Review

On the Use of Parsing for Named Entity Recognition

Miguel A. Alonso *, Carlos Gómez-Rodríguez and Jesús Vilares

Grupo LyS, Departamento de Ciencias da Computación e Tecnoloxías da Información, Universidade da Coruña and CITIC, 15071 A Coruña, Spain; carlos.gomez@udc.es (C.G.-R.); jesus.vilares@udc.es (J.V.)
* Correspondence: miguel.alonso@udc.es

Abstract: Parsing is a core natural language processing technique that can be used to obtain the structure underlying sentences in human languages. Named entity recognition (NER) is the task of identifying the entities that appear in a text. NER is a challenging natural language processing task that is essential to extract knowledge from texts in multiple domains, ranging from financial to medical. It is intuitive that the structure of a text can be helpful to determine whether or not a certain portion of it is an entity and if so, to establish its concrete limits. However, parsing has been a relatively little-used technique in NER systems, since most of them have chosen to consider shallow approaches to deal with text. In this work, we study the characteristics of NER, a task that is far from being solved despite its long history; we analyze the latest advances in parsing that make its use advisable in NER settings; we review the different approaches to NER that make use of syntactic information; and we propose a new way of using parsing in NER based on casting parsing itself as a sequence labeling task.

Keywords: natural language processing; named entity recognition; parsing; sequence labeling

1. Introduction

Named entity recognition (NER) is a task originally defined at the 6th Message Understanding Conference in 1996 [1], and it consists in finding relevant named entities in the text belonging to a set of predefined categories. Typically, the categories considered include personal names, organizations, locations, dates or times (e.g., [2]), but they can be more fine-grained in specialized settings. For example, information about protein–protein interactions can be extracted by relating protein entities [3], drug–drug interactions from drug entities [4], or adverse drug events by relating drug entities to disease entities [4,5]. As a result, NER is a challenging problem that requires advanced *natural language processing* (NLP) techniques, as entities tend to have numerous synonyms and variations that include long phrases and abbreviations [6].

Currently, NER is essential to any information extraction task, while also being the basis of other related or dependent tasks, from relation and event extraction to knowledge discovery and management [7], semantic indexing or question answering [8], with their performance being conditioned by the effectiveness of the entity recognition process. Effective NER is also crucial for the anonymization of documents required in some domains (e.g., clinical documents) before making them available for research purposes [9]. All these tasks can be applied after recognizing the entities involved through a pipeline architecture [10] or by using joint models to learn entities, relations, and/or events at the same time [4].

Most approaches to NER are shallow, sequence labeling systems that are directly trained to recognize entities without regard for the structure or meaning of the text. However, analyzing said structure is helpful for NER, as it provides cues both for detecting the presence of entities (e.g., a direct object of the verb "prescribe" in English, or "pautar" in Spanish, will typically indicate the presence of an entity of type drug) and for delimiting

the exact span of entities (e.g., in "post-COVID-19 pneumonia pulmonary fibrosis," the constituent boundaries of "pulmonary fibrosis" delimit an entity distinct from "pneumonia").

Syntactic parsing, the task of determining the structure of a sentence, is a key task in many NLP applications that need to process text or utterances beyond a shallow level, extracting meaning or relations between objects, entities, or events referred to in the text. Syntactic parsing takes two main forms, depending on the intended output: in *dependency parsing*, the syntactic structure is expressed by means of binary directed relations between words, called dependencies; while in *constituency parsing* (or phrase structure parsing), it is represented as a phrase structure tree that divides the sentence recursively into its constituent units. *Semantic parsing* goes a step further by converting natural language sentences to logical forms following various representation languages, such as Abstract Meaning Representations (AMRs) [11].

In this article, we focus on the use of the information resulting from the parsing process in the NER task. Toward this aim, we start in Section 2 by defining NER, discussing the use of sequence labeling in NLP, and framing NER as a sequence labeling task. We also review the main resources applied to the task and the evaluation measures commonly used. In Section 3 we define parsing and analyze the latest developments in this area that make parsing a convenient tool to be used in NER. We continue in Section 4 with a study of the most relevant NER systems that use parsing information, including both those that use it as a source of features in a sequence labeling setting and those that make use of parsing results to guide the NER process. In Section 5 we discuss the results achieved, summarize the different approaches in tabular form, present a new proposal for the use of parsing in NER based on the consideration of the parsing process itself as a sequential labeling process, and we compare it with a sequence-to-sequence approach. In Section 6 we analyze work related to this article, to finally elaborate the conclusions in Section 7.

2. Named Entity Recognition

Named entity recognition is the task of locating references to entities in texts and classifying them into predefined categories. NER is a crucial component of any text mining application, as has been repeatedly shown in many different domains, from fashion industry intelligence [12] to legal document mining [13], although it has proven especially relevant in biomedical applications [6,14–17], where the categories of entities to extract include chemical terms [18,19], pharmacological substances [20,21], other drug-related information such as dosages or adverse events [22], diseases [23], problems, tests and treatments [24,25], or genes and proteins [15], among others. From an implementation point of view, NER has been linked since its inception with the sequence labeling paradigm.

2.1. NER as a Sequence Labeling Task

Sequence labeling (also known as sequence tagging) is a type of classification task where the input is a sequence of observed values and the output is a categorical label for each member of the sequence. Sequence labeling has long been used to model and solve NLP tasks, where the input values are typically words, although, depending on the task, they can also be smaller units like individual characters [26] or larger units like sentences [27]. To a certain extent, in the context of NLP, sequence labeling can be considered analogous to a software-engineering design pattern [28], since it provides a template of how to solve a problem that can be reused and adapted to different tasks.

It is worth clarifying that sequence labeling models are not to be confused with sequence-to-sequence (seq2seq) models [29]. The main difference is that while sequence labeling models assign exactly one categorical label to each word of the sentence, seq2seq models generate another sequence as output, which can be of arbitrary length. This means that seq2seq models can be used for tasks that do not fit the sequence labeling framework, such as machine translation. Unfortunately, it also means that seq2seq models require

more complex architectures to run (with the use of neural attention being a must) and are significantly slower.

Part-of-Speech (PoS) tagging is probably the most archetypical example of sequence labeling applied to NLP, because the task itself is defined as assigning one label (a part of speech, such as verb or noun) to each element (word) in a sequence. Thus, any implementation of PoS tagging could be said, by definition, to be performing a form of sequence labeling, although pioneering implementations [30] were task-specific (as well as language- and tagset-specific), since they were based on handwritten rules. However, more modern approaches feature trainable models that learn the correspondence between words and tags in context as a supervised learning problem, and thus are instantiations of true generic sequence labeling models, applied to the particular task of PoS tagging. These include early statistical taggers [31], trainable rule-based taggers [32], Hidden Markov Model (HMM) taggers [33], maximum-entropy taggers [34] and, lately, deep learning approaches [35].

Chunking, sometimes referred to as shallow parsing, was incorporated into the sequence labeling paradigm in the mid 90s. This task consists in finding relevant phrases (typically, verb, noun and/or prepositional phrases) in a text. While early approaches were ad hoc [36], the problem was reformulated as a tagging problem by Ramshaw and Marcus [37], who introduced an encoding scheme called IOB (or BIO) tagging. In this approach each word is assigned a tag "I", "O" or "B" depending on whether it occurs "Inside," "Outside," or at the "Beginning" of a chunk. Since then, most approaches to chunking, including machine learning approaches [38] and more recent deep learning approaches [39], have used the IOB tagging scheme or variants of it.

Advances in machine learning (and in particular, new architectures for sequence labeling) have made it possible to apply the pattern to a wider range of tasks. In this line, after the introduction of Conditional Random Fields (CRF) [40] and the averaged perceptron [41], sequence labeling was applied to model a broader spectrum of problems. In sentiment analysis, Choi et al. [42] used a CRF architecture for opinion source extraction, while Jakob and Gurevych [43] did so for opinion target extraction. In shallow discourse parsing, Ghosh et al. [44] used a sequence labeling model to extract arguments, given discourse connectives; and in question answering, Yao et al. [45] presented the first sequence labeling approach to answer extraction. All of these approaches obtained competitive results at their time and were based in IOB tagging or variants thereof. With a different kind of encoding, Frazee [46] applied a CRF architecture to semantic role labeling, predicting argument labels directly (and empty labels for words that did not play the role of arguments). Similarly, language identification in code-switching texts was addressed by Sikdar and Gambäck [47] using a CRF architecture where each word was directly tagged with its corresponding language. On a different note, the task of extractive summarization was also addressed as a sequence labeling task by Shen et al. [48]. In this case, the elements of the input sequence are sentences rather than words and the output labels are binary, representing whether each sentence is chosen to be part of the summary or not.

NER is closely related to chunking, in the sense that the goal is to extract relevant segments of words in a text (although the nature of those segments changes), and the concept of IOB tagging can also be applied. Therefore, it is not surprising that sequence labeling approaches to NER, based on variants of IOB tagging, have been popular from early years and that they have followed the general evolution of sequence labeling architectures outlined above. This way, early approaches used HMM [49] followed by techniques like Maximum Entropy (MaxEnt) [50], CRF [51], Support Vector Machines (SVM) [17], and Structural Support Vector Machines (S-SVM) [52]. Lately, following the general trends in NLP, deep learning techniques have become popular [53,54], including approaches like Convolutional Neural Networks (CNN) [55], Capsule Networks [56,57], Bidirectional Long Short-Term Memory (BiLSTM) [58], or the combination of BiLSTM and CRF [59]. Neural techniques also made it viable to define sequence labeling models on characters instead of words. Misawa et al. [60] used an architecture combining Long Short-Term

Memory (LSTM), CNN, and CRF for a NER task in Japanese where entities do not necessarily follow word boundaries. Krantz et al. [26] used a similar architecture combining LSTM, CNN, and CRF for language-agnostic syllabification.

2.2. Shared Tasks and Data Sets for NER

Since its inception, the NER task has been characterized by the gradual development of collaborative annotated resources that were materialized mainly through the organization of *shared tasks*. In these competitive evaluation workshops, the organizers provide annotated data sets for training in advance, which are used by participating teams from all over the world to fine-tune their systems. Later, test data sets are released for a limited period of time before the official results are provided by the participants. After a shared task has finished, these data sets are used to evaluate emerging new systems, thus enabling comparison between systems. A list of the most relevant shared tasks on NER follows:

- MUC-6 (https://cs.nyu.edu/cs/faculty/grishman/muc6.html), the Sixth Message Understanding Conference, held in 1995 [1,61]. Its data set consists of news from the *Wall Street Journal*.
- MUC-7 (https://www-nlpir.nist.gov/related_projects/muc/index.html), the Seventh Message Understanding Conference, held in 1998 [62]. This time the data set consists of news from the *New York Times*.
- SemEval (https://www.aclweb.org/anthology/venues/semeval/), the series of International Workshops on Semantic Evaluation, an evolution of the Evaluation exercises for Word Sense Disambiguation (SensEval), has held shared tasks since 1998 [63] on text mining in several languages. Usually, some of the tasks proposed each year involve the use of NER.
- IREX (https://nlp.cs.nyu.edu/irex/index-e.html), an information retrieval and information extraction contest for the Japanese language [64]. The CRL NE, IREX General, and WEB NE Japanese data sets are derived from IREX.
- CoNLL-2002 Shared task on Language-Independent Named Entity Recognition (https://www.clips.uantwerpen.be/conll2002/ner/) [65]. The Spanish data are a collection of newswire articles made available by the Spanish EFE News Agency. The Dutch data set consists of four editions of the Belgian newspaper *De Morgen*.
- CoNLL-2003 Shared Task on Language-Independent Named Entity Recognition (https://www.clips.uantwerpen.be/conll2003/ner/) [66]. The data sets consist of Reuters news stories in English and *Frankfurter Rundshau* news in German.
- ACE (https://www.ldc.upenn.edu/collaborations/past-projects/ace), the Automatic Content Extraction events, held from 1999 to 2008 [67]. Throughout these years, their datasets covered languages such as English, Spanish, Arabic, and Chinese.
- BioCreAtIvE (https://biocreative.bioinformatics.udel.edu/), Critical Assessment of Information Extraction systems in Biology. This series of shared tasks started in 2004 [68] in order to develop a community-wide effort for evaluating text mining and information extraction systems applied to the biological domain. Several NER data sets have been derived from this shared task series, namely, BC5CDR [69], a corpus for the evaluation of biomarker chemical–disease relations from the biomedical literature in support of biocuration, new drug discovery, and drug safety surveillance; and the BC2GM (https://github.com/spyysalo/bc2gm-corpus) corpus for gene mention tagging.
- BioNLP Shared Tasks (https://2019.bionlp-ost.org/), dedicated to computational tasks in the field of biomedical text mining, with editions in 2009 [70], 2011 [71], 2013 [72], 2016 [73], and 2019 [74]. The first editions were supported by the GENIA project (http://www.geniaproject.org/) [75].

- N2c2 (https://n2c2.dbmi.hms.harvard.edu/), the National NLP Clinical Challenges, continuation of the i2b2 (Informatics for Integrating Biology & the Bedside) challenges (https://www.i2b2.org/NLP/DataSets/Main.php) held since 2006. Some of these shared tasks on clinical text mining involve the use of NER.
- ShARe/CLEF eHealth Evaluation Lab (https://sites.google.com/site/clefehealth/datasets), a series of shared tasks that started in 2013 [76], focused on NLP and Information Retrieval for clinical care, including several tracks on NER.
- GermEval (https://germeval.github.io/tasks/), the series of evaluation campaigns that focus on NLP for the German language, with the 2014 edition dedicated to NER [77].
- W-NUT (http://noisy-text.github.io/), the Workshops on Noisy User-generated Text, held annually since 2015 [78], are focused on noisy text found in social media, online reviews, crowdsourced data, web forums, clinical records, and language learner essays. Except for a break in 2018–2019, all editions feature a NER-based shared task.
- CCKS (http://www.cipsc.org.cn/sigkg/?cat=9), the series of China Conferences on Knowledge Graph and Semantic Computing, resulting from the merger in 2016 [79] of the former Chinese Knowledge Graph Symposium (CKGS) and Chinese Semantic Web and Web Science Conference (CSWS). CCKS include shared tasks on Chinese NER for several domains (http://sigkg.cn/ccks2020/?page_id=69; http://www.ccks2019.cn/?page_id=62; http://ceur-ws.org/Vol-2242/; http://ceur-ws.org/Vol-1976/; https://dblp.org/db/conf/ccks/ccks2016.html).
- BSNLP (http://bsnlp.cs.helsinki.fi/bsnlp-2019/shared_task.html), the series of Balto-Slavic Natural Language Processing shared tasks on multilingual NER. The data set of 2017 [80] comprised texts in Croatian, Czech, Polish, Russian, Slovak, Slovene and Ukrainian; while the 2019 edition [81] was dedicated to Bulgarian, Czech, Polish, and Russian.
- VNER 2018 (https://vlsp.org.vn/vlsp2018/eval/ner), VLSP 2018 Shared Task on NER for Vietnamese [82]. Texts in the data set were collected from online newspaper websites.
- NSURL-2019 (http://nsurl.org/tasks/task-7-named-entity-recognition-ner-for-farsi/), the First International Workshop on NLP Solutions for Under Resourced Languages, included among its shared tasks one dedicated to NER for Farsi (Task 7) [83].
- IberLEF (https://sites.google.com/view/iberlef2020), the series of Iberian Languages Evaluation Forum workshops formerly known as IberEval (Workshops on Evaluation of Human Language Technologies for Iberian Languages), and dedicated to NLP in Iberian Languages, has always paid attention to entity recognition and processing. At first they proposed tasks on abbreviation recognition and resolution in the Spanish biomedical domain [84,85] and, after that, they decided to include tracks on Portuguese NER [86], Spanish NER [87] and Spanish eHealth Knowledge Discovery [88].

In addition, the following are some of the data sets that are routinely used by NER researchers for evaluation purposes:

- GENIA (http://www.geniaproject.org/genia-corpus) [89], a collection created to support the development and evaluation of information extraction and text mining systems for the molecular biology domain. It contains subcorpora annotated with PoS, constituency (phrase structure), terms, events, relations, and coreferences.
- OntoNotes (https://www.gabormelli.com/RKB/OntoNotes_Corpus) [90], a data set on English, Chinese, and Arabic resulting from the project with the same name. Apart from named entities from several domains (weblogs, news, talk shows, broadcast, Usenet newsgroups, and conversational telephone speech) it also contains structural information (constituency trees and predicate argument structure) and shallow semantics (word sense linked to an ontology and coreference). There are five releases of OntoNotes, from 1.0 to 5.0.

- CHEMDNER (https://biocreative.bioinformatics.udel.edu/resources/biocreative-iv/chemdner-corpus/) [91], a corpus for evaluating text mining and information extraction systems applied to the chemical domain.
- NCBI disease corpus (https://www.ncbi.nlm.nih.gov/CBBresearch/Dogan/DISEASE/) [92], a collection of PubMed abstracts fully annotated at both the mention and concept levels.
- Species-800 (https://github.com/shreyashub/BioFLAIR/tree/master/data/ner/s800), a corpus of PubMed abstracts that contains identified organism mentions.
- Europeana NER corpora for Dutch, French, and German (https://github.com/EuropeanaNewspapers/ner-corpora) [93], with manually annotated historical newspapers.
- LeNER-Br (https://cic.unb.br/~teodecampos/LeNER-Br/), a data set for Named Entity Recognition in Brazilian legal texts [94].
- Swedish NER corpus (https://www.kaggle.com/andreasklintberg/swedish-ner-corpus) with 8000 sentences in Swedish annotated for NER.

2.3. Evaluation Measures for NER Systems

Three measures are usually considered for evaluating NER systems: precision, recall, and F-measure. In the context of NER, the entities labeled in the test data set are considered as ground truth. To compute the metrics, we must consider the numbers of true positives (*TP*), false positives (*FP*), and false negatives (*FN*) with respect to said ground truth, where:

- A true positive is counted for each entity that is returned by a NER system and also appears in the ground truth;
- A false positive is counted for each entity that is returned by a NER system but does not appear in the ground truth;
- A false negative is counted for each entity that is not returned by a NER system but does appear in the ground truth.

Precision, *P*, refers to the percentage of a system's results that correspond to correctly-recognized entities and is computed as indicated below in Equation (1). Recall, *R*, refers to the percentage of the total entities in a text that are successfully recognized by a system and is computed as indicated in Equation (2). Finally, *F-measure* combines precision and recall by means of their harmonic mean, as shown in Equation (3).

$$P = \frac{TP}{TP + FP}, \tag{1}$$

$$R = \frac{TP}{TP + FN}, \tag{2}$$

$$F\text{-}measure = \frac{2 \times P \times R}{P + R}. \tag{3}$$

Nowadays, NER is far from being a solved task. Even in a well-studied and resource-rich language like English, state-of-the-art approaches obtain *F-measures* around 80% in some biomedical data sets (e.g., 80.5% on GENIA [95] (Yu et al., 2020) or 77.1% on the ShARe/CLEF eHealth Task 1 Corpus [96]). As such, NER is currently a highly active topic in NLP research, the subject of frequent shared tasks with a high number of participants. Examples from 2020 are the Spanish CAMTEMIST-NER shared task, with 23 participating teams, featured in IberLEF 2020 [97]; or the English W-NUT-2020 Task 1, with 13 participants, featured in EMNLP 2020 [98].

3. Syntax and Semantic Parsing as Building Blocks for NLP Applications

In spite of some claims about the possibility that large language models could make explicit syntax redundant, which are based on artificial benchmarks [99], the trend in real applications is just the opposite. As a matter of fact, in recent years, improvements in syntactic parsing models have made the constructions resulting from parsing more and more commonly used in various downstream tasks like machine translation [100], opinion mining [101], relation and event extraction [102], question answering [103], or summarization [104]. On the semantic side, semantic parsers have increased their accuracy to the point of becoming useful in applications like summarization [105] or machine translation [106].

In order to make successful use of parsing in NLP applications, we need (1) efficient and accurate parsing models, and (2) an adequate way of using the obtained structures.

3.1. Recent Advancements in Parsing Efficiency and Accuracy

The availability of more powerful machine learning architectures has greatly improved syntactic parsing accuracy, both in dependency parsing [107,108] and in constituency parsing [109]. Moreover, parsing algorithms have been subject to a process of simplification, which has resulted in models that are simpler, more generic and easier to tune:

- In the context of dependency parsing, transition-based parsers that used to require rich feature models to attain an acceptable accuracy with pre-neural models [110] have become viable with a minimal set of generic features by using BiLSTMs [111].
- Mildly non-projective exact-inference parsers that were barely implementable in practice due to the complexity of features needed [112] have become viable, too [113].
- In the case of constituency parsing, transition-based parsers with simple features [114] and reduced transition sequences [115] now obtain good results.

On the semantic side, semantic parsers have also begun to have sufficient accuracy to prove useful in NLP applications [105,106], hinting that semantic representations can be as versatile (or more so) as syntax, as long as accuracy becomes good enough.

3.2. Recent Advancements in the Representation of Parsing Results

The use of the linguistic structures resulting from parsing in NLP applications is not a trivial problem at all. This is in contrast to PoS tagging, for which the fact that PoS tags are simply a sequence of one tag per word, provides a highly standard and universal way of using such information in any neural NLP model: in the form of embeddings that are plugged as input to the network [116]. This makes it extremely easy to plug them into different models, regardless of architecture, or to try embeddings of different kinds of linguistic units (e.g., fine or coarse-grained PoS tags, lemmata, etc.). However, with constructions resulting from parsing, the situation is very different: since syntactic trees (unlike PoS tags) are structures that go beyond the linear order of the words in the sentence, it is not so obvious how they can be used in a way that takes full advantage of syntax while being modular and pluggable, i.e., not conditioning the rest of the model or requiring special resources.

A classic way to use parsing is to extract features from parse trees (such as individual dependencies from dependency trees) and inject them into the application model, as Joshi and Penstein-Rosé [117] or Vilares et al. [118] do with dependency trees for opinion mining. While this approach is relatively simple and pluggable, it does not really make full use of syntax as it is applying a lossy encoding of dependency trees into "bags of dependencies," without regard for the overall structure or the relations between dependencies. In a much more involved approach, Socher et al. [119] syntactically annotated a sentiment treebank to then train a recursive neural network that learns how to apply semantic composition for sentiment analysis. Although this does exploit syntax more fully, it has no modularity as it requires training a special ad hoc model, apart from requiring an ad hoc corpus where tree nodes are annotated with sentiment, a rarely available resource. Vilares et al. [120] forwent the need of special corpora by just computing a dependency parse of

the target sentence and then applying handwritten rules to extract polarity from it, but, again, this is an ad hoc approach and cannot be used to improve existing models or to try different syntactic representations. In a similar way, several recent papers have used tree-specific encoders like graph convolutional networks [101]; but these are specialized models that are not easily pluggable.

Recently, a new paradigm of *parsing as sequence labeling* has arisen; in it, parsing is performed by encoding syntactic trees into sequences of one categorical label per word. This paradigm has been successfully applied to both constituency parsing [121,122] and dependency parsing [123], or even both at once [124]. Apart from providing very fast and environmentally friendly [125] parsers for practical applications, this approach makes it possible to plug syntax into models in the same generic way as for PoS tags, as well as opening possibilities for multitask learning with other sequence labeling tasks, such as NER. Unfortunately, to the best of our knowledge, no sequence labeling approach is available for semantic parsing at the moment.

4. Parsing for NER

As explained before, syntactic information has been shown to be an important asset to improve the accuracy of various NLP tasks. Thus, it is to be expected that such information can be used to improve NER as well as related tasks like relation extraction. However, existing approaches to integrate syntax into NER systems have been affected by the difficulty of using syntax in downstream applications in a pluggable way, as we discussed in the previous section. In the same line as before, some models use syntactic information extracted from parse trees as a feature for standard sequence labeling NER architectures. However, these strategies are limited to using very specific syntactic features and cannot take advantage of the whole parse tree. On the other end of the spectrum, there are models that do use complete parse trees, but have to resort to ad hoc, complex architectures to do so. The most relevant recent references for both approaches are discussed in the rest of this section.

4.1. Syntactic Information as a Feature for Sequence Labeling NER

Sasano and Kurohashi [126] presented a system for Japanese NER based on an SVM classifier that uses several types of structural information, such as that obtained from the head verb of a sentence by means of a syntactic parser and the surface case of the phrase that includes a target entity. To deal with head verbs that do not appear in the training data, case frames are introduced. Case frames describe what kinds of cases each verb can have and what kinds of nouns can fill a case slot. They are learned from a corpus of five hundred million sentences: firstly, entities are detected by a primitive NER system that uses only local features; secondly, case frames are constructed from the sentences containing such entities. Thus, if a given threshold percentage of the examples of a case are classified as pertaining to a certain entity class, the corresponding label is attached to the case. By using all structural information, the performance improves significantly for all data sets, which means that structural information improves the performance of Japanese NER. In particular, syntactic features improve the performance not dramatically, but consistently and independently from the data set. This result also shows that case frame features are general features that can be effective for data from different domains.

In their work [127], Ling and Weld described FIGER, a fine-grained entity recognizer that identifies references to entities in natural language text and labels them with appropriate tags from a set of 112 tags. The training set for these tags is created by exploiting the anchor links in Wikipedia text to automatically label entity segments with suitable tags. A CRF model is trained for segmentation, identifying the boundaries of each text segment that mentions an entity. An adapted perceptron algorithm is used as the final classifier in charge of assigning tags to the detected entities, considering both word-based features (including unigram and bigram features) and syntax-based features such as the head of the segment containing the entity and the syntactic dependency of said head.

Compared to standard NER systems, FIGER shows a higher performance. An error analysis detects that most errors originate from noise in the training data. It is worth mentioning that training data were created without resorting to parsing.

The proposed approach of Luo et al. [128] to chemical NER is based on a neural network. The base classifier is an attention-based bidirectional LSTM with a conditional random field layer, thus trying to leverage document-level global information obtained by an attention mechanism to enforce tagging consistency across multiple instances of the same token in a document. This approach achieves better performance on chemical compound and drug name recognition than other state-of-the-art methods, while requiring little feature engineering. In particular, the authors investigated the effect of linguistic features such as PoS tags and chunks obtained through shallow parsing. The baselines take word and character embeddings as inputs to the model while the additional features are introduced into the deep learning classifier as additional embeddings. Without the attention mechanism, the highest F-score is achieved when only the chunking embedding is added, the main reason being that some entity boundary errors can be revised by the chunking information. When only the PoS embedding is added, the model achieves a smaller improvement. However, with the attention mechanism in place, the contribution of these features to the performance of the model is negligible.

Although these studies demonstrate the potential benefits of incorporating syntactic information, they are limited in either treating noisy syntactic information as gold references for training their taggers, or using direct concatenation to combine that information with context information without weighing it with respect to its contribution to the NER task. Tian et al. [129] tried to find a better way to incorporate syntactic information into deep learning models for NER. For this purpose, they built BioKMNER, a NER model for biomedical texts based on Key-Value Memory Networks (KVMN) [130]. They parsed biomedical text sentences to extract three types of morpho-syntactic information: PoS tags, syntactic constituents, and dependency relations. The KVMN weighs the corresponding syntactic information (values) according to the importance of context features (keys) and combines the weighted syntactic information with the output of the encoder Bidirectional Encoder Representations from Transformers for Biomedical Text Mining (bioBERT) [131]. Finally, the decoder receives the combined embedding and tags the input sequence accordingly. BioKMNER outperforms baselines without memories and achieves new state-of-the-art results on four biomedical data sets.

The system by Tian et al. [129] is based on bioBERT [131], a pre-trained biomedical language model designed for biomedical text mining tasks. It is worth remarking that there is a recent trend in end-to-end NLP systems that use powerful pre-trained language models with huge parameter spaces based on transformers to solve a variety of tasks, as in the case of the Bidirectional Encoder Representations from Transformers (BERT) [132] or Generative Pre-trained Transformer 3 (GPT-3) models [133]. It is important to highlight that it has been shown that explicit syntactic information helps these pre-trained transformer models [134], despite the fact that some authors have questioned the use of explicit syntax in these settings [99].

4.2. Using Complete Parse Trees for NER

In the case of Shi et al. [135], they proposed to use a statistical parsing technique to simultaneously identify biomedical named entities and extract subcellular localization relations for bacterial proteins. In their approach, sentences are automatically annotated by a statistical parser. Then, the constituency parse trees are decorated with annotations on relevant protein, bacterium, and location named entities; and annotations on the path linking related entities in the parse tree of each sentence. Experiments with purely supervised learning showed that in order to be effective, the model required a large curated set to minimize the sparse data problem. Unfortunately, domain-specific annotated corpora are

rare and expensive, so the authors decided to add noisy texts (i.e., with automatically labeled named entities and relations) to the training set. By doing this, the system reaches a competitive performance.

Finkel and Manning [136] proposed a joint model of context-free parsing and named entity recognition, based on a discriminative CRF-based constituency parser. They found that combining parsing and NER improves performance on both tasks, the joint model producing an output that has consistent parse structure and named entity spans and, moreover, is also doing a better job at both tasks than separate models with the same features. This joint model proceeds as follows. In a first phase, the set of constituency trees from the OntoNotes 2.0 data set is modified. As named entities correspond to phrasal nodes, the labels of such phrasal nodes and their descendants in each parse tree are augmented with the type of those named entities. Some additional manual modifications are required, such as removing final periods from the entity annotation, flattening nested noun phrases, and moving adjectives from nested noun phrases into the main noun phrase. The augmented nodes give place to extra copies of the source grammar where named entities are taken into account. This representation will even be able to handle nested named entities in a natural way, although the data set used for evaluation does not contain such entities. This grammar is used by a CRF-based parser that considers features over both the parse rules and the named entities. In the testing phase, the parser analyzes each sentence, and the named entities are extracted from the parse tree. As we can see, this approach differs from the others described in this article in that NER is not considered as a sequence labeling task but rather as a by-product of a parsing process. One of the practical difficulties of this approach is that the size of the corpus employed is much smaller than the treebanks on which parsers are routinely trained, at least for English.

Instead of constituency parsers, Jie et al. [137] proposed a NER model guided by a dependency parser. The basis of this approach is that named entities tend to be covered by single or multiple adjacent dependency arcs, since certain internal structures are expected to exist for most named entities that convey semantically meaningful information. As a result, words inside each named entity typically do not have dependencies with words outside the entities, except for certain words such as head words, which often have incoming arcs from outside words. Thus, the authors derive a semi-Markov CRF model by restricting the space of all possible combinations of entities to those that strictly contain only valid spans, where a valid span either consists of a single word, or is a word sequence that is covered by a chain of dependency arcs where no arc is covered by another. This model performs competitively with respect to conventional linear CRF-based models and exhibits the same time complexity.

Finally, Yu et al. [95] introduced a method to handle both flat and nested named entities by adopting ideas from the biaffine dependency parsing model [107]. The particularity of this system is that instead of using the information resulting from the parsing process, it uses the parsing process to derive the plausibility of each of the possible entities found in the text, without actually building a parse tree. Toward this aim, the authors use a biaffine model on top of a multilayer BiLSTM to assign scores to all possible spans in a sentence. The results are used to rank the candidate spans by their scores and return the top-ranked spans that comply with constraints for flat or nested NER. The experiments show that the system improves over state-of-the-art results on three nested NER corpora and five flat NER corpora. The biaffine mapping and the BERT embedding used as input to the BiLSTM are the components that contribute most to the accuracy of the system.

5. Discussion

Parsing makes it possible to represent the structure of a text and has been repeatedly shown to be useful for improving NER accuracy. Table 1 compiles a summary of the main characteristics of the most relevant NER systems that make use of information derived from parsing. As we explained in the previous section, existing approaches have been limited by the difficulty of integrating hierarchical information such as a parse tree into a

task that is linear in nature. Thus, either they make limited use of such syntactic information [128,129], or they develop ad hoc architectures that result in more complex, less generic, and less efficient models [95,136,137]. What we know for sure is that the use of information from parsers is beneficial but, since they have been tested on different data sets, it is difficult to determine which of those approaches for incorporating parsing information is more effective in general terms. Paradoxically, this issue stems from the success of the NER task: In recent years NER has been applied to so many domains that it has been necessary to create at least one data set for each of them, which has the pernicious effect that the research community has dispersed, with researchers creating systems to work effectively in a particular domain.

Table 1. Main characteristics of named entity recognition (NER) systems using parsing information.

Reference	Language	Base Classifier	Parsing Information	Data Set	Performance F-Measure
Sasano and Kurohashi (2008) [126]	Japanese	SVM	Head verb, case frames	CRL NE	89.40
				IREX General	87.71
				WEB NE	71.03
Ling and Weld (2012) [127]	English	CRF+Perceptron	Syntactic dependencies, head of phrases	Ad hoc from Wikipedia	53.20
Luo et al. (2018) [128]	English	Attention+BiLSTM+CRF	Chunks	CHEMDNER	91.14
				BC5CDR	92.57
Tian et al. (2020) [129]	English	BioBERT	Syntactic constituents, dependency relations	BC2GM	85.67
				BC5CDR (only chemical entities)	94.22
				NCBI-disease	90.11
				Species-800	76.33
Shi et al. (2007) [135]	English	Statistical context-free parser	Phrase structure (constituency) trees	Ad hoc from MEDLINE	83.20
Finkel and Manning (2009) [136]	English	CRF context-free grammar parser	Phrase structure (constituency) trees	English portion of OntoNotes 2.0	74.91 to 88.11 [1]
Jie et al. (2017) [137]	English	Semi-Markov CRF	Dependency trees	English portion of OntoNotes 5.0	79.4 [2]
				SemEval-2010 Task 1 OntoNotes English corpus	75.10 [2]
Yu et al. (2020) [95]	English, German, Spanish, Dutch	BiLSTM	Biaffine dependency parsing model	English OntoNotes [3]	91.3
				CoNLL 2003 (EN)	93.5
				CoNLL 2003 (DE) [4]	86.4
				CoNLL 2003 (ES)	90.3
				CoNLL 2003 (NL)	93.7
				ACE 2004	86.7
				ACE 2005	85.4
				GENIA	80.5

[1] Results are provided for six portions of the English part of OntoNotes. [2] Results reported in this table correspond to predicted dependency trees. [3] Supposedly OntoNotes 2.0, although the version is not clearly identified in the article. [4] On a revised version with more consistent annotations, F = 90.3.

At this point, a promising line of research that has not been tried yet is to integrate the hierarchical information provided by parsing processes into a linear setting by casting syntactic parsing itself as a sequence labeling task [121,123]. Until recently, full syntactic parsing was considered infeasible in practice within the sequence labeling framework: While it was theoretically possible to cast it as sequence labeling, learning algorithms like averaged perceptron or CRF were not powerful enough to achieve practical results. As an example, Spoustová and Spousta [138] presented a sequence labeling approach to dependency parsing as early as 2010, but they reported accuracies 5%–10% behind the state of the

art of the time. Thus, while their work was an interesting exploration and proof of concept, it could hardly be considered a competitive system. It was not until recent years, with the popularization of dense vector representations of linguistic units (embeddings [139]) and the use of recurrent neural networks (especially BiLSTMs) to enrich these representations with context information [140], that generic sequence labeling models started being capable of doing full syntactic parsing. In a pioneering work, Gómez-Rodríguez and Vilares [121] introduced an encoding to represent any constituency tree for a sentence of length n as a sequence of n labels. Several sequence labeling architectures were tried, showing that BiLSTMs were capable of achieving good parsing accuracy (and very fast speeds) where simpler architectures failed. In later work [123], the same approach was tried with dependency parsing, exploring four different ways to cast the problem as sequence labeling and achieving competitive accuracies with two of them, including the one where Spoustová and Spousta [138] had previously obtained impractical results using pre-neural techniques.

This new paradigm of parsing as sequence labeling will be usable not only to integrate deep syntactic trees with NER, but to do it in such a way that we will effectively use the full syntactic information without needing to forgo the standard sequence labeling architectures of NER. Thus, we will obtain NER systems that are fast, scalable, and easily integrable with upstream tasks while also boosting accuracy, thanks to the use of deep syntax. This approach can also be extended to semantic parsing, which generates meaning representations that go beyond syntax trees. However, this will require a reduction of semantic parsing to sequence labeling.

Some NER systems, notably [129], resort to pre-trained language models. End-to-end-models based on large pre-trained language models suffer from high computational costs, with the associated environmental costs [141]; reduced inclusivity in multilingual settings (e.g., GPT-3 is currently only available for English, and training it for a new language has been estimated to cost more than USD 4 million with current hardware [142]); as well as lack of explainability, which can be provided with parsing. In this respect, a practical characteristic of sequence labeling approaches to parsing is that they are more efficient than seq2seq models. For example, the single-core speeds of the seq2seq constituent parsers of Fernández-González and Gómez-Rodríguez [143], albeit optimized for speed, are an order of magnitude slower than those of sequence labeling constituent parsers [121,122]. This is compounded by the fact that sequence labeling is much easier to parallelize, so that the differences can be even larger in multi-core settings. For all these reasons, and while recognizing the usefulness of end-to-end setups and large pre-trained models, non-end-to-end setups that use intermediate tasks explicitly are still preferable if we wish to achieve efficient, green, inclusive, and explainable systems, and will continue to be in the foreseeable future.

6. Related Work

There have been a number of articles reviewing the state of the art in NER in a given moment, but none of them had the use of information derived from parsing processes as their main focus, as in this case.

The work of Nadeau and Sekine [144], for example, is a classical reference that reviews 15 years of research in NER, from 1991 to 2006. They detected that early systems were making use of handcrafted rule-based algorithms, while modern systems most often resorted to machine learning techniques. Handcrafted systems provided good performance at a relatively high system engineering cost. For machine learning systems, a prerequisite was the availability of a large collection of annotated data, a rather rare resource and limited in domain and language coverage. Indeed, most of the work at that point had concentrated on limited domains and textual genres such as news articles and web pages. The application of syntactic information was limited to the use of fixed syntactic constructions for finding candidate named entities and to the use of syntactic relations (e.g., subject–object) to discover more accurate contextual evidence around entities.

Regarding the reviews carried out in the last decade, Vazquez et al. [145] studied the achievements in the recognition of chemical entities mentioned in text, the determination of their chemical structures, and the identification of relationships between chemicals and other entities. It must be taken into account that chemicals may be referenced in documents in a variety of forms: systematic nomenclatures, common names, trade names, database identifiers, or IUPAC International Chemical Identifier strings; with different types of names having different word morphologies. They classified NER approaches into three categories: dictionary-based, morphology-based, and context-based, the latter category being the only one that involves some form of syntactic parsing guided by manual rules, in contrast to current parsing techniques based on treebank data. At that time, hand-made context-free rules had been proposed to describe a kind of "chemical language." Shallow or template-based parsing had also been considered to mine relationships for entities such as proteins and genes, pharmacogenomics entities, or drug and cytochrome proteins. As a result, parsing was limited most of the time to determining certain components of sentences (e.g., subjects), which were then used in a template matching strategy. NER for the chemical domain was reviewed again a few years later by Eltyeb and Salim [146]. They considered a different classification of NER systems in this domain: dictionary-based, rule-based, machine learning-based, and hybrid approaches. Rule-based systems used a set of hand-made rules to extract the names of entities. The handcrafted models consisted of pattern-based and context-based rules, the latter involving, as before, the use of shallow parsing.

In the biomedical domain, Campos et al. [147] analyzed machine learning tools for NER in this context, where it is used to detect entities such as gene, protein, drug, and disease names. It is a complex domain, where many entities are descriptive (e.g., "normal thymic epithelial cells"), several entity names can share one head noun, one entity name can have several forms of spelling, and ambiguous abbreviations are frequently used, among other phenomena. The authors detected that three approaches were used at that time to deal with this variety in entity forms: rule-based approaches for names with a strongly defined orthographic and morphological structure, dictionary-based approaches for closely defined vocabularies of names (e.g., diseases and species), and machine learning approaches for highly dynamic vocabularies of names exhibiting strong variability (e.g., genes and proteins). They focused their survey on the latter and they detected that shallow syntactic parsing benefits pre-processing of gene and protein names, particularly when using *chunking* to divide the text into syntactically correlated parts of words (e.g., noun or verb phrases). They also observed that, given that these linguistic units only provide a local analysis of some tokens in the sentence, additional information can be derived from dependency parsing to collect the relations between a wider range of tokens. NER for the biomedical domain was reviewed again several years later by Alshaikhdeeb and Ahmad [148]. At that time, most methods were relying on machine learning techniques and they reviewed some features that could be used in such techniques, such as morphological features, dictionary-based features, lexical features and distance-based features, but not syntactic features.

A different perspective was taken by Marrero et al. [149], who analyzed the evolution of NER from a theoretical and practical point of view, arguing that the task was actually far from being solved and showing the consequences for the development and evaluation of tools. They focused their review around what the task of NER is in itself, analyzing the different meanings of the term *named entity*. They also analyzed the resources and metrics that were used to solve the task and to measure the results attained, concluding that systems were overfitting to the training corpora, leading to serious limitations in the external validity of NER evaluations, given that systems did not perform well in general but for a particular user and document type.

Another context-specific review work is that of Shaalan [150], who studied the features of common tools used for NER in the Arabic language. This language poses particular challenges for NER, such as the use of the Arabic script; the co-existence of Classic

Arabic, Modern Standard Arabic, and Colloquial Arabic dialects; lack of capitalization; lack of uniformity in writing styles; optional short vowels; and agglutination. As a sample of the complexity of the task in this language, we have that two mentions to entities may appear in one word, given that a pronominal can appear as a suffix pronoun to a nominal. One of the primary approaches for Arabic NER was based on handcrafted local grammatical rules. The structure of Arabic sentences allows a named entity to appear anywhere in the sentence and at different distances from lexical triggers, which complicated the structure of the rules. This led to using base-phrase chunks such as noun phrases and verb phrases, identified by means of shallow syntactic parsing. The other primary approach was based on machine learning classifiers, where syntactic information could also be used, giving rise to hybrid approaches.

Goyal et al. [151] presented the status of NER techniques developed by the research community and identified the issues and challenges (nested entities, ambiguity, annotation of training data, lack of resources) as well as factors (language, text genre, text domain) affecting NER performance, all of them to be considered when designing these systems. They found that earlier systems were most often based on handcrafted rules, including rules based on syntactic–lexical patterns to identify and classify named entities. These systems are highly efficient because they exploit the properties of language-related knowledge, employing domain-specific features to obtain sufficient accuracy. However, they are quite expensive, domain-specific and non-portable. In the case of NER systems based on machine learning, some of them consider chunks of text detected by means of shallow parsing as features.

Névéol et al. [152] offered an overview of clinical NLP for non-English languages. In the case of NER, they found that, similar to approaches for English, the methods for other languages are rule-based, statistical, or a combination of both. Although they did not consider parsing to be one of the most widely used resources, they cited it as one of the NLP techniques used in NER systems. In the clinical domain, NER essentially focuses on two types of entities: personal health identifiers in the context of clinical document de-identification and clinical entities such as diseases, signs/symptoms, procedures or medications, as well as their context of occurrence: negation, assertions, and experiencer (i.e., whether the entities are relevant to the patient or a third party such as a family member or organ donor). They claimed that negation may be easily adapted between languages of the same family that express negation using similar syntactic structures.

Yadav and Bethard [54] presented a survey of deep neural network architectures for NER and contrasted them with previous approaches to NER based on feature engineering and other supervised or semi-supervised learning algorithms. With respect to the use of parsing and syntax, they only briefly cited that shallow syntactic knowledge can be useful as a feature for unsupervised NER systems.

More recently, Hahn and Oleynik [153] reported the latest developments in medical NER for two selected semantic classes, diseases and drugs (or medications), and relations between them. They focused their review on the methodological paradigm shift from standard machine learning techniques to deep learning. They concluded that deep-learning-based approaches outperform classical machine learning ones but, at the same time, small-sized and access-limited corpora create intrinsic problems for data-greedy deep learning. The same applies to special linguistic phenomena of medical sublanguages that have to be overcome by adaptive learning strategies. No mention was made of the use of syntactic information beyond indicating that clinical notes and reports often exhibit syntactically ill-formed, telegraphic language.

Finally, Li et al. [53] reviewed in detail existing deep learning techniques for NER, systematically categorizing approaches based on a taxonomy along three axes: distributed representations for input, context encoder, and tag decoder. Although several deep artificial neural network models try to represent long-distance dependencies that have a clear syntactic component, they do so in an implicit way, through a sequential process in which relevant information is remembered and propagated so that the representation associated

with a given word can include non-local information coming from a different location in the sentence. However, some systems resort to information derived by parsers, such as dependency roles, to build complex distributed representations of words. Minaee et al. [154] presented a comprehensive survey of deep learning models for other classification tasks.

IOB is the most widely used labeling scheme in NER, but it is not the only one. Recently, Zhong et al. [155] proposed several constituent-based (The term *constituent* in *Constituent-based tagging scheme* does not refer to constituents in the sense of constituency parsing, but to each of the elements that constitute (are part of) a named entity or time expression) labeling schemes instead of the traditional IOB positional labeling scheme. More specifically, they define a TOMN scheme to model temporal expressions, where T refers to Time token, M to Modifier, N to Numeral and O to Outside time expressions; and an UGTO scheme to model named entities, where U refers to Uncommon word, G to General modifier, T to Trigger word and O to Outside named entities. Experimental results show that CRF-based methods using these constituent-based labeling schemes perform equally to, or more effectively than, representative state-of-the-art methods on time expression extraction and named entity extraction.

7. Conclusions

Written text is the fundamental element by which human beings record their ideas, desires, aspirations, creations, and the events that occur in their environment; it is ultimately the main medium by which knowledge is transmitted. A clear example of this can be found in articles published in scientific journals like this one. The huge amount of text that is currently generated on a daily basis makes its manual examination unfeasible, making it necessary to create automatic tools, that is, NLP tools, to extract knowledge from it. NER emerges as a basic NLP task for this purpose.

NER is a difficult task. Probably because of this, most systems have opted for an approach based on sequence labeling that makes a limited use of the inherent structure of text. Although this approach has not been able to solve the NER task, it has managed to yield systems with sufficient performance to be applicable in practice. In particular, recent developments in neural architectures have allowed us to increase sequence labeling performance in NLP tasks [27,35,156–158], in part due to the use of contextualized embeddings from language models like Embeddings from Language Models (ELMo) or BERT [159,160]. However, to continue improving the performance of NER systems, it is necessary to incorporate the information provided by the techniques that can analyze, process, and elaborate the structural information of sentences—in other words, parsing. Throughout this article we have shown how NER systems that use parsing information manage to improve over the results of those that do not use it, and how improvements in parsing techniques in recent years allow a smoother and more efficient incorporation of structural information to NER systems.

Regarding the future evolution of the integration of syntactic and semantic information in NER systems, we advise to cast parsing itself as a sequence labeling task, making it much more straightforward to integrate with NER. This way, we will be able to use complete syntactic trees, while at the same time not having to resort to non-standard architectures and retaining the simplicity, genericity, and efficiency of a sequence labeling architecture for NER. This approach can also be used to apply the parsing component of a NER system in a multilingual setting, thanks to the availability of Universal Dependencies (https://universaldependencies.org/) (UD), a unified parsing framework that currently supports 111 languages, with 32 more to be added soon. For example, this approach can be applied to incorporate a parsing component to a NER system in Arabic [161], Persian [162], or French [163] by training a parser with the UD treebank for each of those languages. Moreover, the strategy used in [120,164] to build multilingual sentiment analysis systems can be applied toward building truly multilingual NER systems.

Finally, we would like to point out what we consider to be the main challenges for the successful application of parsing in NER systems:

- A standard framework for NER resources. Although the availability of UD makes it possible to have a multilingual parsing component with common annotation criteria across languages, the same is not the case for the rest of components of a NER system. The NER community needs to move in this direction, which will also facilitate the creation of truly multi-domain NER systems.
- Larger data sets. The deep learning techniques that currently represent the state of the art in both parsing and NER require the largest possible data sets to exploit their full potential.
- Semantic parsing. The latest developments in syntactic parsing indicate that the proposed approach is fast enough to be applicable in large-scale NER systems, as well as accurate enough to provide useful information for the task. It is still unknown whether similar performance can be achievable with respect to semantic parsing, although the prospects are encouraging.

Author Contributions: Conceptualization, M.A.A., C.G.-R. and J.V.; methodology, M.A.A., C.G.-R. and J.V.; formal analysis, M.A.A., C.G.-R. and J.V.; investigation, M.A.A., C.G.-R. and J.V.; resources, M.A.A., C.G.-R. and J.V.; writing—original draft preparation, M.A.A., C.G.-R. and J.V.; writing—review and editing, M.A.A., C.G.-R. and J.V.; supervision, M.A.A., C.G.-R. and J.V.; project administration, M.A.A., C.G.-R. and J.V.; funding acquisition, M.A.A., C.G.-R. and J.V. All authors have read and agreed to the published version of the manuscript.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Grishman, R.; Sundheim, B. Message understanding conference- 6: A brief history. In Proceedings of the 16th International Conference on Computational Linguistics, COLING 1996, Center for Sprogteknologi, Copenhagen, Denmark, 5–9 August 1996; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 1996; pp. 466–471.
2. Borthwick, A. A Maximum Entropy Approach to Named Entity Recognition. Ph.D. Thesis, New York University, New York, NY, USA, 1999.
3. Pyysalo, S.; Ginter, F.; Heimonen, J.; Björne, J.; Boberg, J.; Järvinen, J.; Salakoski, T. BioInfer: A corpus for information extraction in the biomedical domain. *BMC Bioinform.* **2007**, *8*, 50, doi:10.1186/1471-2105-8-50.
4. Li, F.; Zhang, M.; Fu, G.; Ji, D. A neural joint model for entity and relation extraction from biomedical text. *BMC Bioinform.* **2017**, *18*, 198, doi:10.1186/s12859-017-1609-9.
5. Oronoz, M.; Gojenola, K.; Pérez, A.; de Ilarraza, A.D.; Casillas, A. On the creation of a clinical gold standard corpus in Spanish: Mining adverse drug reactions. *J. Biomed. Inform.* **2015**, *56*, 318–332, doi:10.1016/j.jbi.2015.06.016.
6. Cho, H.; Lee, H. Biomedical named entity recognition using deep neural networks with contextual information. *BMC Bioinform.* **2019**, *20*, 735.
7. García Cumbreras, M.A.; Gonzalo, J.; Martínez Cámara, E.; Martínez Unanue, R.; Rosso, P.; Carrillo-de-Albornoz, J.; Montalvo, S.; Chiruzzo, L.; Collovini, S.; Guitiérrez, Y.; et al. (Eds.) *Iberian Languages Evaluation Forum Co-Located, Proceedings of the 35th Conference of the Spanish Society for Natural Language Processing (SEPLN 2019), Bilbao, Spain, 24 September 2019*; CEUR-WS.org: Aachen, Germany, 2019; Volume 2421.

8. Krallinger, M.; Krithara, A.; Nentidis, A.; Paliouras, G.; Villegas, M. BioASQ at CLEF2020: Large-scale biomedical semantic indexing and question answering. In *Advances in Information Retrieval, Proceedings of the 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, 14–17 April 2020*; Part II; Jose, J.M., Yilmaz, E., Magalhães, J., Castells, P., Ferro, N., Silva, M.J., Martins, F., Eds.; Springer: Cham, Switzerland, 2020; Volume 12036, pp. 550–556.

9. Marimon, M.; Gonzalez-Agirre, A.; Intxaurrondo, A.; Rodriguez, H.; Martin, J.L.; Villegas, M.; Krallinger, M. Automatic de-identification of medical texts in Spanish: The MEDDOCAN track, corpus, guidelines, methods and evaluation of results. In *Iberian Languages Evaluation Forum Co-Located, Proceedings of the 35th Conference of the Spanish Society for Natural Language Processing, IberLEF@SEPLN 2019, Bilbao, Spain, 24 September 2019*; Cumbreras, M.Á.G., Gonzalo, J., Cámara, E.M., Martínez-Unanue, R., Rosso, P., Carrillo-de-Albornoz, J., Montalvo, S., Chiruzzo, L., Collovini, S., Gutiérrez, Y., et al., Eds.; CEUR-WS.org: Aachen, Germany, 2019; Volume 2421, pp. 618–638.

10. Fundel, K.; Küffner, R.; Zimmer, R. RelEx—Relation extraction using dependency parse trees. *Bioinformatics* **2007**, *23*, 365–371, doi:10.1093/bioinformatics/btl616.

11. Banarescu, L.; Bonial, C.; Cai, S.; Georgescu, M.; Griffitt, K.; Hermjakob, U.; Knight, K.; Koehn, P.; Palmer, M.; Schneider, N. Abstract meaning representation for sembanking. In *LAW@ACL*; Dipper, S., Liakata, M., Pareja-Lora, A., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2013; pp. 178–186.

12. Arous, I.; Khayati, M. FashionBrain D2.1: Named Entity Recognition and Linking Methods v3.0. Deliverable of the H2020 Project. In *Understanding Europe's Fashion Data Universe*; FashionBrain (Grant Agreement No. 732328); Fashion Brain: Sheffield, UK, 2019.

13. Dozier, C.; Kondadadi, R.; Light, M.; Vachher, A.; Veeramachaneni, S.; Wudali, R. Named entity recognition and resolution in legal text. In *Semantic Processing of Legal Texts: Where the Language of Law Meets the Law of Language*; Francesconi, E., Montemagni, S., Peters, W., Tiscornia, D., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6036, pp. 27–43.

14. Colic, N.; Furrer, L.; Rinaldi, F. Annotating the pandemic: Named entity recognition and normalisation in COVID-19 literature. In *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020.

15. Gridach, M. Character-level neural network for biomedical named entity recognition. *J. Biomed. Inform.* **2017**, *70*, 85–91, doi:10.1016/j.jbi.2017.05.002.

16. Segura-Bedmar, I.; Martínez, P.; Herrero-Zazo, M. SemEval-2013 task 9: Extraction of drug-drug interactions from biomedical texts (DDIExtraction 2013). In Proceedings of the Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Seventh International Workshop on Semantic Evaluation (SemEval 2013), Atlanta, GA, USA, 14–15 June 2013; Association for Computational Linguistics: Stroudsburg, PA, USA, 14–15 June 2013; pp. 341–350.

17. Ju, Z.; Wang, J.; Zhu, F. Named entity recognition from biomedical text using SVM. In Proceedings of the 2011 5th International Conference on Bioinformatics and Biomedical Engineering, Wuhan, China, 10–12 May 2011; pp. 1–4.

18. Huang, L.; Ling, C. Representing multiword chemical terms through phrase-level preprocessing and word embedding. *ACS Omega* **2019**, *4*, 18510–18519.

19. Krallinger, M.; Leitner, F.; Rabal, O.; Vazquez, M.; Oyarzabal, J.; Valencia, A. CHEMDNER: The drugs and chemical names extraction challenge. *J. Cheminformatics* **2015**, *7*, S1, doi:10.1186/1758-2946-7-S1-S1.

20. Gonzalez-Agirre, A.; Marimon, M.; Intxaurrondo, A.; Rabal, O.; Villegas, M.; Krallinger, M. PharmaCoNER: Pharmacological Substances, Compounds and Proteins Named Entity Recognition Track. In Proceedings of the 5th Workshop on BioNLP Open Shared Tasks, BioNLP-OST@EMNLP-IJNCLP 2019, Hong Kong, China, 4 November 2019; Kim, J.-D., Nédellec, C., Bossy, R., Deléger, L., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 1–10.

21. Korkontzelos, I.; Piliouras, D.; Dowsey, A.W.; Ananiadou, S. Boosting drug named entity recognition using an aggregate classifier. *Artif. Intell. Med.* **2015**, *65*, 145–153, doi:10.1016/j.artmed.2015.05.007.

22. Ju, M.; Nguyen, N.T.H.; Miwa, M.; Ananiadou, S. An ensemble of neural models for nested adverse drug events and medication extraction with subwords. *J. Am. Med. Inform. Assoc.* **2020**, *27*, 22–30, doi:10.1093/jamia/ocz075.

23. Leaman, R.; Lu, Z. TaggerOne: Joint named entity recognition and normalization with semi-markov models. *Bioinformatics* **2016**, *32*, 2839–2846, doi:10.1093/bioinformatics/btw343.

24. Tang, B.; Feng, Y.; Wang, X.; Wu, Y.; Zhang, Y.; Jiang, M.; Wang, J.; Xu, H. A comparison of conditional random fields and structured support vector machines for chemical entity recognition in biomedical literature. *J. Cheminform.* **2015**, *7*, S8, doi:10.1186/1758-2946-7-S1-S8.

25. Si, Y.; Wang, J.; Xu, H.; Roberts, K.E. Enhancing clinical concept extraction with contextual embeddings. *J. Am. Med. Inform. Assoc.* **2019**, *26*, 1297–1304, doi:10.1093/jamia/ocz096.

26. Krantz, J.; Dulin, M.; Palma, P.D. Language-agnostic syllabification with neural sequence labeling. In Proceedings of the 18th IEEE International Conference on Machine Learning And Applications, ICMLA 2019, Boca Raton, FL, USA, 16–19 December 2019; Wani, M.A., Khoshgoftaar, T.M., Wang, D., Wang, H., Seliya, N., Eds.; IEEE: New York, NY, USA, 2019; pp. 804–810.

27. Nallapati, R.; Zhai, F.; Zhou, B. SummaRuNNer: A recurrent neural network based sequence model for extractive summarization of documents. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; Singh, S.P., Markovitch, S., Eds.; AAAI Press: Palo Alto, CA, USA, 2017; pp. 3075–3081.

28. Gamma, E.; Helm, R.; Johnson, R.E. *Design Patterns: Elements of Reusable Object-Oriented Software*, 1st ed.; Reprint; Addison-Wesley Longman: Amsterdam, The Netherlands, 1994; ISBN 0-201-63361-2.

29. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to Sequence learning with neural networks. In *Advances in Neural Information Processing Systems 27, Proceedings of the Annual Conference on Neural Information Processing Systems 2014, Montreal, QC, Canada, 8–13*

December 2014; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q., Eds.; NeurIPS Proceedings: Montreal, QC, Canada, 2014; pp. 3104–3112.

30. Greene, B.B.; Rubin, G.M. *Automatic Grammatical Tagging of English*; Department of Linguistics of Brown University: Providence, RI, USA, 1971.

31. Marshall, I. Choice of grammatical word-class without global syntactic analysis: Tagging words in the LOB corpus. *Comput. Humanit.* **1983**, *17*, 139–150, doi:10.1007/BF02259886.

32. Brill, E. A Simple rule-based part-of-speech tagger. In Proceedings of the Third Conference on Applied Natural Language Processing (ANLP'92), Trento, Italy, 31 March–3 April 1992; pp. 152–155.

33. Brants, T. TnT—A statistical part-of-speech tagger. In Proceedings of the Sixth Applied Natural Language Processing (ANLP-2000); Seattle, WA, USA, 29 April–4 May 2000; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA 2000.

34. Toutanova, K.; Klein, D.; Manning, C.D.; Singer, Y. Feature-rich part-of-speech tagging with a cyclic dependency network. In Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, Edmonton, AB, Canada, 27 May–1 June 2003; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2003; pp. 252–259.

35. Akbik, A.; Blythe, D.; Vollgraf, R. Contextual string embeddings for sequence labeling. In Proceedings of the 27th International Conference on Computational Linguistics (COLING), Santa Fe, NM, USA, 20–26 August 2018; pp. 1638–1649.

36. Abney, S.P. Parsing by chunks. In *Principle-Based Parsing: Computation and Psycholinguistics*; Berwick, R.C., Abney, S.P., Tenny, C., Eds.; Kluwer: Dordrecht, The Netherlands, 1991; pp. 257–278.

37. Ramshaw, L.A.; Marcus, M. Text chunking using transformation-based learning. In Proceedings of the Third Workshop on Very Large Corpora, VLC@ACL 1995, Cambridge, MA, USA, 30 June 1995; Yarowsky, D., Church, K., Eds.; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 1995.

38. Kudo, T.; Matsumoto, Y. Chunking with support vector machines. In Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics, Pittsburgh, PA, USA, 2–7 June 2001; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2001.

39. Huang, Z.; Xu, W.; Yu, K. Bidirectional LSTM-CRF Models for Sequence Tagging. *arXiv* **2015**, arXiv:1508.01991.

40. Lafferty, J.D.; McCallum, A.; Pereira, F.C.N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001), Williams College, Williamstown, MA, USA, 28 June–1 July 2001; Brodley, C.E., Danyluk, A.P., Eds.; Morgan Kaufmann: Burlington, MA, USA, 2001; pp. 282–289.

41. Collins, M. Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms. In Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, Philadelphia, PA, USA, 6–7 July 2002; Association for Computational Linguistics: Stroudsburg, PA, USA, 2002; Volume 10, pp. 1–8.

42. Choi, Y.; Cardie, C.; Riloff, E.; Patwardhan, S. Identifying sources of opinions with conditional random fields and extraction patterns. In Proceedings of the HLT-EMNLP 2005, Vancouver, BC, Canada, 6–8 October 2005; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2005; pp. 355–362.

43. Jakob, N.; Gurevych, I. Extracting opinion targets in a single and cross-domain setting with conditional random fields. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, Cambridge, MA, USA, 18–25 June 2010; Association for Computational Linguistics: Stroudsburg, PA, USA, 2010; pp. 1035–1045.

44. Ghosh, S.; Johansson, R.; Riccardi, G.; Tonelli, S. Shallow discourse parsing with conditional random fields. In Proceedings of the Fifth International Joint Conference on Natural Language Processing, IJCNLP 2011, Chiang Mai, Thailand, 8–13 November 2011; The Association for Computer Linguistics: Stroudsburg, PA, USA, 2011; pp. 1071–1079.

45. Yao, X.; Van Durme, B.; Callison-Burch, C.; Clark, P. Answer extraction as sequence tagging with tree edit distance. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Atlanta, GA, USA, 9–14 June 2013; Association for Computational Linguistics: Stroudsburg, PA, USA, 2013; pp. 858–867.

46. Frazee, J. Dynamic conditional random fields for semantic role labeling. In Proceedings of the Fifth Midwest Computational Linguistics Colloquium (MCLC-5), East Lansing, MI, USA, 10–11 May 2008.

47. Sikdar, U.K.; Gambäck, B. Language identification in code-switched text using conditional random fields and babelnet. In Proceedings of the Second Workshop on Computational Approaches to Code Switching, Austin, TX, USA, 1 November 2016; Association for Computational Linguistics: Stroudsburg, PA, USA, 2016; pp. 127–131.

48. Shen, D.; Sun, J.-T.; Li, H.; Yang, Q.; Chen, Z. Document summarization using conditional random fields. In Proceedings of the IJCAI 2007—20th International Joint Conference on Artificial Intelligence, Hyderabad, India, 6–12 January 2007; pp. 2862–2867.

49. Bikel, D.M.; Miller, S.; Schwartz, R.; Weischedel, R. Nymble: A high-performance learning name-finder. In Proceedings of the Fifth Conference on Applied Natural Language Processing, Washington, DC, USA, 31 March–3 April 1997; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 1997; pp. 194–201.

50. Curran, J.; Clark, S. Language independent ner using a maximum entropy tagger. In Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL, Edmonton, AB, Canada, 31 May–1 June 2003; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2003; pp. 164–167.

51. Xu, Z.; Qian, X.; Zhang, Y.; Zhou, Y. CRF-based hybrid model for word segmentation, NER and even POS tagging. In Proceedings of the Sixth SIGHAN Workshop on Chinese Language Processing, Hyderabad, India, 11–12 January 2008; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2008.

52. Tang, B.; Cao, H.; Wu, Y.; Jiang, M.; Xu, H. Clinical entity recognition using structural support vector machines with rich features. In Proceedings of the ACM Sixth International Workshop on Data and Text Mining in Biomedical Informatics, DTM-BIO@CIKM 2012, Maui, HI, USA, 29 October 2012; Kim, D., Ananiadou, S., Song, M., Xu, H., Eds.; ACM: New York, NY, USA, 2012; pp. 13–20.

53. Li, J.; Sun, A.; Han, J.; Li, C. A survey on deep learning for named entity recognition. *IEEE Trans. Knowl. Data Eng.* **2020**, 1. doi:10.1109/TKDE.2020.2981314.

54. Yadav, V.; Bethard, S. A survey on recent advances in named entity recognition from deep learning models. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, NM, USA, 20–26 August 2018; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 2145–2158.

55. Zhu, Q.; Li, X.; Conesa, A.; Pereira, C. GRAM-CNN: A deep learning approach with local context for named entity recognition in biomedical text. *Bioinformatics* **2018**, *34*, 1547–1554, doi:10.1093/bioinformatics/btx815.

56. Jaiswal, A.K.; Tiwari, P.; Garg, S.; Hossain, M.S. Entity-aware capsule network for multi-class classification of big data: A deep learning approach. *Future Gener. Comput. Syst.* **2021**, *117*, 1–11, doi:10.1016/j.future.2020.11.012.

57. Zhao, W.; Peng, H.; Eger, S.; Cambria, E.; Yang, M. Towards scalable and reliable capsule networks for challenging NLP applications. In Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, 28 July–2 August 2019; Korhonen, A., Traum, D.R., Màrquez, L., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; Volume 1, pp. 1549–1559.

58. Zhou, D.; Miao, L.; He, Y. Position-aware deep multi-task learning for drug-drug interaction extraction. *Artif. Intell. Med.* **2018**, *87*, 1–8, doi:10.1016/j.artmed.2018.03.001.

59. Li, F.; Zhang, B.; Gao, D. Chinese named entity recognition for hazard and operability analysis text. In Proceedings of the 2020 Chinese Control and Decision Conference (CCDC), Hefei, China, 20–24 August 2020; IEEE: New York, NY, USA, 2020; pp. 374–378.

60. Misawa, S.; Taniguchi, M.; Miura, Y.; Ohkuma, T. Character-based bidirectional LSTM-CRF with words and characters for japanese named entity recognition. In Proceedings of the First Workshop on Subword and Character Level Models in NLP, Copenhagen, Denmark, 7 September 2017; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 97–102.

61. Grishman, R.; Sundheim, B. Design of the MUC-6 evaluation. In Proceedings of the 6th Conference on Message Understanding, Columbia, MD, USA, 6–8 November 1995; Association for Computational Linguistics: Stroudsburg, PA, USA, 1995; pp. 1–11.

62. Chinchor, N.A. Overview of MUC-7. In Proceedings of the Seventh Message Understanding Conference (MUC-7), Fairfax, VA, USA, 29 April–1 May 1998; NIST: Gaithersburg, MD, USA, 1998.

63. Kilgarriff, A.; Palmer, M. (Eds.). *Pilot SENSEVAL*; Association for Computational Linguistics: Stroudsburg, PA, USA, 1998.

64. Sekine, S.; Isahara, H. IREX: IR & IE evaluation project in Japanese. In Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00), Athens, Greece, 31 May–2 June 2000; European Language Resources Association (ELRA): Stroudsburg, PA, USA, 2000.

65. Tjong Kim Sang, E.F. Introduction to the CoNLL-2002 shared task: Language-independent named entity recognition. In Proceedings of the CoLING-02: The 6th Conference on Natural Language Learning 2002 (CoNLL-2002), Taipei, Taiwan, 31 August–1 September 2002; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2002.

66. Tjong Kim Sang, E.F.; De Meulder, F. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003, Edmonton, AB, Canada, 31 May–1 June 2003; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2003; pp. 142–147.

67. Doddington, G.; Mitchell, A.; Przybocki, M.; Ramshaw, L.; Strassel, S.; Weischedel, R. The Automatic Content Extraction (ACE) program—Tasks, data, and evaluation. In Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04), Lisbon, Portugal, 26–28 May 2004; European Language Resources Association (ELRA): Stroudsburg, PA, USA, 2004.

68. Hirschman, L.; Yeh, A.S.; Blaschke, C.; Valencia, A. Overview of BioCreAtIvE: Critical assessment of information extraction for biology. *BMC Bioinform.* **2005**, *6* (Suppl. S1), S1 doi:10.1186/1471-2105-6-S1-S1.

69. Li, J.; Sun, Y.; Johnson, R.J.; Sciaky, D.; Wei, C.-H.; Leaman, R.; Davis, A.P.; Mattingly, C.J.; Wiegers, T.C.; Lu, Z. BioCreative V CDR Task Corpus: A Resource for Chemical Disease Relation Extraction. *Database J. Biol. Databases Curation* **2016**, *2016*, baw068, doi:10.1093/database/baw068.

70. Tsujii, J. (Ed.) *BioNLP 2009 Workshop Companion Volume for Shared Task*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2009.

71. Kim, J.-D.; Pyysalo, S.; Ohta, T.; Bossy, R.; Nguyen, N.; Tsujii, J. Overview of BioNLP shared task 2011. In Proceedings of BioNLP Shared Task 2011 Workshop, Portland, OR, USA, 24 June 2011; Association for Computational Linguistics: Stroudsburg, PA, USA, 2011; pp. 1–6.

72. Nédellec, C.; Bossy, R.; Kim, J.-D.; Kim, J.; Ohta, T.; Pyysalo, S.; Zweigenbaum, P. Overview of BioNLP shared task 2013. In Proceedings of the BioNLP Shared Task 2013 Workshop, Sofia, Bulgaria, 9 August 2013; Association for Computational Linguistics: Stroudsburg, PA, USA, 2013; pp. 1–7.

73. Nédellec, C.; Bossy, R.; Kim, J.-D. (Eds.) *4th BioNLP Shared Task Workshop*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2016.

74.　Ben Abacha, A.; Shivade, C.; Demner-Fushman, D. Overview of the MEDIQA 2019 shared task on textual inference, question entailment and question answering. In Proceedings of the 18th BioNLP Workshop and Shared Task, Florence, Italy, 1 August 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 370–379.

75.　Collier, N.; Park, H.S.; Ogata, N.; Tateishi, Y.; Nobata, C.; Ohta, T.; Sekimizu, T.; Imai, H.; Ibushi, K.; Tsujii, J. The GENIA project: Corpus-based knowledge acquisition and information extraction from genome research papers. In Proceedings of the Ninth Conference of the European Chapter of the Association for Computational Linguistics, Bergen, Norway, 8–12 June 1999; Association for Computational Linguistics: Stroudsburg, PA, USA, 1999; pp. 271–272.

76.　Suominen, H.; Salanterä, S.; Velupillai, S.; Chapman, W.W.; Savova, G.K.; Elhadad, N.; Pradhan, S.; South, B.R.; Mowery, D.L.; Jones, G.J.F.; et al. Overview of the ShARe/CLEF eHealth evaluation lab 2013. In *Information Access Evaluation. Multilinguality, Multimodality, and Visualization, Proceedings of the 4th International Conference of the CLEF Initiative, CLEF 2013, Valencia, Spain, 23–26 September 2013*; Forner, P., Müller, H., Paredes, R., Rosso, P., Stein, B., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8138, pp. 212–231.

77.　Benikova, D.; Biemann, C.; Kisselew, M.; Pado, S. GermEval 2014 named entity recognition shared task: Companion paper. In *GermEval 2014 NER Shared Task*; Institut für Informationswissenschaft und Sprachtechnologie: Hildesheim, Germany, 2014; pp. 104–112.

78.　Xu, W.; Han, B.; Ritter, A. (Eds.) *Workshop on Noisy User-Generated Text*; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA, 2015.

79.　Chen, H.; Ji, H.; Sun, L.; Wang, H.; Qian, T.; Ruan, T. (Eds.). *Knowledge Graph and Semantic Computing: Semantic, Knowledge, and Linked Big Data—First China Conference, CCKS 2016, Beijing, China, 19–22 September 2016, Revised Selected Papers*; Communications in Computer and Information Science; Springer: Singapore: 2016; Volume 650, ISBN 978-981-10-3167-0.

80.　Piskorski, J.; Pivovarova, L.; Šnajder, J.; Steinberger, J.; Yangarber, R. The first cross-lingual challenge on recognition, normalization, and matching of named entities in Slavic Languages. In Proceedings of the 6th Workshop on Balto-Slavic Natural Language Processing, Valencia, Spain, 4 April 2017; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 76–85.

81.　Piskorski, J.; Laskova, L.; Marcińczuk, M.; Pivovarova, L.; Přibáň, P.; Steinberger, J.; Yangarber, R. The second cross-lingual challenge on recognition, normalization, classification, and linking of named entities across Slavic Languages. In Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing, Florence, Italy, 2 August 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 63–74.

82.　VLSP. *Fifth International Workshop on Vietnamese Language and Speech Processing*; Association for Vietnamese Language and Speech Processing: Hanoi, Vietnam, 2018.

83.　Taghizadeh, N.; Borhanifard, Z.; Pour, M.G.; Farhoodi, M.; Mahmoudi, M.; Azimzadeh, M.; Faili, H. NSURL-2019 Task 7: Named entity recognition for Farsi. In Proceedings of the First International Workshop on NLP Solutions for Under Resourced Languages (NSURL 2019) Co-Located with ICNLSP 2019—Short Papers, Trento, Italy, 11–12 September 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 9–15.

84.　Intxaurrondo, A.; Pérez-Pérez, M.; Rodríguez, G.P.; López-Martín, J.A.; Santamaría, J.; de la Peña, S.; Villegas, M.; Akhondi, S.A.; Valencia, A.; Lourenço, A.; et al. The Biomedical Abbreviation Recognition and Resolution (BARR) track: Benchmarking, evaluation and importance of abbreviation recognition systems applied to Spanish biomedical abstracts. In Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017) Co-Located with 33th Conference of the Spanish Society for Natural Language Processing (SEPLN 2017), Murcia, Spain, 19 September 2017; Martínez, R., Gonzalo, J., Rosso, P., Montalvo, S., de Albornoz, J.C., Eds.; CEUR-WS.org: Aachen, Germany, 2017; Volume 1881, pp. 230–246.

85.　Intxaurrondo, A.; Marimon, M.; Gonzalez-Agirre, A.; López-Martín, J.A.; Rodriguez, H.; Santamaría, J.; Villegas, M.; Krallinger, M. Finding mentions of abbreviations and their definitions in Spanish clinical cases: The BARR2 shared task evaluation results. In Proceedings of the Third Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018) Co-Located with 34th Conference of the Spanish Society for Natural Language Processing (SEPLN 2018), Sevilla, Spain, 18 September 2018; Rosso, P., Gonzalo, J., Martínez, R., Montalvo, S., de Albornoz, J.C., Eds.; CEUR-WS.org: Aachen, Germany, 2018; Volume 2150, pp. 280–289.

86.　Collovini, S.; Neto, J.F.S.; Consoli, B.S.; Terra, J.; Vieira, R.; Quaresma, P.; Souza, M.; Claro, D.B.; Glauber, R. IberLEF 2019 portuguese named entity recognition and relation extraction tasks. In Proceedings of the Iberian Languages Evaluation Forum Co-Located with 35th Conference of the Spanish Society for Natural Language Processing, IberLEF@SEPLN 2019, Bilbao, Spain, 24 September 2019; Cumbreras, M.Á.G., Gonzalo, J., Cámara, E.M., Martínez-Unanue, R., Rosso, P., Carrillo-de-Albornoz, J., Montalvo, S., Chiruzzo, L., Collovini, S., Gutiérrez, Y.; et al., Eds.; CEUR-WS.org: Aachen, Germany, 2019; Volume 2421, pp. 390–410.

87.　Porta-Zamorano, J.; Espinosa-Anke, L. Overview of CAPITEL shared tasks at IberLEF 2020: Named entity recognition and universal dependencies parsing. In *Iberian Languages Evaluation Forum (IberLEF 2020) Co-Located, Proceedings of the 36th Conference of the Spanish Society for Natural Language Processing (SEPLN 2020), Malaga, Spain, 23–25 September 2020*; CEUR-WS.org: Aachen, Germany, 2020; Volume 2664, pp. 31–38.

88.　Piad-Morffis, A.; Gutiérrez, Y.; Consuegra-Ayala, J.P.; Estevez-Velarde, S.; Almeida-Cruz, Y.; Muñoz, R.; Montoyo, A. Overview of the eHealth knowledge discovery challenge at IberLEF 2019. In *Iberian Languages Evaluation Forum Co-Located, Proceedings of the 35th Conference of the Spanish Society for Natural Language Processing, IberLEF@SEPLN 2019, Bilbao, Spain, 24 September 2019*;

Cumbreras, M.Á.G., Gonzalo, J., Cámara, E.M., Martínez-Unanue, R., Rosso, P., Carrillo-de-Albornoz, J., Montalvo, S., Chiruzzo, L., Collovini, S., Gutiérrez, Y., et al., Eds.; CEUR-WS.org: Aachen, Germany, 2019; Volume 2421, pp. 1–16.

89. Kim, J.-D.; Ohta, T.; Tateisi, Y.; Tsujii, J. GENIA corpus—A semantically annotated corpus for bio-textmining. In Proceedings of the Eleventh International Conference on Intelligent Systems for Molecular Biology, Brisbane, Australia. 29 June–3 July 2003; pp. 180–182.

90. Hovy, E.H.; Marcus, M.P.; Palmer, M.; Ramshaw, L.A.; Weischedel, R.M. OntoNotes: The 90% solution. In Proceedings of the Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, New York, NY, USA, 4–9 June 2006; Moore, R.C., Bilmes, J.A., Chu-Carroll, J., Sanderson, M., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2006.

91. Krallinger, M.; Rabal, O.; Leitner, F.; Vazquez, M.; Salgado, D.; Lu, Z.; Leaman, R.; Lu, Y.; Ji, D.; Lowe, D.M.; et al. The CHEMDNER corpus of chemicals and drugs and its annotation principles. *J. Cheminform.* **2015**, *7*, S2, doi:10.1186/1758-2946-7-S1-S2.

92. Dogan, R.I.; Leaman, R.; Lu, Z. NCBI disease corpus: A resource for disease name recognition and concept normalization. *J. Biomed. Inform.* **2014**, *47*, 1–10, doi:10.1016/j.jbi.2013.12.006.

93. Neudecker, C. An open corpus for named entity recognition in historic newspapers. In Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, 23–28 May 2016; Calzolari, N., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., et al., Eds.; European Language Resources Association (ELRA): Luxemburg, 2016.

94. Luz de Araujo, P.H.; de Campos, T.E.; de Oliveira, R.R.R.; Stauffer, M.; Couto, S.; Bermejo, P. LeNER-Br: A dataset for named entity recognition in Brazilian legal text. In *Computational Processing of the Portuguese Language*; Villavicencio, A., Moreira, V., Abad, A., Caseli, H., Gamallo, P., Ramisch, C., Gonçalo Oliveira, H., Paetzold, G.H., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 313–323.

95. Yu, J.; Bohnet, B.; Poesio, M. Named entity recognition as dependency parsing. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 6470–6476.

96. Peng, Y.; Yan, S.; Lu, Z. Transfer learning in biomedical natural language processing: An evaluation of BERT and ELMo on ten benchmarking datasets. In Proceedings of the 18th BioNLP Workshop and Shared Task, BioNLP@ACL 2019, Florence, Italy, 1 August 2019; Demner-Fushman, D., Cohen, K.B., Ananiadou, S., Tsujii, J., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 58–65.

97. Miranda-Escalada, A.; Farré, E.; Krallinger, M. Named entity recognition, concept normalization and clinical coding: Overview of the Cantemist track for cancer text mining in Spanish, corpus, guidelines, methods and results. In *Iberian Languages Evaluation Forum (IberLEF 2020) Co-Located, Proceedings of the 36th Conference of the Spanish Society for Natural Language Processing (SEPLN 2020), Málaga, Spain, 23 September 2020*; Cumbreras, M.Á.G., Gonzalo, J., Cámara, E.M., Martínez-Unanue, R., Rosso, P., Zafra, S.M.J., Zambrano, J.A.O., Miranda, A., Zamorano, J.P., Gutiérrez, Y., et al., Eds.; CEUR-WS.org: Aachen, Germany, 2020; Volume 2664, pp. 303–323.

98. Tabassum, J.; Xu, W.; Ritter, A. WNUT-2020 task 1 overview: Extracting entities and relations from wet lab protocols. In Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020), Online, 19 November 2020; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 260–267.

99. Glavas, G.; Vulic, I. Is supervised syntactic parsing beneficial for language understanding? An empirical investigation. *arXiv* **2020**, arXiv:2008.06788.

100. Yang, B.; Wong, D.F.; Chao, L.S.; Zhang, M. Improving tree-based neural machine translation with dynamic lexicalized dependency encoding. *Knowl. Based Syst.* **2020**, *188*, 105042, doi:10.1016/j.knosys.2019.105042.

101. Zhang, B.; Zhang, Y.; Wang, R.; Li, Z.; Zhang, M. Syntax-aware opinion role labeling with dependency graph convolutional networks. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 3249–3258.

102. Nguyen, D.Q.; Verspoor, K. From POS tagging to dependency parsing for biomedical event extraction. *BMC Bioinform.* **2019**, *20*, 72:1–72:13, doi:10.1186/s12859-019-2604-0.

103. Cao, Q.; Liang, X.; Li, B.; Lin, L. Interpretable visual question answering by reasoning on dependency trees. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, 1. doi:10.1109/TPAMI.2019.2943456.

104. Balachandran, V.; Pagnoni, A.; Lee, J.Y.; Rajagopal, D.; Carbonell, J.G.; Tsvetkov, Y. StructSum: Incorporating latent and explicit sentence dependencies for single document summarization. *arXiv* **2020**, arXiv:2003.00576.

105. Takase, S.; Suzuki, J.; Okazaki, N.; Hirao, T.; Nagata, M. Neural headline generation on abstract meaning representation. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; Association for Computational Linguistics: Stroudsburg, PA, USA, 2016; pp. 1054–1059.

106. Song, L.; Gildea, D.; Zhang, Y.; Wang, Z.; Su, J. Semantic neural machine translation using AMR. *Trans. Assoc. Comput. Linguist.* **2019**, *7*, 19–31, doi:10.1162/tacl_a_00252.

107. Dozat, T.; Manning, C.D. Deep biaffine attention for neural dependency parsing. In Proceedings of the 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, 24–26 April 2017.

108. Fernández-González, D.; Gómez-Rodríguez, C. Left-to-right dependency parsing with pointer networks. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MI, USA, 2–7 June 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 710–716.

109. Fernández-González, D.; Gómez-Rodríguez, C. Discontinuous constituent parsing with pointer networks. In Proceedings of the the Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, 7–12 February 2020; AAAI Press: Palo Alto, CA, USA, 2020; pp. 7724–7731.

110. Zhang, Y.; Nivre, J. Transition-based dependency parsing with rich non-local features. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA, 19–24 June 2011; Association for Computational Linguistics: Stroudsburg, PA, USA, 2011; pp. 188–193.

111. Shi, T.; Huang, L.; Lee, L. Fast(Er) exact decoding and global training for transition-based dependency parsing via a minimal feature set. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, 9–11 September 2017; Palmer, M., Hwa, R., Riedel, S., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 12–23.

112. Gómez-Rodríguez, C.; Carroll, J.; Weir, D. Dependency parsing schemata and mildly non-projective dependency parsing. *Comput. Linguist.* **2011**, *37*, 541–586, doi:10.1162/COLI_a_00060.

113. Gómez-Rodríguez, C.; Shi, T.; Lee, L.; Pado global transition-based non-projective dependency parsing. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 15–20 July 2018; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 2664–2675.

114. Liu, J.; Zhang, Y. In-order transition-based constituent parsing. *Trans. Assoc. Comput. Linguist.* **2017**, *5*, 413–424, doi:10.1162/tacl_a_00070.

115. Fernández-González, D.; Gómez-Rodríguez, C. Faster shift-reduce constituent parsing with a non-binary, bottom-up strategy. *Artif. Intell.* **2019**, *275*, 559–574, doi:10.1016/j.artint.2019.07.006.

116. Goldberg, Y.; Hirst, G. *Neural Network Methods in Natural Language Processing*; Morgan & Claypool Publishers: San Rafael, CA, USA, 2017; ISBN 1-62705-298-4.

117. Joshi, M.; Penstein-Rosé, C. Generalizing dependency features for opinion mining. In Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, Singapore, 2–7 August 2009; Association for Computational Linguistics: Stroudsburg, PA, USA, 2009; pp. 313–316.

118. Vilares, D.; Alonso, M.A.; Gómez-Rodríguez, C. On the usefulness of lexical and syntactic processing in polarity classification of twitter messages. *J. Assoc. Inf. Sci. Technol.* **2015**, *66*, 1799–1816, doi:10.1002/asi.23284.

119. Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C.D.; Ng, A.Y.; Potts, C. Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18–21 October 2013; Association for Computational Linguistics: Stroudsburg, PA, USA, 2013; pp. 1631–1642.

120. Vilares, D.; Gómez-Rodríguez, C.; Alonso, M.A. Universal, unsupervised (rule-based), uncovered sentiment analysis. *Knowl. Based Syst.* **2017**, *118*, 45–55, doi:10.1016/j.knosys.2016.11.014.

121. Gómez-Rodríguez, C.; Vilares, D. Constituent parsing as sequence labeling. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 1314–1324.

122. Vilares, D.; Abdou, M.; Søgaard, A. Better, faster, stronger sequence tagging constituent parsers. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MI, USA, 2–7 June 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 3372–3383.

123. Strzyz, M.; Vilares, D.; Gómez-Rodríguez, C. Viable dependency parsing as sequence labeling. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MI, USA, 2–7 June 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 717–723.

124. Strzyz, M.; Vilares, D.; Gómez-Rodríguez, C. Sequence labeling parsing by learning across representations. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 5350–5357.

125. Schwartz, R.; Dodge, J.; Smith, N.A.; Etzioni, O. Green AI. *arXiv* **2019**, arXiv:1907.10597.

126. Sasano, R.; Kurohashi, S. Japanese named entity recognition using structural natural language processing. In Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-II, Jeju Island, Korea, 11–13 October 2008; Association for Computational Linguistics: Stroudsburg, PA, USA, 2008.

127. Ling, X.; Weld, D.S. Fine-grained entity recognition. In Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, ON, Canada, 22–26 July 2012; Hoffmann, J., Selman, B., Eds.; AAAI Press: Palo Alto, CA, USA, 2012.

128. Luo, L.; Yang, Z.; Yang, P.; Zhang, Y.; Wang, L.; Lin, H.; Wang, J. An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition. *Bioinformatics* **2018**, *34*, 1381–1388, doi:10.1093/bioinformatics/btx761.

129. Tian, Y.; Shen, W.; Song, Y.; Xia, F.; He, M.; Li, K. Improving biomedical named entity recognition with syntactic information. *BMC Bioinform.* **2020**, preprint.

130. Miller, A.; Fisch, A.; Dodge, J.; Karimi, A.-H.; Bordes, A.; Weston, J. Key-value memory networks for directly reading documents. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; Association for Computational Linguistics: Stroudsburg, PA, USA, 2016; pp. 1400–1409.

131. Lee, J.; Yoon, W.; Kim, S.; Kim, D.; Kim, S.; So, C.H.; Kang, J. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* **2020**, *36*, 1234–1240, doi:10.1093/bioinformatics/btz682.

132. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MI, USA, 2–7 June 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 4171–4186.

133. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *arXiv* **2020**, arXiv:2005.14165.

134. Sachan, D.S.; Zhang, Y.; Qi, P.; Hamilton, W. Do syntax trees help pre-trained transformers extract information? *arXiv* **2020**, arXiv:2008.09084.

135. Shi, Z.; Sarkar, A.; Popowich, F. Simultaneous identification of biomedical named-entity and functional relation using statistical parsing techniques. In Proceedings of the Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers, Rochester, New York, NY, USA, 22–27 April 2007; Association for Computational Linguistics: Stroudsburg, PA, USA, 2007; pp. 161–164.

136. Finkel, J.R.; Manning, C.D. Joint parsing and named entity recognition. In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Boulder, CO, USA, 31 May–5 June 2009; Association for Computational Linguistics: Stroudsburg, PA, USA, 2009; pp. 326–334.

137. Jie, Z.; Muis, A.O.; Lu, W. Efficient dependency-guided named entity recognition. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; Singh, S.P., Markovitch, S., Eds.; AAAI Press: Palo Alto, CA, USA, 2017; pp. 3457–3465.

138. Spoustová, D.J.; Spousta, M. Dependency parsing as a sequence labeling task. *Prague Bull. Math. Linguist.* **2010**, *94*, 7–14.

139. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In Proceedings of the Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013, Lake Tahoe, NV, USA, 5–8 December 2013; pp. 3111–3119.

140. Cross, J.; Huang, L. Incremental parsing with minimal features using bi-directional LSTM. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Berlin, Germany, 7–12 August 2016; Association for Computational Linguistics: Stroudsburg, PA, USA, 2016; pp. 32–37.

141. Anderson, M.; Gómez-Rodríguez, C. Distilling neural networks for greener and faster dependency parsing. In *IWPT 2020*; Bouma, G., Matsumoto, Y., Oepen, S., Sagae, K., Seddah, D., Sun, W., Søgaard, A., Tsarfaty, R., Zeman, D., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 2–13.

142. Li, C. OpenAI's GPT-3 Language Model: A Technical Overview. Blog Post. 2020. Available online: https://lambdalabs.com/blog/demystifying-gpt-3/ (accessed on 4 January 2021).

143. Fernández-González, D.; Gómez-Rodríguez, C. Enriched in-order linearization for faster sequence-to-sequence constituent parsing. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 4092–4099.

144. Nadeau, D.; Sekine, S. A survey of named entity recognition and classification. *Lingvisticæ Investig.* **2007**, *30*, 3–26, doi:10.1075/li.30.1.03nad.

145. Vazquez, M.; Krallinger, M.; Leitner, F.; Valencia, A. Text mining for drugs and chemical compounds: Methods, tools and applications. *Mol. Inform.* **2011**, *30*, 506–519, doi:10.1002/minf.201100005.

146. Eltyeb, S.; Salim, N. Chemical named entities recognition: A review on approaches and applications. *J. Cheminform.* **2014**, *6*, 17, doi:10.1186/1758-2946-6-17.

147. Campos, D.; Matos, S.; Oliveira, J.L. Biomedical named entity recognition: A survey of machine-learning tools. In *Theory and Applications for Advanced Text Mining*; Sakurai, S., Ed.; IntechOpen: Rijeka, Croatia, 2012.

148. Alshaikhdeeb, B.; Ahmad, K. Biomedical named entity recognition: A review. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2016**, *4*, 889–895.

149. Marrero, M.; Urbano, J.; Sánchez-Cuadrado, S.; Morato, J.; Gómez-Berbís, J.M. Named entity recognition: Fallacies, challenges and opportunities. *Comput. Stand. Interfaces* **2013**, *35*, 482–489, doi:10.1016/j.csi.2012.09.004.

150. Shaalan, K. A survey of Arabic named entity recognition and classification. *Comput. Linguist.* **2014**, *40*, 469–510, doi:10.1162/COLI_a_00178.

151. Goyal, A.; Gupta, V.; Kumar, M. Recent named entity recognition and classification techniques: A systematic review. *Comput. Sci. Rev.* **2018**, *29*, 21–43, doi:10.1016/j.cosrev.2018.06.001.

152. Névéol, A.; Dalianis, H.; Velupillai, S.; Savova, G.; Zweigenbaum, P. Clinical natural language processing in languages other than English: Opportunities and challenges. *J. Biomed. Semant.* **2018**, *9*, 12:1–12:13, doi:10.1186/s13326-018-0179-8.

153. Hahn, U.; Oleynik, M. Medical information extraction in the age of deep learning. *Yearbook Med. Inform.* **2020**, *29*, 208–220, doi:10.1055/s-0040-1702001.

154. Minaee, S.; Kalchbrenner, N.; Cambria, E.; Nikzad, N.; Chenaghlu, M.; Gao, J. Deep learning based text classification: A comprehensive review. *arXiv* **2020**, arXiv:2004.03705.

155. Zhong, X.; Cambria, E.; Hussain, A. Extracting time expressions and named entities with constituent-based tagging schemes. *Cogn. Comput*. **2020**, *12*, 844–862, doi:10.1007/s12559-020-09714-8.

156. Yang, J.; Zhang, Y. NCRF++: An open-source neural sequence labeling toolkit. In Proceedings of ACL 2018, System Demonstrations, Melbourne, Australia, 15-20 July 2018; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 74–79.

157. Zhong, M.; Liu, P.; Wang, D.; Qiu, X.; Huang, X. Searching for effective neural extractive summarization: What works and what's next. In Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, 28 July–2 August 2019; Korhonen, A., Traum, D.R., Màrquez, L., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 1049–1058.

158. Li, X.; Bing, L.; Li, P.; Lam, W. A unified model for opinion target extraction and target sentiment prediction. In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, the Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, the Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, HI, USA, 27 January–1 February 2019; AAAI Press: Palo Alto, CA, USA, 2019; pp. 6714–6721.

159. Li, X.; Bing, L.; Zhang, W.; Lam, W. Exploiting BERT for end-to-end aspect-based sentiment analysis. In Proceedings of the 5th Workshop on Noisy User-Generated Text (W-NUT 2019), Hong Kong, China, 4 November 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 34–41.

160. He, H.; Choi, J.D. Establishing strong baselines for the new decade: Sequence tagging, syntactic and semantic parsing with BERT. In Proceedings of the Thirty-Third International Florida Artificial Intelligence Research Society Conference, North Miami Beach, FL, USA, 17–20 May 2020; Barták, R., Bell, E., Eds.; AAAI Press: Palo Alto, CA, USA, 2020; pp. 228–233.

161. Shaalan, K.; Raza, H. Arabic named entity recognition from diverse text types. In *Advances in Natural Language Processing, Proceedings of the 6th International Conference, GoTAL 2008, Gothenburg, Sweden, 25–27 August 2008*; Nordström, B., Ranta, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; Volume 5221, pp. 440–451.

162. Dashtipour, K.; Gogate, M.; Adeel, A.; Algarafi, A.; Howard, N.; Hussain, A. Persian named entity recognition. In Proceedings of the 16th IEEE International Conference on Cognitive Informatics & Cognitive Computing, ICCI*CC 2017, Oxford, UK, 26–28 July 2017; Howard, N., Wang, Y., Hussain, A., Hamdy, F., Widrow, B., Zadeh, L.A., Eds.; IEEE Computer Society: Washington, DC, USA, 2017; pp. 79–83.

163. Azpeitia, A.; Cuadros, M.; Gaines, S.; Rigau, G. NERC-Fr: Supervised named entity recognition for French. In *Text, Speech and Dialogue, Proceedings of the 17th International Conference, TSD 2014, Brno, Czech Republic, 8–12 September 2014*; Sojka, P., Horák, A., Kopecek, I., Pala, K., Eds.; Springer: Cham, Switzerland, 2014; Volume 8655, pp. 158–165.

164. Vilares, D.; Alonso, M.A.; Gómez-Rodríguez, C. Supervised sentiment analysis in multilingual environments. *Inf. Process. Manag*. **2017**, *53*, 595–607, doi:10.1016/j.ipm.2017.01.004.