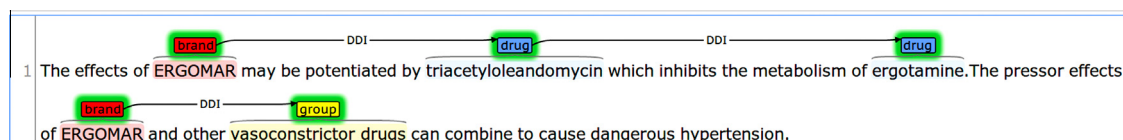
Fig. 1. A fragment from the annotated corpus.

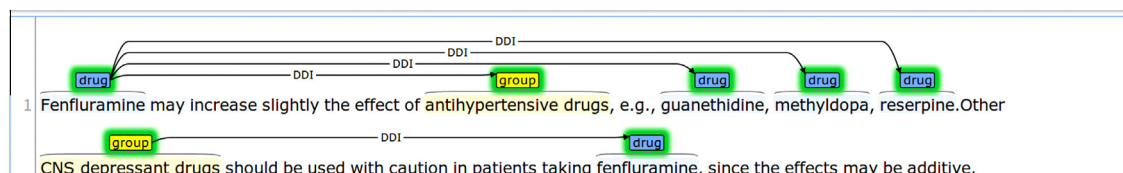**Fig. 2.** Examples of DDIs: effect and mechanism.

Fig. 3. Examples of DDIs: effect and advice.

interaction, whereas PK descriptions are very common in these documents [35]. On the other hand, DDI compendia (such as Stockley's drug interactions [6] or Drug Interaction Facts [7]) also contain considerable information on advice regarding drug interactions.

In summary, the annotation guidelines provide detailed definitions and examples of what substances and interactions should be annotated and which ones should not. For example metabolites and excipients should be annotated as drug_n entities, while drug-protein interactions or endogenous substances should not be included in the annotation. Additionally the annotation guidelines contain rules to deal with the annotation of specific linguistic phenomena such as discontinuous names, abbreviations, nested named entities, synonyms, adjectives, hypernymic propositions, and interacting entities. The annotation guidelines can be found at: http://labda.inf.uc3m.es/ddicorpus.

3.3. Annotating the corpus

This section describes the process followed in the annotation of drugs and their interactions in the DDI corpus. Two expert pharmacists with a substantial background in Pharmacovigilance carried out the annotation task. The first contributed to the creation of annotation guidelines and marked up the whole corpus, while the second annotated a total of 1600 randomly selected sentences from the DDI-DrugBank corpus and 400 ones from the DDI-MedLine corpus. A text miner with background in pharmacovigilance assisted annotators in technical aspects, such as the use of the annotation tool, and participated in the harmonization process, giving advice to resolve conflicting annotations between the two annotators after the annotation process. XML Notepad was the annotation tool used in the process. As explained in Section 3.1, all documents were pre-annotated with pharmacological substance entities by the MetaMap tool (all entities were annotated with drug type). Then, the annotators manually reviewed these labels and added new ones and removed and modified (in particular the type of entity) when necessary. All mentions of pharmacological substances were annotated (even those that were not involved in a DDI). Finally, DDIs were manually annotated at a sentence level.

The annotation process relied on annotation guidelines. They were created in an iterative process. In their earliest development stage, the first annotator studied and annotated a set of 30 documents from DrugBank and 10 Medline abstracts. The previously defined annotation schema and initial annotation guidelines were discussed and established by a multidisciplinary team with two text mining experts and two pharmacists. The text mining experts

checked and reviewed those sentences containing disagreements, which were classified according to the main reason for discrepancy (e.g. missed entity annotation, partial matching, different entity type assigned, missed DDI annotation, different DDI type assigned). These cases were studied and discussed between the two annotators and the two text mining experts, who helped to achieve consensus on the final corpus. Changes in the corpus were made accordingly to the consensus-driven decisions. On the same way, annotation guidelines were modified to include new rules and examples.

3.4. Inter-annotator agreement

To assess the consistency and quality of the corpus as well as the complexity of the annotation task, the inter-annotator agreement was measured in terms of the standard Kappa statistic [37]. From each dataset, DDI-DrugBank and DDI-MedLine, a set of documents was randomly selected and annotated by two different annotators. We should note that the IAA scores were measured after a rigorous process to define strict, comprehensive and clear guidelines. For this reason, IAA scores were calculated under exact match criteria, that is, the annotations should overlap completely. For the entities, their annotations should overlap completely and annotators should also agree on the assigned types. As regards the interactions, the annotators should agree in the annotation of the interacting drugs as well as the type assigned to the interaction. IAA results are shown in the following section.

4. Results

Based on the sentence splitting during preprocessing, the DDI-DrugBank corpus contains 6795 sentences, and the DDI-MedLine corpus is made up of 2147 sentences. Table 2 shows the number of the named entity types annotated in each corpus. The most common type was drug (63%) in both corpora. However, the numbers of other types of entity differ between both sub-corpora. For example, while the second most common type in the DDI-MedLine corpus was drug-n (23%), these substances hardly ever occur in the

Table 2
Numbers of the annotated entities in the DDI corpus.

	DDI-DrugBank	DDI-MedLine	Total
DRUG	9901 (63%)	1745 (63%)	11,646 (63%)
BRAND	1824 (12%)	42 (1.5%)	1866 (10%)
GROUP	3901 (25%)	324 (12%)	4225 (23%)
DRUG_N	130 (1%)	635 (23%)	765 (4%)
TOTAL	15,756	2746	18,502

Table 3
Numbers of the annotated relationships in each corpus.

	DDI-DrugBank	DDI-MedLine	Total
EFFECT	1855 (39.4%)	214 (65.4%)	2069 (41.1%)
MECHANISM	1,539 (32.7%)	86 (26.3%)	1625 (32.3%)
ADVICE	1035 (22%)	15 (4.6%)	1050 (20.9%)
INT	272 (5.8%)	12 (3.7%)	284 (5.6%)
TOTAL	4701	327	5028

Table 5
IAA results of the annotated relationships in the DDI corpus.

	DDI-DrugBank	DDI-MedLine
K_{EFFECT}	0.7525	0.5548
$K_{MECHANISM}$	0.4214	0.5577
K_{ADVICE}	0.9428	0.5587
K_{INT}	0.9558	0.7252
K	0.8385	0.6213

DDI-DrugBank corpus. As regards relationships, Table 3 shows the numbers of the annotated relationships in each corpus. Effect was the dominant relationship found in the whole DDI corpus. Advice, accounting for 20.9% of the whole relationship, showed an even greater disproportion between both corpora because the ratio of relationship being advice interactions in the DDI-DrugBank corpus (22%) is much higher than in the DDI-Medline corpus (5%). These results are further discussed in the next section.

Table 4 presents the results for the agreement per type of entity. For the DDI-DrugBank corpus, IAA results show very high agreement (0.95) for drug, brand and group entities compared to moderate agreement for drug-n entities (0.44). For the DDI-MedLine corpus, all types of entities show close and good IAA scores. In general, IAA was higher for the DDI-DrugBank corpus than for the DDI-MedLine corpus, except for the drug-n entity that exhibits greater IAA in the DDI-MedLine corpus. The overall k per type of interaction (see Table 5) is greater for the DDI-DrugBank corpus (0.8385) than for the DDI-MedLine corpus (0.6213). The int type exhibits the highest IAA in both corpora. While in the DDI-MedLine corpus, the other types (effect, mechanism and advice) show very close scores to each other (0.55), in the DDI-DrugBank corpus, the mechanism type shows a lower agreement than the other types. These results are further discussed in the next section.

5. Discussion

Recently, there have been several attempts to build annotated corpora of DDIs [24,26]. These corpora vary in size, type of annotation and type of document.

The size of the DDI corpus is significantly larger than that of other corpora annotated with drugs and DDIs, both in the number of documents, as well as in the total number of annotated entities and relationships.

The DDI corpus is made up of two different types of text: MedLine abstracts and documents describing DDIs from the DrugBank database. Thus, the corpus covers two different styles of biomedical text: while the texts taken from the DrugBank database are completely focused on the description of DDIs, the main topic of the scientific texts would not necessarily be on DDIs. Moreover, while abstracts are usually written in a very scientific language, the language used in the texts from DrugBank is similar to the language used in package inserts.

Our annotation guidelines could serve as a standard for annotating drug names. Further, it is important to note that the DDI corpus is the only available corpus specifically annotated with groups

Table 4
IAA results of the annotated entities in the DDI corpus.

	DDI-DrugBank	DDI-MedLine
K_{DRUG}	0.9534	0.8467
K_{BRAND}	0.9569	0.8853
K_{GROUP}	0.9563	0.8299
K_{DRUG_n}	0.4422	0.8122
K	0.9104	0.7962

of drugs and substances not approved for human use. The PK and of DDIs occurring through a PK mechanism. According to the authors of the PK-DDI corpus, the vocabulary used to describe this type of DDIs is significantly different from that used to describe PD DDIs because they are discovered in distinct ways. In contrast to these previous corpora, the DDI corpus is annotated with all DDIs described in its texts, including both PK and PD DDIs. Additionally, the proposed classification of DDIs is based on the information requirements for the effective management of DDIs.

The different nature of the texts determines that the types of entity and relationship have different ratios in the two subcorpora. For example, substances not approved for human use are the second most common type of entity in DDI-MedLine, these substances account for only about 1 percent of the entities in the DDI-DrugBank corpus. Similarly, brand drugs are about 12% of the entities in the DDI-DrugBank corpus; however this type had the lowest frequency in the DDI-MedLine corpus. These observations make sense because MedLine abstracts usually describe results from laboratory experiments, while DrugBank texts are mainly compiled from repositories of drug interactions. As regards the distribution of the relationships, the main difference between two subcorpora is that the advice relationship is far more frequent in DDI-DrugBank than in DDI-MedLine corpus. This is also consistent with the fact that the texts from DrugBank seem to be aimed at health-care professionals because these texts usually contain recommendations to avoid any drug interactions and their side effects.

The most common type of relationship in the corpus is effect. Thus, this corpus is annotated with a large amount of information describing PD mechanisms and interaction effects. At the same time, the corpus contains a lot of information on PK DDIs. However, both DrugBank and MedLine documents in the corpus present a low frequency of management recommendations. These results agree with the characteristics of the main DDI information sources [35,36].

Both the quality and consistency of the corpus were evaluated by measuring the IAA scores, which allow the complexity of the annotation task to be determined as well as providing insights into the quality of the guidelines developed. Moreover, IAA also provides an upper bound on the performance of the automatic systems for the detection of pharmacological substances and the interactions between them.

As regards the type of source, IAA was higher for the DDI-DrugBank corpus than for the DDI-MedLine corpus in both entities and relationships. One explanation for this is that MedLine abstracts have far more complexity than texts from the DrugBank database, which are usually expressed in simple sentences.

Similar to other annotated corpora [22,2], IAA scores are higher for entities than for relationships. As regards the type of entity, the highest IAA score is obtained for the brand type in both DDI-DrugBank and DDI-MedLine corpora (see Table 4). This may be because branded drug names are carefully selected by the manufacturer to be short, unique and easy to remember [38]. A high level of agreement is observed for drugs and groups. High IAA scores may indicate that these types are more clearly defined than others in the annotation guidelines [39]. For example, our annotators found the identification of experimental drugs (e.g. *pempidine*),

which should be annotated as drug_n type, more difficult than the identification of the names referring to approved drugs or groups of drugs.

On the other hand, IAA results show moderate agreement for drug_n entities. These results can be due to the large variety of substances included in this type. Additionally, since some of these substances can be both endogenous (produced inside an organism) and exogenous (produced outside the body) (e.g. *calcium* or *dopamine*), their recognition depends substantially on the context in which they appear. We should comment that the mentions of endogenous substances should not be annotated as pharmacological substances. In particular, the agreement was lower for drug_n in the DrugBank corpus than in the DDI-MedLine. Metabolite names (e.g. *descarboethoxyloratadine*), which are very similar to drug names (e.g. *loratadine*), are very frequent in DrugBank texts. Annotators often have difficulty distinguishing between both types. Similarly another main reason for disagreement between annotators was the classification of substances such as vitamins since some of them can be considered as group of drugs (e.g. *vitamin A*) while others are drugs (e.g. *betacarotene*). These differences were resolved in the harmonization process and more accurate explanations were included in the annotation guidelines.

In conclusion, the IAA scores show that annotation guidelines have been successfully developed and validated for the annotation of complex drug names such as stereoisomer (e.g. *S-warfarin*), salts (e.g. *oxycodone hydrochloride*) or nested named terms (e.g., *thiazide diuretics*). Therefore, the DDI corpus may be a valuable resource for developing systems for pharmacological substance recognition.

In general terms, fairly high IAA results are obtained per type of interaction (see Table 5). The int type presents the highest IAA scores in both DDI-DrugBank and DDI-Medline datasets. However, this is the lowest common type of relationship (less than 6%) annotated in the corpus (see Table 3). The second DDI relationship with higher IAA results is the advice type. This type of DDI information is very clear and can be easily identified by manual annotators in both types of document. On the other hand, two main reasons for disagreement in DDI type advice have been observed. Firstly, annotators are frequently confused with sentences containing a recommendation for a specific DDI effect, as in the following sentence:'*Consider additive sedative effects and confusional states to emerge if chlorprothixene is given with benzodiazepines or barbiturates*'. Similarly annotators also had problems with sentences describing a PK mechanism and suggesting a posology change to avoid undesired consequences. For example, the sentence: '*Fenofibrate should be taken at least 1 h before or 4–6 h after a bile acid binding resin to avoid impeding its absorption*' was considered as advice by one annotator while the another one classified it as mechanism. As shown in Table 5, the mechanism type shows the lowest IAA scores in DDI-DrugBank corpus. One reason for this result is that annotators find it difficult to distinguish between sentences describing a PD mechanism or an effect. This observation led to the final annotation of PD interactions with the effect type.

In general, some disagreements may be because many sentences provide various textual evidence of the same interaction and each piece of textual evidence may correspond to a different type of drug interaction. This is very common in complex sentences because subordinate clauses often describe different properties of the same interaction. In these cases, the guidelines have proposed a priority rule to assign the type of interaction (see the guidelines section). However, sometimes the annotators incorrectly applied this rule, and they often tended to assign the first type described in the sentence instead of the type according to the priority rule. On the other hand, the guidelines state that clauses in compound sentences should be considered as independent sentences, and thereby, annotators should annotate each of drug interactions described in their clauses.

6. Conclusion

We describe our efforts to build a manually annotated corpus for DDIs in biomedical texts. To the best of our knowledge, the problem of producing an annotated corpus for DDI extraction has not been explored to the depth and extent reported in this work, and the resulting corpus is the most richly semantically annotated resource for pharmacological text processing built to date. The DDI corpus can encourage the NLP community to research the development of automatic tools to support to healthcare professionals in the early detection of DDIs.

To ensure the quality of the corpus and the consistency of the annotation, we decided to focus in the fine-grained annotation described in this paper. In view of the preliminary results obtained in the SemEval 2013 DDI Extraction Task, the main objective of our work, the creation of a gold standard for the recognition and classification of different drug names and their interactions, has been achieved. Once this stage has been reached, a significant challenge for future work is the annotation of relevant features such as drug dosages, time interval between administration of the drugs, and the mechanism and effect of an interaction. These features would be very useful to characterize the previously identified drug interactions. In the future, we would also like to increase the size of the corpus using other textual sources such as package inserts, patient records, case studies, discharge summaries, among others. Additionally we will annotate the interactions at document level for capturing those interactions spanning several sentences.

The resources described in this work, including both the annotated corpus and the annotation guidelines, are available from http://labda.inf.uc3m.es/ddicorpus.

References

[1] Thamrongrattanarit A, Shafir M, Crivaro M, Borukhov B, Meteer M. What can NLP tell us about BioNLP? In: Proceedings of the 2012 workshop on biomedical natural language processing. Association for Computational Linguistics; 2012. p. 122–9.
[2] Van Mulligen EM, Fourrier-Reglat A, Gurwitz D, Molokhia M, Nieto A, Trifiro G, et al. The EU-ADR corpus: annotated drugs, diseases, targets, and their relationships. J Biomed Inform 2012;45:879–84.
[3] Aronson J. Communicating information about drug interactions. Br J Clin Pharmacol 2007;63(6):637–9.
[4] Jankel C, McMillan J, Martin B. Effect of drug interactions on outcomes of patients receiving warfarin or theophylline. Am J Health-Syst Pharm 1994;51(5):661–6.
[5] Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 2006;34(suppl 1):D668–72.
[6] Baxter K, Stockely IH. Stockley'sdrug interactions. 9th ed. London: Pharmaceutical Press London; 2010.
[7] Tatro D. Drug interaction facts 2010: the authority on drug interactions. St. Louis MO: Wolters Kluwer Health; 2010.
[8] Rodríguez-Terol A, Caraballo M, Palma D, Santos-Ramos B, Molina T, Desongles T, et al. Quality of interaction database management systems. Farmacia Hospitalaria (English Edition) 2009;33(3):134–46.
[9] Paczynski RP, Alexander GC, Chinchilli VM, Kruszewski SP. Quality of evidence in drug compendia supporting off-label use of typical and atypical antipsychotic medications. Int J Risk Safety Med 2012;24(3):137–46.
[10] National Library of Medicine: Fact Sheet Medline. <http://www.nlm.nih.gov/pubs/factsheets/medline.html> [Accessed 02.07.13].
[11] Stricker BH, Psaty BM. Detection, verification, and quantification of adverse drug reactions. BMJ: Br Med J 2004;329(7456):44.
[12] Hansten PD. Drug interaction management. Pharm World Sci 2003;25(3):94–7.
[13] Segura-Bedmar I, Martinez P, Herrero-Zazo M. SemEval-2013 Task 9: extraction of drug–drug interactions from biomedical texts. In: Proceedings

of the 7th international workshop on semantic evaluation (SemEval 2013); 2013.

[14] Doan S, Kawazoe A, Conway M, Collier N. Towards role-based filtering of disease outbreak reports. J Biomed Inform 2009;42(5):773–80.

[15] Del éger L, Grouin C, Zweigenbaum P. Extracting medical information from narrative patient records: the case of medication-related information. J Am Med Infor Assoc 2010;17(5):555–8.

[16] Warrer P, Hansen EH, Juhl-Jensen L, Aagaard L. Using text-mining techniques in electronic patient records to identify ADRs from medicine use. Br J Clin Pharmacol 2012;73(5):674–84.

[17] Xu R, Wang Q. Large-scale extraction of accurate drug-disease treatment pairs from biomedical literature for drug repurposing. BMC Bioinformatics 2013;14(1):181.

[18] Sutton N, Wojtulewicz L, Mehta N, Gonzalez G. Automatic approaches for gene-drug interaction extraction from biomedical text: corpus and comparative evaluation. In: Proceedings of the 2012 workshop on biomedical natural language processing. Association for Computational Linguistics; 2012. p. 214–22.

[19] Segura-Bedmar I, Martinez P, de Pablo-Sánchez C. Using a shallow linguistic kernel for drug–drug interaction extraction. J Biomed Inform 2011;44(5):789–804.

[20] Roberts A, Gaizauskas R, Hepple M, Demetriou G, Guo Y, Roberts I. Building a semantically annotated corpus of clinical texts. J Biomed Inform 2009;42(5):950–66.

[21] Rosario B, Hearst MA. Classifying semantic relations in bioscience texts. In: Proceedings of the 42nd annual meeting on association for computational linguistics. Association for Computational Linguistics; 2004. p. 430.

[22] Gurulingappa H, Rajput AM, Roberts A, Fluck J, Hofmann-Apitius M, Toldo L. Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. J Biomed Inform 2012;45:885–92.

[23] Alex B, Grover C, Haddow B, Kabadjov M, Klein E, Matthews M, et al. The ITI TXM corpora: tissue expressions and protein-protein interactions. In: Proceedings of LREC workshop on building and evaluating resources for biomedical text mining Citeseer, vol. 8. 2008. p. 11–8.

[24] Boyce R, Gardner G, Harkema H. Using natural language processing to identify pharmacokinetic drug–drug interactions described in drug package inserts. In: Proceedings of the 2012 workshop on BioNLP. Association for Computational Linguistics; 2012. p. 206–13.

[25] Rubrichi S, Quaglini S. Summary of product characteristics content extraction for a safe drugs usage. J Biomed Inform 2012;45(2):231–9.

[26] Karnik S, Subhadarshini A, Wang Z, Rocha LM, Li L. Extraction of drug–drug interactions using all paths graph kernel. In: Proceedings of the 1st Challenge task on drug–drug interaction Extraction (DDIExtraction 2011); 2011. p. 83–88.

[27] Uzuner Ö, Solti I, Cadag E. Extracting medication information from clinical text. J Am Med Inform Assoc 2010;17(5):514–8.

[28] Jagannathan V, Mullett CJ, Arbogast JG, Halbritter KA, Yellapragada D, Regulapati S, et al. Assessment of commercial NLP engines for medication information extraction from dictated clinical notes. Int J Med Inform 2009;78(4):284.

[29] Segura-Bedmar I, Martınez P, Sánchez-Cisneros D. The 1st DDIExtraction-2011 challenge task: extraction of drug–drug interactions from biomedical texts. In: Proceedings of the 1st Challenge task on drug–drug interaction Extraction (DDIExtraction 2011), vol. 2011. 2011. p. 1–9.

[30] Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In: Proc AMIA symp. American Medical Informatics Association; 2001. p. 17.

[31] Pyysalo S, Airola A, Heimonen J, Björne J, Ginter F, Salakoski T. Comparative analysis of five protein-protein interaction corpora. BMC Bioinformatics 2008;9(Suppl 3):S6.

[32] Herrero-Zazo M, Segura-Bedmar I, Martínez P. Annotation issues in pharmacological texts. Procedia – Soc Behav Sci (in press).

[33] Hochman M, Hochman S, Bor D, McCormick D. News media coverage of medication research: reporting pharmaceutical company funding and use of generic medication names. JAMA 2008;300(13):1544–50.

[34] Steinman MA, Chren M-M. What'sin a name? Use of brand versus generic drug names in United States outpatient practice. J Gen Intern Med 2007;22(5):645–8.

[35] Bergk V, Haefeli WE, Gasse C, Brenner H, Martin-Facklam M. Information deficits in the summary of product characteristics preclude an optimal management of drug interactions: a comparison with evidence from the literature. Eur J Clin Pharmacol 2005;61(5-6):327–35.

[36] Aronson J. Drug interactions-information education and the British National Formulary. Br J Clin Pharmacol 2004;57(4):371–2.

[37] Cohen J et al. A coefficient of agreement for nominal scales. Educ Psych Meas 1960;20(1):37–46.

[38] Boring D. The development and adoption of nonproprietary, established, and proprietary names for pharmaceuticals. Drug Inf J 1997;31(3):621–34.

[39] Pustejovsky J, Stubbs A. Natural language annotation for machine learning. O'Reilly 2012.