Instituto Politécnico Nacional

Centro de Investigación en Computación

TESIS:

Unrestricted Bridging Anaphora Resolution using Lexical Information

QUE PARA OBTENER EL GRADO DE:

Maestría en Ciencias de la Computación

PRESENTA:

Ing. Jason Efrain Angel Gil

DIRECTOR DE TESIS:

Dr. Alexander Gelbukh

Roberto Navigli

Centro de Investigación en Computación
Instituto Politécnico Nacional

México, CDMX Diciembre 2019

SIP-14 bis

INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México siendo las 13:00 horas del día 11 del mes de octubre de 2019 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

Centro de Investigación en Computación

para examinar la tesis titulada:

"Unrestricted Bridging Anaphora Resolution using Lexical Information"

Presentada por el alumno:

ANGEL	**GIL**	**JASON EFRAIN**
Apellido paterno	Apellido materno	Nombre(s)

Con registro: | **B** | **1** | **7** | **1** | **1** | **4** | **3** |

aspirante de: **MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN**

Después de intercambiar opiniones los miembros de la Comisión manifestaron ***APROBAR LA TESIS***, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA
Directores de Tesis

Dr. Alexander Gelbukh

Dr. Roberto Navigli

Dr. Grigori Sidorov

Dr. Ildar Batyrshin

Dra. Olga Kolesnikova

Dr. Luis Manuel Vilches Blázquez

PRESIDENTE DEL COLEGIO DE PROFESORES

Dr. Marco Antonio Ramírez Salinas

INSTITUTO POLITECNICO NACIONAL
CENTRO DE INVESTIGACION
EN COMPUTACION
DIRECCION

INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México el día 02 de diciembre del año 2019, el que suscribe Jason Efrain Angel Gil alumno del Programa de Maestría en Ciencias de la Computación con número de registro B171143, adscrito al Centro de Investigación en Computación, manifiesta que es autor intelectual del presente trabajo de Tesis bajo la dirección del Dr. Alexander Gelbukh y Dr. Roberto Navigli y cede los derechos del trabajo titulado "Unrestricted Bridging Anaphora Resolution Using Lexical Information", al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección ajason08@gmail.com. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

Jason E. Angel

Jason Efrain Angel Gil

Firma

Resumen

El lenguaje natural es el medio de comunicación humana más utilizado, por lo tanto, juega un papel muy importante para una amplia gama de actividades humanas. En las últimas décadas, los avances en inteligencia artificial han abierto una puerta a la creación de máquinas capaces de procesar el lenguaje natural, ofreciendo resultados competitivos en comparación a los humanos en varias tareas, sin embargo, aún están muy lejos de lograr una comprensión profunda del discurso. Quizá la razón principal de esta dificultad se debe a ciertos aspectos del lenguaje que aún son demasiado complejos para ser modelados por medios computacionales, especialmente los niveles semántico y pragmático en textos más largos que una oración. Uno de esos aspectos son las anáforas indirectas, las cuales son utilizadas en casi cualquier discurso, y detectarlas es de alta relevancia para comprender los mensajes. En este trabajo enfrentamos el problema de resolver anáforas indirectas mediante la inclusión de información léxica, específicamente empleamos representaciones de sentidos (*sense embeddings*) para este fin. Dichas representaciones han demostrado su utilidad en varias tareas de PLN, sin embargo, no han sido utilizadas previamente en resolución de anáforas indirectas. Nuestros hallazgos muestran que la información léxica, aunque no es suficiente para resolver las anáforas, resulta necesaria para detectar anáforas indirectas presentes en escenarios sin restricciones, que son lingüísticamente más complejos y suelen ser ignorados en la literatura.

Abstract

Natural language is probably the most used human means of communication; hence it plays a very important role for a wide range of human activities. In recent decades advances in artificial intelligence have opened a door for creating machines able to process natural languages offering competitive results versus humans in several NLP tasks, yet they are still too far from achieving a deep understanding of discourse. The main reason for these shortcomings could be due to certain aspects of language which are still too complex to model by computational means, especially the semantic and pragmatic levels in texts longer than a sentence. One of those aspects are the bridging anaphoras, they are indirect references used in almost any discourse, and detect them is highly relevant to understand the communication. In this work we face the problem of bridging anaphora resolution by the inclusion of lexical information, specifically we employed sense embeddings to model it. Such embeddings are a special kind of word embedding which achieve good results in a variety of NLP tasks but have not been used before for bridging anaphora resolution. The results show that lexical information although not enough to solve anaphoras, is still necessary to detect complex bridging anaphora in unrestricted scenarios, which are linguistically more complex than the kind of anaphora usually found in most of bridging literature.

Acknowledgements

The time I shared with the people in the Centro de investigación en computación (CIC) has been the most rewarding of my life. I have learned and grown a lot as a professional, and as a person. This time was enough to realize that I really want to become a scientist in the NLP field. One of the outcomes is the present work, in which many people and organizations made their contribution. I would like to express my gratitude particularly to the following ones:

I thank my advisors during master's degree, *Dr. Alexander Gelbukh* for his invaluable research guidance and strategies about how to proceed and create opportunities, as well as his generosity and unconditional support. He is a very remarkable person with an interesting way of seeing the world. I thank *Dr. Roberto Navigli*, for all the conversations we hold about lexical semantics, because of him this thesis was possible. I am also grateful for his hospitality in Rome, and his advice about the researcher life. I hope to continue learning from them as their PhD student

A lot of appreciation to my good friend Segun Aroyehun, whom I admire as a person and as a researcher, and Dr. Grigori Sidorov for their irreplaceable support during the entire time of my master programme.

I thank my family, namely my parents, my sister, and my girlfriend, as well as my friend Alexis for their emotional and spiritual support. I thank my friend Mr. Jacob for everyday life advice, and unrestricted support.

Thanks very much to Consejo Nacional de Ciencia y Tecnología (CONACYT) for grants me a 2-years scholarship and Instituto Politécnico Nacional (IPN) for their tireless assistance at administrative and cultural aspects. I would also thank to this marvelous country for the kindness of their people and their focus on scientific research.

Content

List of figures

List of tables

1 Introduction

This is the First Section of this thesis, here we briefly describe what a bridging anaphora is and introduce the reader to this thesis, this Section starts by describing the problem we are interested in and the hypothesis we formulated for this purpose, this Section also discloses the objective, justification and motivation for the present development, as well as the contribution of this work. Finally, we offer a short description of the structure of this thesis for the succeeding Sections.

1.1 Background

Commonly discourses messages are expressed with dependency of previously mentioned ideas and concepts, or they occur in total isolation, i.e., the meaning of a new message is in function of previous interpretations, then, in order to achieve a good understanding of discourse it is necessary to infer the relationships between words and phrases. It occurs that textual inference is easy to perform for humans but a very hard task for machines, this may be due to machines lack of world knowledge, as well as reasoning systems to process such knowledge. Particularly exist a pragmatic inference called bridging anaphora which we are interested in, this is a special kind of anaphora whose references link to a previous entity or phrase in implicit or indirect fashion, that is to say, a bridging relation does not express identity or direct relations (better known as coreference). Instead, a bridging relation connects two textual expressions holding some kind of association. In synthesis, a bridging anaphora does not express explicitly its antecedent, and capture this indirect relation it is an indispensable step to achieve a fully understanding of discourse.

May be due to the abstraction and complexity of bridging anaphora as a linguistic phenomenon, it has been called by several different names throughout the years, namely “associative anaphora” (Hawkins, 1978), “inferrables” (Prince, 1981), “implicit anaphora” (Saeboe, 1996), “indirect anaphora” (Gelbukh and Sidorov, 1999). The term “bridging” has been used more frequently in recent years. The following are some examples of bridging anaphora[1].

[1] Here, and for the rest of the document, the bridging anaphora will appear in bold type, whereas its antecedent (the referent) will appear underlined.

(1.1) Lawmakers upheld controversial agreements made by <u>a House-Senate</u> conference earmarking community development funds for more than 40 projects backed by **often influential members**.

(1.2) <u>Vegetables</u> are abundant and full of flavor in Poland the pickles and sauerkraut sublime **the state monopolies** long broken.

Therefore, in (1.1) "often influential members" indirectly refers to "a House Senate" since the members are part of the House Senate. Likewise, in (2.2) “the state monopolies” indirectly refers to “vegetables”.

We will call a mention to any word or phrase already said in the discourse, it can include nouns, verbs, pronouns and connectives, as well as their phrase counterparts like nominal and verbal phrases. So, a bridging anaphora is solved when its antecedent is determined among all the possible mentions; such possibilities (including the real reference) are called candidates. However, since grammatical phrases usually overlap between them, it is quite common to find bridging anaphoras having an extensive list of candidates, especially for long documents. The explosion of candidates can be considered the main difficulty for solve the anaphor. For instance, “a house-senate conference” is another candidate for example (1.1), which, in spite of its similarity with “a house-senate” is not the antecedent for the aforementioned anaphor. Another useful term to introduce for this work is "bridging pair", which for this work indicates the twosome, i.e., the bridging anaphora with any of its correspondent candidates. In this context, <"often influential members" - "House-Senate"> is the bridging pair for example (1.1). Finally, we must point out that most of the bridging examples cited below are intentionally selected to be short for clarity of the explanations, these examples have a distance of at most 2 sentences between the bridging anaphora and its antecedent. However, longer distance are quite common in real scenarios.

Full bridging anaphora resolution consist of two subtasks: bridging anaphora identification and bridging antecedent selection. The former consists of identifying mentions which potentially are a bridging anaphora, whereas the latter consists of identifying the entity or phrase that anaphora is referring to. This separation suggests that an anaphora can be detected without point out the reference antecedent, such an approach is shown in Hou (2013a).

1.2 Problem formulation

This thesis focuses on the latter subtask, bridging antecedent selection, starting from a list of mentions (words or phrases) and bridging anaphora annotations freely available in ISNotes dataset (Hou et al. 2013b). To this end, a bridging antecedent solver should consider all the previous mentions as candidates for a given bridging annotation and determine which of them is the antecedent (or correct candidate). In spite of the simplicity of the statement, this step causes several challenges, such as discriminate between relative identical candidates because of the overlapping or the coreference between them, further to this the search problem becomes even more demanding inasmuch as the length of the document increases.

Formally, bridging relations are defined by a relation between mentions and bridging anaphoras. Let $A = \{A1,\ A2,\ \ldots Ai\}$ be the set of mentions for a given document, and $B = \{B1,\ B2,\ \ldots Bj\}$ the set of bridging anaphoras, that refer to candidates antecedent Ai forming a bridging pair (Ai, Bj) where $Ai \in A$ and $Bj \in B$. The next image describes graphically some bridging relations for a given document.

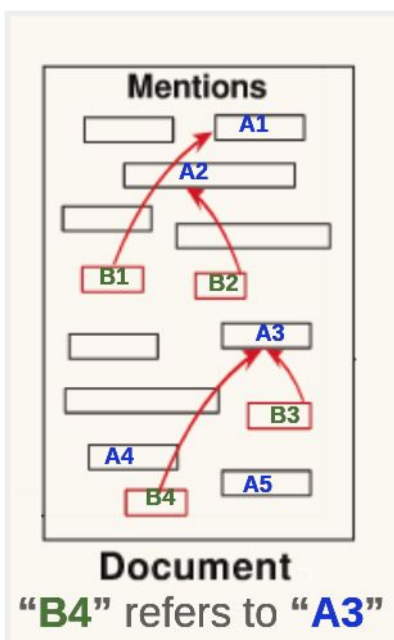

Figure 1. Graphical view of bridging relations in a document

Anaphoras in general follow a many to many cardinality, thus, an antecedent can be referenced by several different anaphoras and a single anaphora can perform references to different antecedents. Typically, mentions (in a bridging anaphora context) used to receive more references than those it gives (Hou et al. 2018a). Also a bridging anaphora can be the antecedent for another bridging anaphora, composing a chain of bridging anaphoras.

1.3 Hypothesis

In this work we plan to assess how beneficial is the use of lexical information for the task of bridging anaphora resolution using embedding representations.

1.4 Objectives

This thesis proposes the use of lexical information for modelling bridging anaphora resolution, specifically, bridging antecedent selection. The model leverages the lexical information included in sense embeddings to measures the association between a pair of vector representations which correspond to the bridging pair, in order to determine the bridging antecedent. To this end, the text is analysed by a Word Sense Disambiguation (WSD) and Entity Linking (EL) system called Babelfy, (Moro et al. 2014), which maps words and multiword expressions as concepts and entities in a lexical repository called BabelNet (Navigli and Ponzetto, 2010).

To evaluate the utility of include lexical information for solving bridging anaphora we compare the results against the embeddings provided by the state-of-the-art, which are specialized designed for this task, as well as other word embedding representations.

1.5 Justification

In spite of the many efforts done on anaphora resolution tasks, the actual progress is very limited. In fact, several approaches have been adapted across the years, starting off early works based on rule and hand-crafted features, (Poesio et al. 2004), to advances on specialized dense representation for anaphora (Hou et al. 2013b); even though the results indicate a slow progress in comparison to other NLP tasks. Furthermore, not many reliable datasets exist, and the language coverage is low. In spite of all these arguments, we believe that research on anaphora resolution tasks, like bridging resolution, deserves more attention for the sake of the progress on natural language processing, specifically we give consideration to the following reasons:

- Bridging anaphora could be the next cornerstone for text understanding, since it is one of the discourse devices responsible for maintaining coherence across the text (Irmer, 2011). Moreover, in general textual coherence is needed to perform language processes involving the discourse such as reading and writing in the way humans do.
- Bridging anaphora resolution could be a bridge for a better understanding of other natural language processing tasks, because of the close relation it shares with them (Rösiger et al. 2018b), such as tasks like relation extraction or aspect extraction for sentiment analysis (Poesio et al. 2016) can be seen as a generalization or a specialization respectively, also implicit semantic role labelling seems to share multiple features with bridging anaphora (Hou et al. 2018a).
- Bridging anaphora resolution hints multiple applications for downstream NLP tasks such as text generation by the use of coherence links (Soricut and Marcu, 2006) textual entailment (Mirkin et al. 2010) or text summarization (Poesio et al. 2016). Additionally, opsince bridging relations are a specialized form of anaphora references, being able to extract them effectively can lead to improvements in several semantic and pragmatic representations of documents and words (Poesio et al. 2016).

1.6 Motivation

Here we propose a method based on lexical information as an attempt to shed light over the bridging anaphora problem, though other researches that have included this kind of information in previous anaphora resolution tasks, to the best of our knowledge this is the first research that focuses on solving bridging anaphora by the use sense embeddings.

The idea for this thesis rise due to two facts: the first is that many bridging anaphoras are expressed via lexico-semantic relations such as whole-part relations and encyclopedic, so the use of lexical resources seems to be an intuitive decision for those examples, in fact, previous works suggest that bridging relations are indeed a manifestation of lexical relations (Irmer, 2011). The second reason lies in the free availability of big lexical resources for research purposes such as WordNet (Miller, 1995) and BabelNet (Navigli and Ponzetto, 2010), which allow us to try these ideas for bridging anaphora resolution.

1.7 Contributions

The main contribution of this thesis are:

1. A new model for bridging anaphora resolution from a lexical perspective, which is able to improve detection of bridging antecedent in some particular cases.
2. Introducing the use of sense embeddings for bridging anaphora resolution.
3. Introducing word sense disambiguation and entity linking together as part of a pipeline for bridging anaphora resolution.
4. A new classification for bridging anaphora from practical perspective for computational models.
5. Additional annotation of the bridging anaphora in the ISNotes dataset.

1.8 Structure of this thesis

The Second Section is the theoretical framework, here a body of knowledge is provided as background, we present different concepts and perspectives which are relevant for understanding the content of this work, such as the lexical perspective bridging anaphora can take, as well as the most used existent bridging anaphora datasets, and needed concepts from machine learning. The Third Section presents a synthesis of related works on bridging anaphora resolution recently developed, considering their point of view for solving this task, the first refers to rule-based systems and hand-crafted feature engineering for the use of machine learning methods, the second is present recent approaches using word embeddings specialized for this task. The Fourth Section describes the additional annotations we did over ISNotes corpus and presents the algorithm we used during this research to introduce sense embeddings as additional information for bridging anaphora resolution. We employed the specialized word embeddings previously mentioned and compare them with embeddings which include lexical information. The Fifth Section show the results obtained for the corpus annotation and reveal the analysis we did for the experiments using our algorithm to find the bridging antecedent. Finally, in the Sixth Section the conclusion for this thesis is given, here we briefly discuss the pros and cons of our approach and expose the main contribution and lesson learned from this work, then some brief statements about future work are mentioned.

2 Theoretical framework

2.1 Reference resolution

Reference resolution is the task of detecting what textual expression another one is referring to. Reference resolution have demonstrated to be one of the most challenging tasks throughout NLP history (Sukthanker et al. 2018). Reference resolution can be understood according to two different perspectives, the first way is to classify reference resolution regarding the position of the referent, thus, it is possible to distinguish between anaphora and cataphora, the former indicates to a previous mentioned word or phrase, and the latter to something that will be mentioned later on the discourse. The second perspective discriminates the references regarding whether they are direct or indirect, by such means it will be named coreference resolution if the goal is to find direct references e.g. the relation of identity expressed between a pronoun and a previously mentioned entity, conversely it will be named bridging anaphora if the objective is to find indirect references, e.g. two concepts holding a part-whole relation. For the purpose of this work we will discuss only the second classification. The following are some examples of direct and indirect anaphora.

(2.1) "I vote for Nader because he was more aligned with my values," she said. [2]

(2.2) Starbucks has a new take on the unicorn frappuccino. **One employee** accidentally leaked a picture of the secret new drink.

The example (2.1) is for coreference, it exposes the coreference groups (also called set and clusters) using colors (red and blue), that is to say, all the elements of a group refers to the same concept or entity. On the other hand, bridging anaphora links mentions using associative or implicit relations, such as worker-organization as example (2.2) shows. These examples evinced the second major difference between direct and indirect anaphora, the former conform clusters while the latter creates reference chains. Another important difference for them comes from psycholinguistic studies disclosed in Singer (1979), in this work the author shows that coreference is understood faster for a reader or hearer than bridging references, may be this difficulty is one of the causes for the slow progress in bridging anaphora concerning both, performance and reliable datasets, in comparison with its counterpart, the coreference task.

[2] Example taken from https://nlp.stanford.edu/projects/coref.shtml

Because of the difficulty of solving bridging anaphora, several complexity scopes have been designed, such as, reckoning only definite nouns, therefore ignoring most of the indefinite noun phrases (Irmer, 2011), other works have been extended to consider nouns and verbs, and just recently some works consider bridging anaphora in unrestricted setting (Hou et al. 2018a), i.e. the consideration of any kind of bridging anaphoras.

The following subSections expose in deeper detail the nature of bridging anaphora resolution, specifically the classification of bridging anaphoras and its relation with lexical information.

2.1.1 Bridging anaphora classification

Bridging anaphora resolution is a difficult NLP task which occurs with a wide range of scenarios. This has motivated several research to find accurate classifications for the bridging phenomenon. In this concern, Clark (1975) is one of the earliest and most important works for bridging classification, in fact was this author who coined the term “bridging”. The brief taxonomy presented in Clark (1975) consider 11 classes and use approximately 3 samples per class to exemplify them. Table 1 offers an excerpt of this taxonomy.

Table 1. Excerpt taken from Clark (1975) taxonomy

Class	**example**
identity	I met a man yesterday. The man told me a story
Set membership	I met two people yesterday. The **woman** told me a story.
Necessary parts	I looked into the room. **The ceiling** was very high.
Probable parts	I went shopping yesterday. **The walk** did me good.
Inducible parts	I walked into the room. **The chandeliers** sparkled brightly.
Necessary roles	I went shopping yesterday. **The time I started** was 3 p.m.
Optional roles	John was murdered yesterday. **The knife** lay nearby.

Reasons	John fell, **what he wanted to do** was scare Mary.
Causes	John came to <u>the party</u>. The one who **invited him** was Mary.
Consequences	John came to <u>the party</u> early. The one **he saw first** was Mary.
Concurrences	Alex went to <u>a party last night</u>. He's going **to get drunk again** tonight.

Examples exposed in Table 1 are using the notation of “bold type” for anaphoras and “underlined” for antecedent, except for examples of class “identity” and “reason”[3]. Although Clark (1975) was highly influential for succeeding works, nowadays its taxonomy puts some inconsistencies on display, starting from its first class “identity” which contradicts the definition of bridging anaphora since it expresses direct reference i.e., coreference. Moreover, some classes in this taxonomy can be difficult to distinguish as Clark (1975) acknowledge, it is hard to separate the classes "parts" from "roles”. In fact, a posterior research (Irmer, 2011), argues that the last four classes presented in Clark (1975), namely reasons, causes, consequences and concurrences are not bridging classes, but coherence relations, a different kind of discourse relations. Further bridging classification proposals which have been addressed these issues can be found in Irmer (2011). Additionally, the present work proposes a new classification in Section 4.1.

2.1.2 Bridging relation from a lexical-semantic perspective

Lexical-semantic is a subfield of linguistics semantics which focuses on the study of the composition, classification of lexical units, as well as the relationship between them. A lexical unit may refer to a single word, part of a word or sequence of words (MWE) that conform the vocabulary of a language. Lexical unit relationships include hyponymy (also called is-a relation), meronomy (also called part-whole relation), synonymy (lexical units with a close meaning), antonymy (lexical units with opposite meaning), among others. From a cognitive view, these relations are learned in context, such learning can be denominated world knowledge (Hovy et al. 2013).

[3] According to Clark (1975) the antecedent for the example shown for the class of ”reason” is not present in the text but implicitly linked to the verb fell, he suggests this explanation “John fell for the reason that he wanted to do something; that something is the antecedent to what he wanted to do.”

The relevance of lexical-semantic studies for bridging anaphora is the notion that determine the antecedent for a given bridging anaphora could imply reasoning operations and world knowledge to some extent. As Irmer (2011) observed, possibly bridging relations are readily available in the presence of strong semantic relationship between bridging antecedent and anaphor. In fact, a tuple antecedent and bridging anaphora can be seen as a lexical unit sequence split for rhetorical purposes. Lexical unit sequences are usually collocations like "food inflation" or a prepositional phrase like "penalties for pollution", hence, they can become a bridging anaphora for a given discourse, as it is observed in the following examples:

> (2.3) On Aug. 1, the state tore up its controls, and <u>food</u> prices leaped. Without buffer stocks, **inflation** exploded.

> (2.4) Conference participants saw these effects as flowing directly from planned economies' inability to control <u>pollution</u> where enterprises are state-owned and **penalties** are paid by the government.

These new constructions show that the first element become either the antecedent for the case of collocation, or a bridging anaphor, for the case of a prepositional phrase, the knowledge of that split is possible is responsible for preserving the coherence along the discourse. This idea was also grasped in Poesio et al. (2004) for relations between nominal phrases using the preposition “of”, and then generalized in Hou et al. (2018a) to “preposition patterns” which consider all the prepositions. The main hypothesis for preposition patterns is that they are not only able to capture the most common lexical relations (like meronomy) but naturally encompass encyclopedic relations such as caused by, located in, part-of, made of, attribute of, etc. As a result of this, some scholars (Nand and Yeap, 2012) claim this relations are in fact, an explicit interpretation for bridging anaphora.

In spite of these arguments, another research Rösiger (2018b) claims not all bridging are based on lexical relations, but may be possible to consider at least two kinds of bridging anaphoras, namely those which can be explained by lexical relations and those which interpretation are highly depending on the context. In this work we consider this observation to craft a classification for bridging anaphora, which consider “context” cases.

2.1.3 Technical challenges for bridging resolution

When solving bridging anaphora, and particularly, bridging antecedent selection, there are a set of issues to be faced at practical time. The following list mentions the most important practical issues identified during this research:

- **Overlapping candidates:** it occurs in most of the textual constructions that phrases overlaps with others, e.g. “food prices” and “food”, this overlap can make difficult to model each candidate having different enough representations. For instance, given a particular text a solver decides to assign “food prices” as the bridging antecedent of “inflation”, yet “food” and “prices” are also good candidates. It could be hard for a computational model to discriminate between those three candidates since the representation is likely similar. Moreover, because of the compositionality of languages, the overlapping increase severely the number of candidates along phrases which comprise others.

- **Coreferenced candidates:** in a similar way, it is hard to procure different representations for coreferenced candidates. Although, each context provides some semantic, pragmatic or discourse argues to favor one candidate over others, it is still difficult to model such rationale. Even more, such coreference can occur without using the same form, as the case for acronyms, for example:
  (2.5) <u>OPEC 's</u> ability to produce more petroleum than it can sell is beginning to cast a shadow over world oil markets. Output from the *Organization of Petroleum Exporting Countries* is already at a high for the year and most member **nations** are running flat out.
  In (2.5) the bridging anaphora “Nations” refers to “OPEC’s” (according to the corpus), yet a solver can wrongly predict Organization of Petroleum Exporting Countries as antecedent because this is the full name for OPEC.

- **Windows context**: this is a common strategy many systems use to alleviate the explosion of possible candidates for a given bridging anaphora. This strategy takes just some few previous sentences from a given bridging anaphora as a windows to search for the bridging antecedent. Particularly, one of the most used datasets for bridging anaphora resolution, called ISNotes, shows that nearly 75% of the bridging antecedent are inside the windows of two sentences before the respective bridging anaphora. This reduces the difficulty of the task at the price of sacrifice an important quantity of samples. The following chart shows how many sentences far are the antecedents from their respective bridging anaphora:

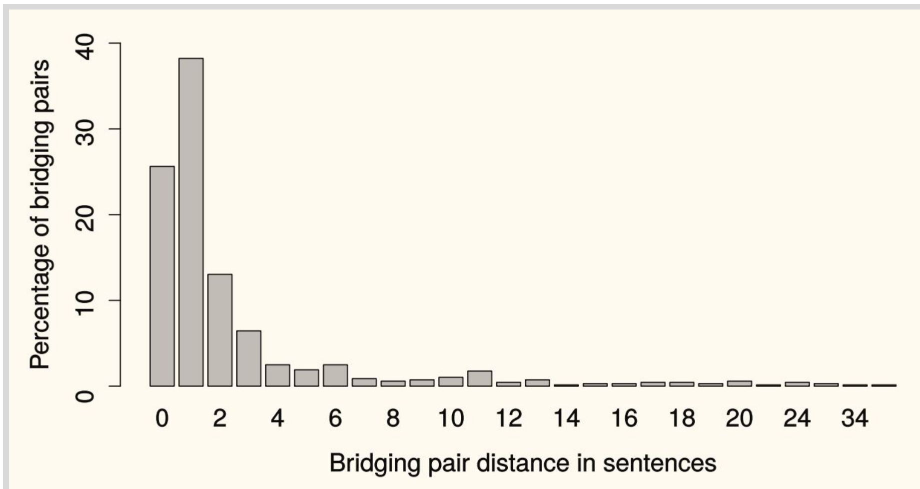

Figure 2. Anaphor–antecedent distances in sentences. (Hou et al. 2018a)

- **Knowledge bottleneck:** to consider world knowledge can benefit bridging anaphora. According to Guha (2017) bridging anaphora needs a significantly greater amount of world knowledge to solve anaphoras compared with coreference resolution, since usually the knowledge needed to discriminate between candidates it is not present in text. Lexical information such as lexical-semantic relations like synonymy, hyponymy or meronomy, encompass an important part of the required knowledge, however the knowledge stored in current knowledge repositories not seems to be enough for demanding tasks like bridging anaphora (Markert and Nissim, 2005). As a matter of fact, solving this gap of knowledge is a major problem known in natural language processing as “knowledge bottleneck” (Hovy et al. 2013).

- **Contextual cases:** there are cases where determine the antecedent is possible just by having a good understanding of the context, i.e. how the words relates semantically between them. Consider the following example:

(2.6) The state quit shoving peasants onto its subsidized farms over 30 years ago. But it never did let up on **the pressure**.

Here, “the state” performs certain action over “peasants”, then in the following sentence, “the state” (which is co-referred using the pronoun “it”) performs a second action over an omitted object, which implicitly correspond to “peasants”. Therefore, other linguistic phenomena such as ellipsis and coreference, should be considered for solving bridging anaphora.

- **Long mentions:** another particular issue of bridging anaphora is resolutions is the existence of long mentions in the role of bridging or candidate. The meaning of long mentions are challenging to represent by computational means, since most aggregation methods are not good at capturing the global meaning. Consider the next example, which is a very difficult one, not only because of the length of the mention but because it suffers from other problems discussed above.

  (2.7) The White House has likewise avoided any involvement in Florida's recent special legislative session on abortion, which anti-abortion forces had regarded as a key test of their ability to get state lawmakers to toughen abortion restrictions. The session failed to enact any new curbs. Now, some see Mr. Bush trapped in **a position he is neither comfortable with nor able to escape**.

2.1.4 Bridging anaphora datasets

Building bridging anaphora datasets is a challenging job, as it has been evinced in natural language processing history, most of the datasets face the dichotomy between suffer from lack of reliability versus being too small as a condition to preserve the inter-annotator agreement (Poesio, 2004). In Guha (2017) the author claims that the difficulty of bridging anaphora could be a reason to restrict the annotation to certain POS tags or biased to them, i.e., nominal phrases. As a result of collecting information, some of the main bridging datasets are cited below.

- **ISNotes corpus:** released by Hou et al. (2013b). This corpus contains near 11K nominal phrases annotation for information status, and 663 bridging samples obtained from 50 documents of newspaper domain. Specifically these 50 documents were taken from OntoNotes 4.0 (Weischedel et al. 2011), hence ISNotes can be seen as an additional annotation layer for those documents on OntoNotes. The main particularity about ISnotes respect to any other bridging dataset is the “unstrictedness” of its bridging samples, which in words of Hou et al. (2018a) means that unlike previous released datasets, they did not impose any constraint on the type of bridging anaphora or relations between anaphor and antecedent.
- **SciCorp**: developed by Rösiger (2016), this is a corpus annotated following information status schema over scientific text about computational linguistics and genetics. It contains 1366 bridging samples. However, it contains several samples which from our point of view are more related to coreference resolution than bridging anaphora, for example: “them . . . their interest.” The dataset can be obtained from official webpage[4]
- **ARRAU**: annotated by Poesio et al. (2013) from 3 domains: newspapers, narrative text and dialogue, but the newspaper domain constitutes the major part of the samples. This corpus contains 5512 bridging samples, however, most of these samples refers to only lexical semantic relations, hence, some authors like Rösiger et al. (2018a) suggest they are not truly anaphoric bridging anaphors, or at least the corpus imposes a big restriction. The data can be obtained from the LDC[5].

2.2 Lexical information

In computational linguistic literature, lexical information refers to all the features or representations which can be obtained from words regarding their lexical properties. However, not all words have a rich lexical information. From a lexical perspective words are commonly discriminated between open- and closed-class words regarding its part-of-speech (Sun and Uszkoreit, 2012). Closed class words are generally function words for structuring grammar, open class instead are associated to a lexical meaning or sense from which several lexical properties are derived. Just nouns, verbs, adjectives and adverbs are widely recognized as words that exhibit senses. Nonetheless, to obtain the sense of an open class word is not

[4] https://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/scicorp.html
[5] https://catalog.ldc.upenn.edu/LDC2013T22

straightforward process, since the meaning can vary enormously according to the context in which they (occur Jurafsky and Martin, 2009), this phenomenon is called polysemy, and the process to retrieve the sense given a word in a context is called Word Sense Disambiguation (WSD).

Lexical information is usually stored into lexicons or lexical databases to make them machine readable. Nowadays, lexical repositories such as Babelnet (Navigli and Ponzetto, 2010) are designed to store millions of senses enriched by lexical information using a graph structure. In such a manner senses are usually represented as nodes, while lexical relations for a pair of senses are usually represented by an edge connecting the respective nodes. Notably, BabelNet is the largest and most complete manually-made lexical repository (Camacho-Collados, 2017), currently it contains near 809M of senses and 277M of lexico-semantic relations.

Another kind of lexical information that have gained importance recently is called sense embeddings (Camacho-Collados and Pilehvar, 2018), which is a specialization of word embeddings for senses. In short, these embeddings are distributed representations in a vector space model. The main motivation for the existence of sense embedding is to have a vector able to represent words according to their meaning in a given context, that is to say, independently whether they use a synonymy or a different form, unlike traditional word embeddings which always provide the same vector for the same words. For this work we are using Nasari a sense embedding resource which leverage BabelNet as repository. Notwithstanding, the use of sense embeddings resources usually requires to perform Word Sense Disambiguation (WSD) and Entity Linking (EL) tasks, since words are not supposed to be already annotated with sense information. The shared goal for these tasks is to determine the sense of a word, or a multi-word expression (MWE) according to a given lexical repository. On the other hand, the difference between them lies on the type of sense they attempt to disambiguate. It is called EL if the system aims to find named entities which are real-world object, such as persons, locations, organizations, products, etc., or it is called WSD if the aim is to find concepts. According to Navigli (2009), a disambiguation task can be considered AI-complete task. i.e. a very challenging artificial intelligence.

We noted in this work that for the case of bridging anaphora, a good WSD process it is necessary for an effective use of lexical information. This fact was also evinced in Markert and

Nissim (2005), in that work the proposed method wanted to avoid consider all the possible senses of a word and multiword expressions. In this work we disambiguate all the documents before anaphora resolution using Babelfy (Moro et al. 2014), an algorithm that perform WSD and EL tasks at the same time in unsupervised fashion.

2.3 Machine learning in NLP

Machine learning is a subfield of artificial intelligence focused on learning programs automatically from sample data, in place of hard-coding all the instructions the machine should follow. Machine learning is used to model a wide range of phenomena and perform predictions over unseen samples, therefore, it has a vast ambit of applications, and in fact, it has provided solutions for most science fields. The main purpose for machine learning in NLP is to perform automatic classification of linguistic examples, to this end each object should be represented as a vector, where the dimensions of the vector corresponds to a set of features, which serve to describe the linguistic object. Such dimensions are known as vector space model, because it models the objects by place them into a vectorial space (Sidorov et al. 2014).

The most common types for machine learning are supervised and unsupervised models. The former relies on training data to learn how to predict over test data, i.e, the algorithm receives a dataset where each sample is annotated with the expected output, and given a vector space model, it learns how the representations are correlated with the corresponding class. The latter unlike supervised models, do not use annotated data, these types of models are designed to discover the classes which each sample belong only using the vector space model. Unsupervised algorithms mostly classify by clustering the sample into groups and then match such groups into the required classes; another alternative is simply predicting the outcome class as the most similar between the possible options, relying on similarity measures over the classes or other samples. As a matter of fact, in any machine learning model, is expected that a better representation uses to lead to better results.

In practice for NLP, machine learning use to leverage different linguistic levels to build the vector space model for the object and the task of interest, the process of manually design the representation is called hand-crafted feature engineering. Although hand-crafted feature engineering approaches are still utilized these days for some tasks, in general they have been substituted by more robust and cleverer algorithms, hence they are also known as traditional (or classical) machine learning. The recent approaches for NLP rely on the distributed representation of words called *word embeddings.* Word embeddings models the distributional information of words as dense vector representation (Mikolov et al. 2013), since it was introduced a decade ago, it has been used as vector space model with impressive and indiscutible benefits for NLP tasks. Moreover in comparison with hand-crafted feature engineering, word embedding as features used to perform better when enough training data is provided, and unlike traditional machine learning features, word embedding are automatically build. However, one advantage of hand-crafted feature engineering respect recent models, is that the former provides better explanations about the nature of the problem, based on the importance of the features it employes to solve the task of interest.

3 Related work

Different ways for building bridging anaphora solvers have been proposed in the literature, the main difference between approaches lay on how the external information is exploded. The following subSections presents a discrimination bewteen classical approaches corresponding to rule based systems or hand crafted feature engineering, and more recent approaches based on distributed models especialized for bridging anaphora.

3.1 Rule-based systems and hand-crafted feature engineering

Rule-based systems were the pioneer strategy in most of the proposed NLP problems, bridging anaphora resolution was not the exception, the early works attempts to leverage the existing theories and machine readable knowledge resources to design rules. However, such rules prove to be insufficient to model bridging phenomenon, thus, with the introduction of machine learning new approaches for NLP tasks new approaches arise leveraging hand-crafted features for bridging anaphora. Hand-crafted performed better than rule-based system, specially for bridging anaphora a remarkable proposal is the use Markov Logic Networks (MLN) (Domingos and Lowd, 2009) which is currently used in combination with recent word embedding advances to become the state-of-the-art (Hou, 2018c).

3.1.1 Using lexical information and web sources

Early bridging anaphora models whe re designed using rules derived from the understanding of the bridging phenomenon. The common strategy was to acquire and exploit external information and several sources were proposed in the literature to take advantage of them. Most of these works leverage information from lexical repositories or statistical analysis as a way to fill the knowledge gap. For instance Poesio et al. (2004), Markert and Nissim (2005), Sasano and Kurohashi (2009), and Lassalle and Denis (2011) which alleviate the sparsity in lexical repositories such as WordNet (Miller, 1995) using the web as a source. These works show the benefits of using distributional information for bridging anaphora resolution. Particularly, in Markert and Nissim (2005) the authors implement Hearst Patterns (Hearst, 1992) to perform lexical knowledge extraction from 8,058M webpages, which was the largest corpus available to the NLP community. We must point out that strategies using WordNet or any other sense-based lexical repository have to perform WSD in order to obtain a context-dependent relation between a pair of lexical units (see Section 2.2 Lexical information).

3.1.2 Markov Logic Networks

Years later Hou et al. (2013b) presents a novel approach for modelling bridging at the global level, considering the full document. The approach was based on a generative model called Markov Logic Networks (MLN) (Domingos and Lowd, 2009) which is able to combine first order logic with Probabilistic Graphical Models (PGM). The MLN model presented in Hou et al. (2013b) associates First Order Logic propositions with weights learnt from training data, these weights are intended to measure the importance for such propositions and conforming the set of features used for a machine learning algorithm. Additionally, this model takes into consideration the proximity between bridging antecedents in order to exploit the pattern of "sibling anaphors" (anaphors that share an antecedent with other bridging anaphors) and because of the graph approach, it is able to classify better long distance antecedents than previous pairwise proposals.

Another work that implemented hand crafted engineering is Hou et al. (2014), here the authors considered eight rules, four of them were designed to capture statistics patterns from big corpus using two notions, the first is called semantic connectivity, it accounts how frequently a pair words co-occurs in a prepositional phrase. The second is called argument-taking ratio, which accounts how likely a nominal phrase is to take arguments. Other two rules were designed to verify whether the bridging reference appears in a list of potential antecedent-anaphor pairs e.g., *prime minister - Japan*. Finally, the other two are intended for face bridging anaphora set cases[6], such as examples (3.1) and (3.2)

> (3.1) 22% of <u>the firms</u> said employees or owners had been robbed on their way to or from work. **Seventeen percent** reported their customers being robbed.
>
> (3.2) <u>Reds and yellows</u> went about their business with a kind of measured grimness. **Some** frantically dumped belongings into pillowcases.

[6] Examples, (3.1) and (3.2) follow the set - element pattern

The model designed in Hou et al. (2014), was later replicated in Rösiger et al. (2018a), with the introduction of a new feature for the purpose of improving results for recall metric, this feature discriminates hyponym from meronym relations between the elements for each bridging pair, the feature was computed by an off-the-shelf classifier which received ConceptNet embeddings (Speer et al. 2017) as input. Results in Rösiger et al. (2018a) show that the system generalises to in-domain corpora if they are of the same type of bridging.

3.2 Specialized word embeddings for bridging anaphora resolution

In the last decade word embedding have become an essential part of the pipeline for almost any NLP model (Mikolov et al. 2013). Sometimes specific word embedding models are designed for solving specific NLP tasks, the common procedure is to bias the model towards specific aspects of the language that are relevant for the task of interest. These embedding are commonly named specialized embeddings. Generally, specialized embeddings take the relevant aspects from knowledge generated by previous works, where it is explained theoretically or empirically how phenomenon works. This is the case for bridging anaphora resolution, the recent research Hou (2018b) and Hou (2018c) have developed word embedding resources specialized in capturing bridging aspects.

In Hou (2018b) it is explained how to build the proposed specialized dense representation over row corpora by matching the morphosyntactic pattern X <prep> Y, where X and Y correspond to nominal or verbal phrases. As it can be noticed, the pattern resemble those used in early works (Poesio et al. 2004), reinforcing the role that prepositions play bridging anaphora resolution. Moreover, Hou (2018b) makes an important claim about the importance of directionality feature that can be captured following this pattern, unlike simple statistical aggregations. To this end, the author add a suffix "_PP" to the first nominal phrase as a mark that indicates the directionality of the nominal phrase in the prepositional pattern. The introduction of this simple feature improves the results by a large marge, then the embeddings were named "embeddings_PP". Not long afterward, Hou (2018c) released "bridging embedding" as an extension of "embeddings_PP" which face the problem of words out-of-vocabulary by including GloVe embeddings (Pennington et al. 2014).

Both works, Hou (2018b) and Hou (2018c) use their respective specialized embedding models for perform unsupervised bridging anaphora resolution. The procedure for testing consists of measuring the cosine similarity for each bridging pair and choose the candidate with the highest similarity for each anaphor. They also applied some hard rules for filtering some candidates, such as ‘to include as candidates only those representing a DATE entity if and only if the anaphor also represents a DATE entity’. Also, in the cases when a mention correspond to more than 1 word they average the embeddings for each of them. Consider the following example:

> (3.3) The space shuttle Atlantis landed at a desert air strip at Edwards Air Force Base , Calif. , ending a five - day mission that dispatched the Jupiter - bound Galileo space probe . **The five astronauts** returned to Earth about three hours early because high winds had been predicted at the landing site.

The first step is to list the possible candidates and compute their vector representation, as well as for the bridging anaphora. Then, a cosine similarity is computed, and the most similar is classified as the antecedent for the aforementioned anaphor. Below there is a graphical description which the procedure.

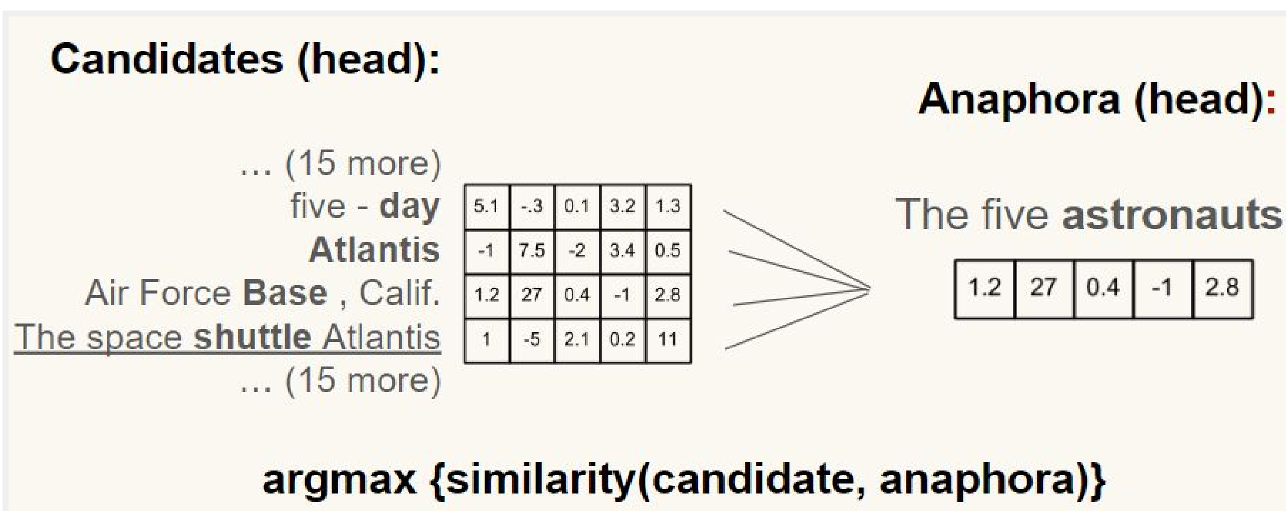

Figure 3. Graphical description of a pairwise setting for bridging anaphora resolution

The results reported in Hou (2018b) show the benefits of using specialized embedding for the task of bridging anaphora. This strategy can be considered the state of the art, although a higher results can be obtained by fusion “bridging embeddings” with the work developed in Hou et al. (2013b), as is shown in Hou (2018c), they report reporting an accuracy of 46.46%. This result indicates that there is still the need of research in bridging anaphora.

4. Methodology

This Section describes the research procedures followed to design the bridging resolution model using lexical information. This Section is a compound of two parts: The first part is about the manual annotation of bridging cases, an informal bridging classification proposal we employed to inform us about the performance of the different approaches we tried to compare with a baseline which considers the results reported in state-of-the-art. The second part is the algorithm for solving bridging anaphoras which offers a clear view about the preprocess operations conducted over the original data, throughout the procedures for generating vector representations and getting predictions for unseen examples.

This research considers the resolution of unrestricted bridging anaphora in unsupervised fashion. Hence, the included bridging anaphora are not limited to particular parts of speech or certain forms of them, such as nouns which are introduced by definite pronouns, nor training process was used for getting predictions.

4.1 Corpus annotation

Considering the several classifications proposed for bridging phenomenon from a theoretical point of view Clark (1975), Irmer (2011), and motivated for the recent research Rösiger et al. (2018a), we propose a classification scheme which follows previous ideas, i.e. discriminate between “lexical cases” and “context cases” but extending to three classes, for the inclusion of distributional class. This classification is intended to guide the classification models to determine which set of features are more likely to benefit the prediction for a given case. In summary here we considered three categories, namely knowledge, distributional and contextual. Briefly, lexical cases indicates lexical-semantic or encyclopedic relations, distributional cases indicates relations which are not evident but occurs frequently across large corpora and contextual indicates the relation is particular of the introduced context and much less frequent across large corpora. Table 2 shows some examples for this classification:

Table 2. Examples of bridging cases annotation

Example	Bridging Case
Currently , <u>Boeing</u> has a backlog of about $80 billion , but production has been slowed by a strike of **55,000 machinists** , which entered its 22nd day today.	Knowledge
<u>The business</u> closed when **the owner** was murdered by robbers.	Knowledge
Conference participants saw these effects as flowing directly from planned economies' inability to control <u>pollution</u> where enterprises are state-owned and **penalties** are paid by the government.	Distributional
Mario Mandina, president of Kansas City Lawyers for Life, says that if abortion foes succeed in using the preamble to escape prosecution for trespass, “This will shut down <u>abortion</u> in Missouri. There 's no risk to **the protesters**, and you ca n't keep an abortion clinic open if there are 3,000 people standing outside every day.”	Distributional
Nobody is sure what will come next in <u>Somalia</u> or whom the successor might be. But as **one expert** tells me: "Whoever it is will have to work pretty damn hard to be worse than Barre."	Contextual
But industry and OPEC officials agree that a handful of members <u>still have</u> enough unused capacity to glut the market and cause an oil-price collapse a few months from now if OPEC doesn't soon adopt a new quota system to corral its chronic cheaters. **As a result**, the effort by some oil ministers to get OPEC to approve a new permanent production-sharing agreement next month is taking on increasing urgency.	Contextual

From this, we can consider knowledge cases exhibit close lexical relations, distributional exhibit less evident relation, but they can be understood by seeing the bridging and its antecedent as a collocation, and lastly contextual cases exhibit very implicit relations which manifest clearer in the particular bridging anaphora sample, they usually correspond to ellipsis escenarios. We can also claim contextual cases are the most difficult cases to solve. Also, it occurs that distributional cases exhibit characteristics of lexical and contextual cases, since some kind of encyclopedic relation or ellipsis phenomenon can be set for distributional collocations.

This annotation allowed us to gain a deeper understanding of bridging phenomenon, particularly in the domain of news (which is present in ISNotes) and comparison between bridging resolution model approaches, to determine their strengths and weaknesses. To this end, we tagged ISNotes corpus manually with this categories (see Section 5.1), and automatic algorithm to perform this prediction is proposed as future work.

4.2 Algorithm

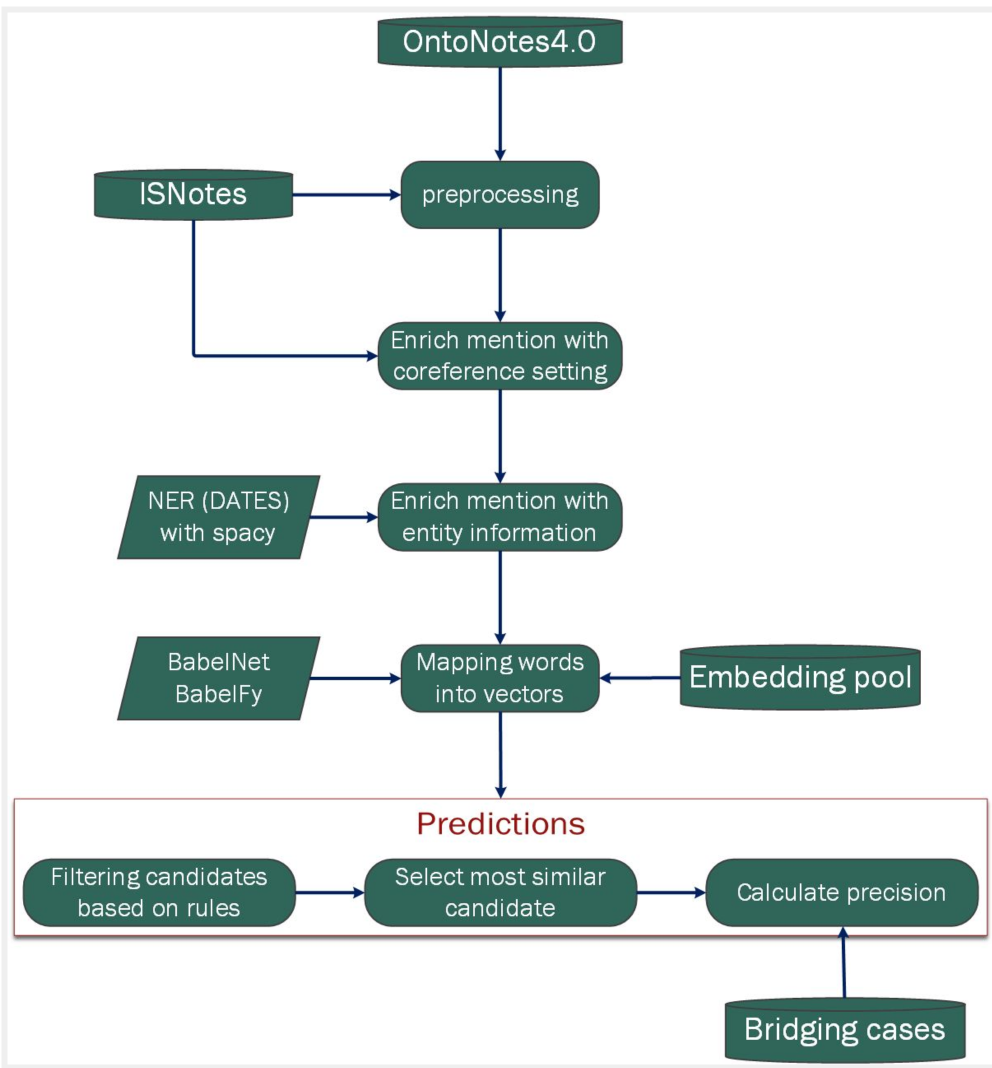

Figure 4. Model architecture outline

This algorithm is end-to-end, that is to say the process starts from receiving a bridging anaphora corpus and ends with a final output which contains the predictions for each bridging anaphora. This process is summarized Into four stages: 1) corpus processing, 2) bridging representation, 3) vector computation and 4) model predictions. A major detail for these stages is reported in the subsequent Sections. Figure 4 offers a graphical outline of these stages. Additionally, the employed tools and resources are included in the outline and a brief description for them is given afterward.

The following is the full list of tools used for the purpose of this research:

- ISNotes (Hou et al. 2013b): is the corpus where the unrestricted bridging annotations come from.
- OntoNotes 4.0 (Weischedel et al. 2011): this is the corpus employed as base for the bridging annotations offered in ISNotes.
- OntoNotes2MMAX.jar (Hou et al. 2013b): this script is used to splice together OntoNotes 4.0 and ISNotes annotations.
- KNIME Analytics Platform 4.0 (Berthold et al. 2007): this is a data science platform, which allows the use of graphical data workflows as an intuitive way to transform data and build machine learning models. This tool is used here to shape the original corpus input into bridging pairs, compute the vector representations and afterwards perform the predictions.
- Spacy (Honnibal and Johnson, 2015): is a free open-source library for Natural Language Processing in Python. We used it for lemmatization.
- Babelfy (Moro et al. 2014): this is an algorithm able to perform Word Sense Disambiguation (WSD) and Entity Linking (EL) over any input text in a multilingual fashion. It links concepts and entity to a common lexical repository called BabelNet which was employed in this research too.
- BabelNet (Navigli and Ponzetto, 2010): it is a multilingual lexicalized semantic network and ontology, which serves as a repository for Babelfy disambiguation.
- Nasari (Camacho-Collados et al. 2015): this is the lexical resource which provide the sense embeddings used here for getting bridging anaphora predictions.

- Other word embeddings resources such as “embeddings_PP” (Hou, 2018b), “bridging_embeddings” (Hou, 2018c), (GloVe Pennington et al. 2014) and Numberbatch (Speer et al. 2017).

4.2.1 Preprocessing ISNotes corpus

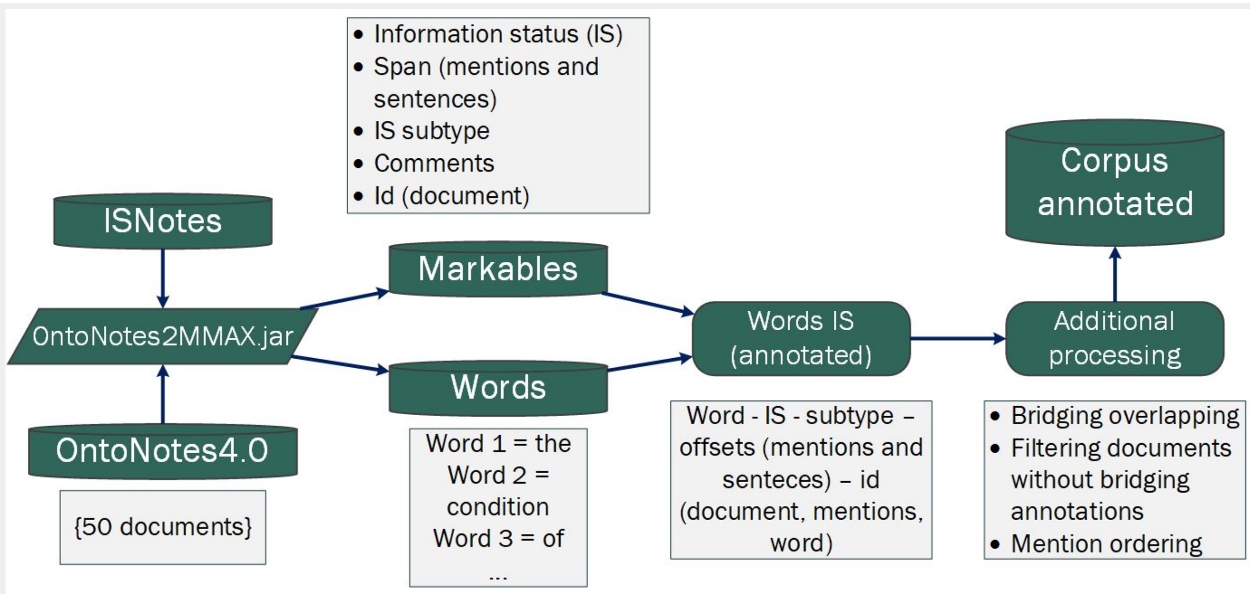

Figure 5. Model architecture: corpus processing

Original ISNotes annotations are a list of XML files which refer to 50 documents published in OntoNotes4.0 corpus, consequently, it is necessary to have access to OntoNotes4.0 corpus in order to read the bridging annotations in their corresponding context. Then, a script[7] is used to splice together annotations (from ISNotes) and context information (from OntoNotes). As a result of this process, ISNotes annotations can be read as a list of markables, words and sentence sources are are listed apart in other XML files. Each markable represents a mention, which refers to a word or multiword expression and corresponding sentences from OntoNotes, they are annotated with type and subtype of information status.

Then, some preprocess steps are done for preparation of the bridging pairs, they include filtering documents which do not contain bridging anaphoras, marking those mentions which overlaps if they are bridging anaphoras, and set an order per mentions for each document. The

[7] https://github.com/nlpAThits/ISNotes1.0

results is a word-based corpus (i.e., each row corresponds to a word) which contains all the information from ISNotes projected over OntoNotes 4.0 documents. The Figure 5 shows a high level diagram of this procedure.

4.2.2 Bridging anaphora representation

A bridging anaphora can be represented in many different ways. Some settings consider the portion of context expressed between the candidate and the anaphor. Other settings ignore the context and assume the bridging pair directly as the representation for the full anaphora, the latter is the most used, several variations have been proposed for it. The particular representation chosen is important, since it has a huge effect in the parameters a computational model will use to predict the bridging antecedent. Here we expose different settings we examined for our experiments:

- Head: this setting consider only the head of the mention instead of consider the full mention. This head is usually defined as the first node in a dependency tree, also it can be understood as the main noun in a nominal phrase resulting after remove noun modifiers, or verb depending the case. In this work the head of each mention was obtained via Spacy.
- Mention: this could be considered the basic setting for bridging anaphora representation as a pair, this setting includes all the words in the mention. Some variations of this setting apply preprocessing functions to remove stopwords or closed words in general.
- Mention and local context: this setting extends the mention representation to consider some context around the original mention. To this end, a window size of n words should be defined. For this work it was set to 5 empirically. Additionally preprocessing steps were done for remove closed words in a similar way to mention representation.

The following image shows a graphical explanation of the representation settings explained above for representing bridging anaphora.

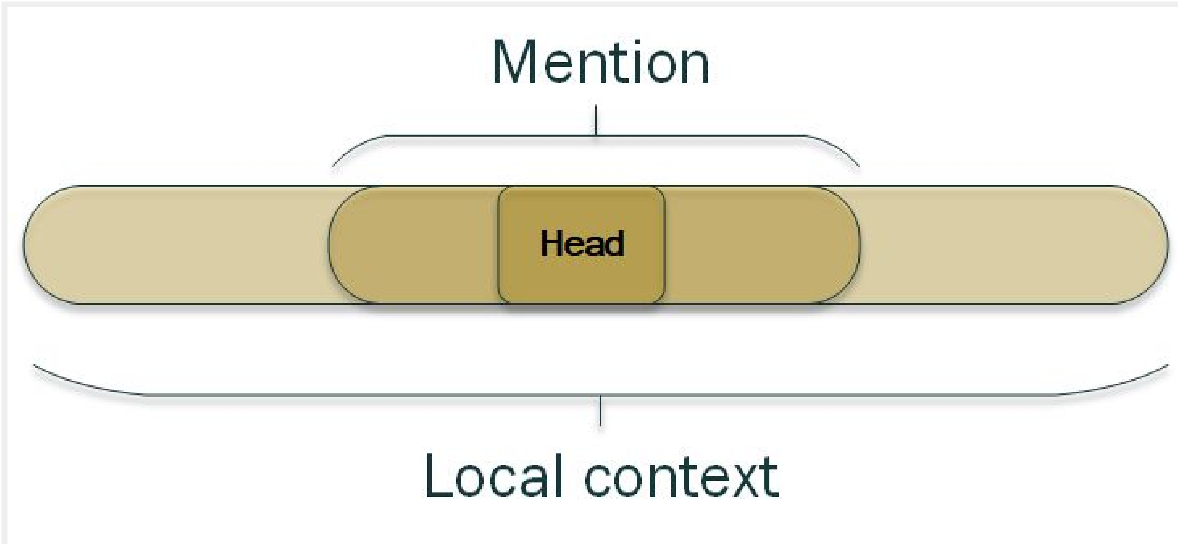

Figure 6. Graphical explanation of the representation settings of a bridging pair element

An additional processing performed during this work with the expectation of improving bridging anaphora representation is called lexical substitution. Lexical substitution is actually an NLP task where the goal is to identify a good substitute for a certain word in a given context. By default lexical substitution does not impose any restriction on which should be the best alternative for the word of interest, however a good selection of words to substitute is an important previous step to obtain improvements from applying lexical substitution. Particularly, for bridging anaphora it was noted that certain substitutions lead to better results at discriminating among candidates.

In this work two substitution strategies were tried, the former was to substitute entity names for the corresponding entity classes. The latter was about enrich mentions with their respective co-referenced versions. The first strategy was approached by the use of Spacy, this off-the-shelf tool made possible to perform Name Entity Recognition (NER), i.e., to identify entities of different classes such as ORGANIZATION, TIME, DATE and PEOPLE. In general these simple substitutions allow the model to generalize better. Particularly, entities of type DATE were useful to include in the set rules as was suggested in Hou (2018b), since it is quite common that bridging anaphora representing a DATE entity selects as antecedent a candidate which represents a DATE entity too. Other lexical substitutions were attempted, such as the relation job-organization, set-element and frame-argument. However, no one gave a better improvement for detecting antecedents than the simple rule of filter candidates with the criteria “DATE refers to DATE”.

The second strategy for substitution used coreference information obtained directly from ISNotes annotations. This annotations consist of a list of all the mentions and the coreference cluster they were associated with. We make use of the following three settings to test this strategy:

- Union of tokens: combine all the mentions which belong to the same coreference cluster and remove repeated words.
- Top n open words: combine all the mentions belonging to the same coreference cluster and select top n most frequent words as a delegate for the mention.
- The largest coreferent: select the largest mention among all the possible mentions for each cluster set.

The following diagram shows how this operations where chained to produce a corpus annotated which IS information and enriched by coreference and entity information.

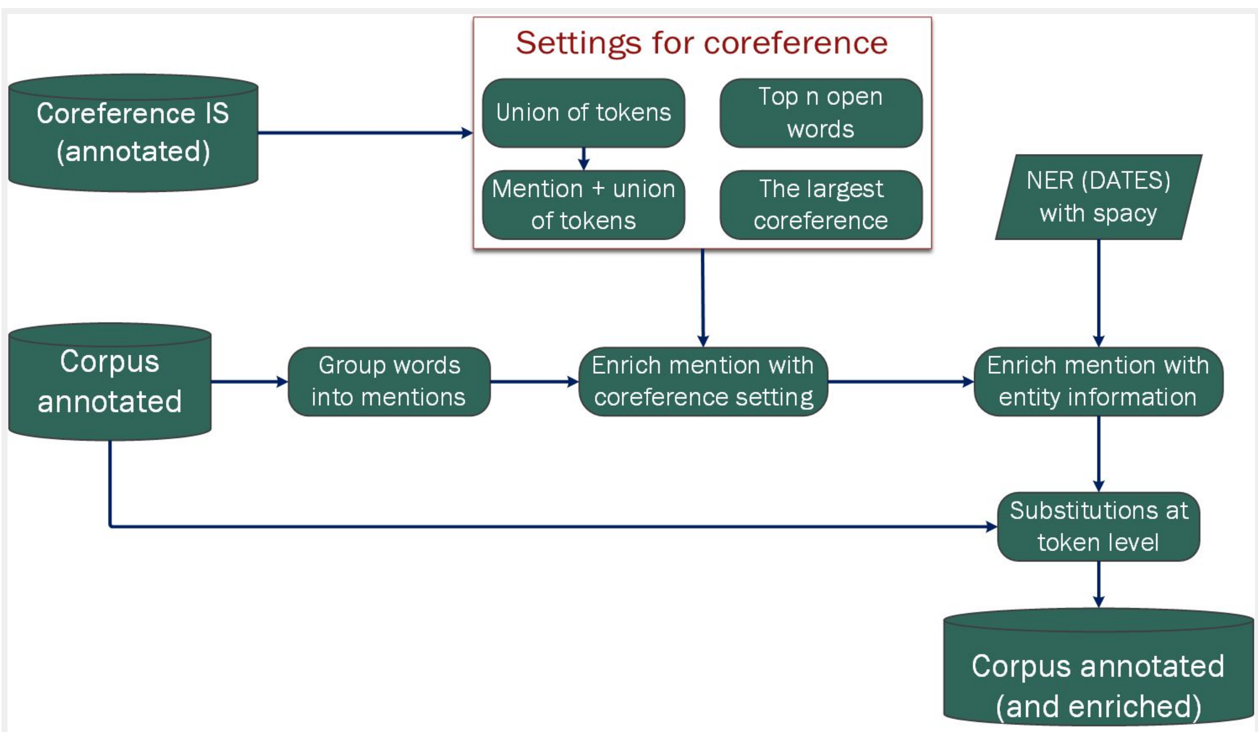

Figure 7. Model architecture: bridging representation

Additionally, experiments removing pronouns were tried for the first and third settings obtaining slightly better results. For second setting, a word was considered open if it exists in BabelNet according to babelfy disambiguation, or closed in other cases. Most of pronouns which coreference are properly annotated in ISNotes corpus. Other linguistic factor exploded, mainly for third setting came from the entity information retrieved from Spacy, since it makes sense to replace proper nouns (like organization and people names) by the text which the name entities are first introduced in order to provide additional information about the relation those entities could have with the bridging anaphora. Afterward, several aggregation methods were tested too such filter stopwords and weighted schemes for words and concepts, these aggregation methods were intended to improve the information of the actual mentioned by merging in an efficient way their coreferences. Finally, it was discovered that the best combination was substitute every possible mention with coreferences annotations by concatenating the actual mention with the unique list of words taken from the coreference forms. However, all this effort to enhance bridging pair representations offered only a slight improvement over the performance. Such results are presented in Section 5.2.

4.2.3 Automatic word sense disambiguation and entity linking

We propose the approach of use the senses of each word in order to get an accurate representation of words. To obtain senses (concepts or entities) it was necessary to disambiguate the full documents in the corpus.

Babelfy was the algorithm used to disambiguate the text. Babelfy is capable of performing both disambiguation tasks, i.e., WSD and EL at the same time. It includes several parameters to set according to the particular situation, and most important, Babelfy offers disambiguation metrics in order to estimate how trustful the result is. For the purpose of making the best with the algorithm capabilities, several settings were tried for the task of predicting bridging anaphora relations. And now we will describe in detail what aspects were considered for performing the sense disambiguation process.

The first aspect to consider is that Babelfy return different results according to the amount of context given for the disambiguation. The more context is provided, the more general the disambiguation is, therefore, the algorithm will return more general senses when a long text is input, but also less concepts and entities. On the contrary, when a small context is given, like a phrase or some sentences where the anaphor and its candidate appears, the algorithm returns more specialized concepts and a higher number of concepts. This proportional phenomenon occurs due to the graph heuristics applied by Babelfy applies during the disambiguate process of the text. The notion is that larger contexts will create a higher number of relations between the disambiguated words, and some heuristics such as the densest subgraph will attempts to select the subgraph which have less nodes but more relations, thus ignoring many senses which less connections that are out of the densest subgraph. This aspect is particularly important in the context of bridging anaphora resolutions, since as was listed in Section *4.2.2,* there are some alternatives for get a textual representation of the bridging pair, each of them will provide different concepts as the senses employed in the text. From this perspective we can try both alternatives, the first consist of input the *full context*, which encompass a windows size of words[8] before the candidate, all the context between the bridging pair, and the same windows size after the anaphora. The second consisting of the *full document* which encompasses all the words in the document. Figure 8. Shows a graphical descriptions of these alternatives. Results at Section 5.2 shows that a combination of both local context, using the first added alternative, and a global context, using the last added.

A second important aspect for considering was to set the parameter "*Enable partial matches*" which is used to set the disambiguation candidate extraction strategy. To enable partial matching allows to consider candidates which share at least one word in the lexicalized form of a concept or entity which consist of a multiword expression. On the other hand, an exact matching reduces the range of possible senses at the price of sacrifice some good disambiguation which can be performed via partial matching, i.e, all those concepts or entities whose lexicalized forms differ of the sequence exhibited in the given text. This parameter was relevant for obtaining good disambiguations for the task of finding the bridging antecedent, since it was noticed that enable partial matching also bias the WSD model towards disambiguate words as entities rather than concepts. Therefore, when partial matching is enable Babelfy tends to disambiguate words as part of entities lexicalized as multiword expressions, for instance, "Pianist" in some example is more probable to be disambiguated as "The pianist" the movie if partial matching is enabled. Nevertheless, there are also good examples when partial

[8] We experiment with several windows size between (from 5 to 15) without meaningful difference in results. Finally we resolved to use 5 words around the bridging pair.

matching can help, for instance, “Teddy Z” was successfully disambiguated as TV show “The famous Teddy Z”, and thanks to that the system did a correct antecedent selection for the particular example in which this entity appeared. Unfortunately, in most of the disambiguation cases, the bias towards entities caused a wrong disambiguation, and as a consequence a wrong antecedent selection for ISNotes documents.

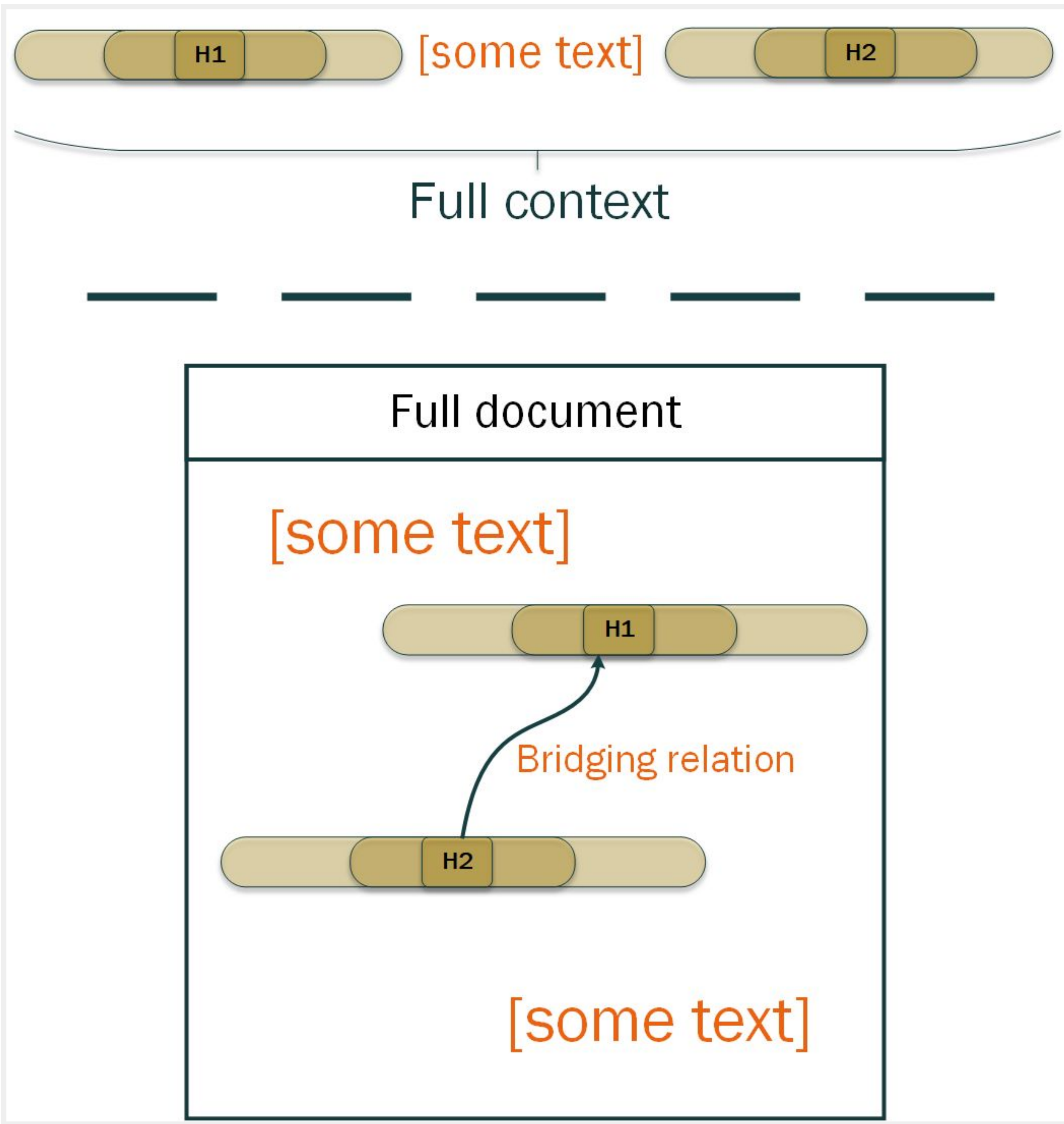

Figure 8. Full context input versus full document input for disambiguation

A third aspect which claim certain relevance was having overlapping senses as Babelfy results. Sense overlaps occurs when certain words in a disambiguated text sequence belongs to more than one sense at the same time. Next image explain graphically this phenomenon.

Word by word		Best sense found	BabelSynset
0	A	---	---
1	patent	patent classification	bn:03127990n
2	classification	patent classification	bn:03127990n
3	is	---	---
4	a	---	---
5	system	system	bn:15125301n
6	for	---	---
7	examiners	examiners	bn:03820659n
8	of	---	---
9	patent	patent offices	bn:00060985n
10	offices	patent offices	bn:00060985n

Figure 9. Aligning of words and concepts for extrac multiword expressions (MWE)

Sense overlap exists inevitably due to language ambiguity and recursivity e.g., compound meaning, but it also occurs due to the degree of concept and entity coverage in the knowledge base, particularly to establish an adequate level of sense granularity (Navigli, 2009) is desirable to avoid having many too specific or almost identical senses in the repository. A high quantity of senses leads the solver to recognize a sense for certain multiword expression and other senses for individual words which compound the multiword expression. Conversely, the lack of senses in the knowledge repository prevent the solver to disambiguate complete multiword expressions and even individual words. As an intuitive decision that works in practice, to choose the longest disambiguation leads to more precise disambiguations (Pilehvar et al. 2013), however we also noted some long multiword disambiguation although more precise from sense perspective, they obtained not that good representation than choosing individual words or less precise multiwords expressions as disambiguation. We hypothesize this occurs because of the low frequency for those high precise senses at the moment of train the embedding representation for them.

Finally, beside Babelfy parameters, two disambiguation scores were taken into consideration as a way to optimize the disambiguations results, they are GlobalScore and CoherenceScore. These scores inform about the result of the graph heuristic operations, and are part of the babelfy outcome per each disambiguation it performs. In particular, the disambiguations score were used a guide to choose better parameters for the disambiguation algorithm, they were useful to decide how much context to input concerning the bridging anaphora representation.

4.2.4 Embeddings pool and predictions

A module called “embedding pool” was designed in order to perform practical experiments using several word representation alternatives (word embeddings). It consists of a single pipeline which receive all the vector representation alternative for represent bridging pairs, therefore, different embeddings were grouped together, according to the type of preprocessing they required as it is described below.

- **word embeddings**, this group include the vanilla word embeddings we use across the experiments. It includes GloVe and Numberbatch. The nature of Numberbatch was interesting for this research, it is not a sense embeddings since it links ConceptNet as repository, which is not a sense embedding. However, it has a strong tendency to capture lexical relationships because of the graph structure of ConceptNet for storage lexical-semantic relations.
- **specialized word embeddings:** this group includes “embeddings_PP”, “bridging embeddings” which needed to add the suffix “_PP” anaphors. Also it was necessary to concatenate GloVe to “bridging_embeddings” as was suggested in Hou (2018c) for the use of the latter specialized word representation.
- **sense embeddings:** this group only includes Nasari embeddings, the processing for Nasari was quite different than the rest, because it needed to align the retrieved senses to the actual words used in the anaphora in order to get the representation of the bridging pair.

Thereupon the bridging pairs, i.e., bridging candidate and the bridging anaphora, were represented by the embedding vectors as the average of the vectors of the individual words which compound them, according to the bridging anaphora representation setting chosen. Thus, a single vector representation was obtained for each bridging anaphora and their respective candidates. In order to deal with unknown vectors e.g., because of words out-of-vocabulary, random vectors were generated accordingly to the dimension of the embeddings to replace the missing ones. The following image describes graphically how these operations we connected to provide the list of bridging pairs with their corresponding vectors.

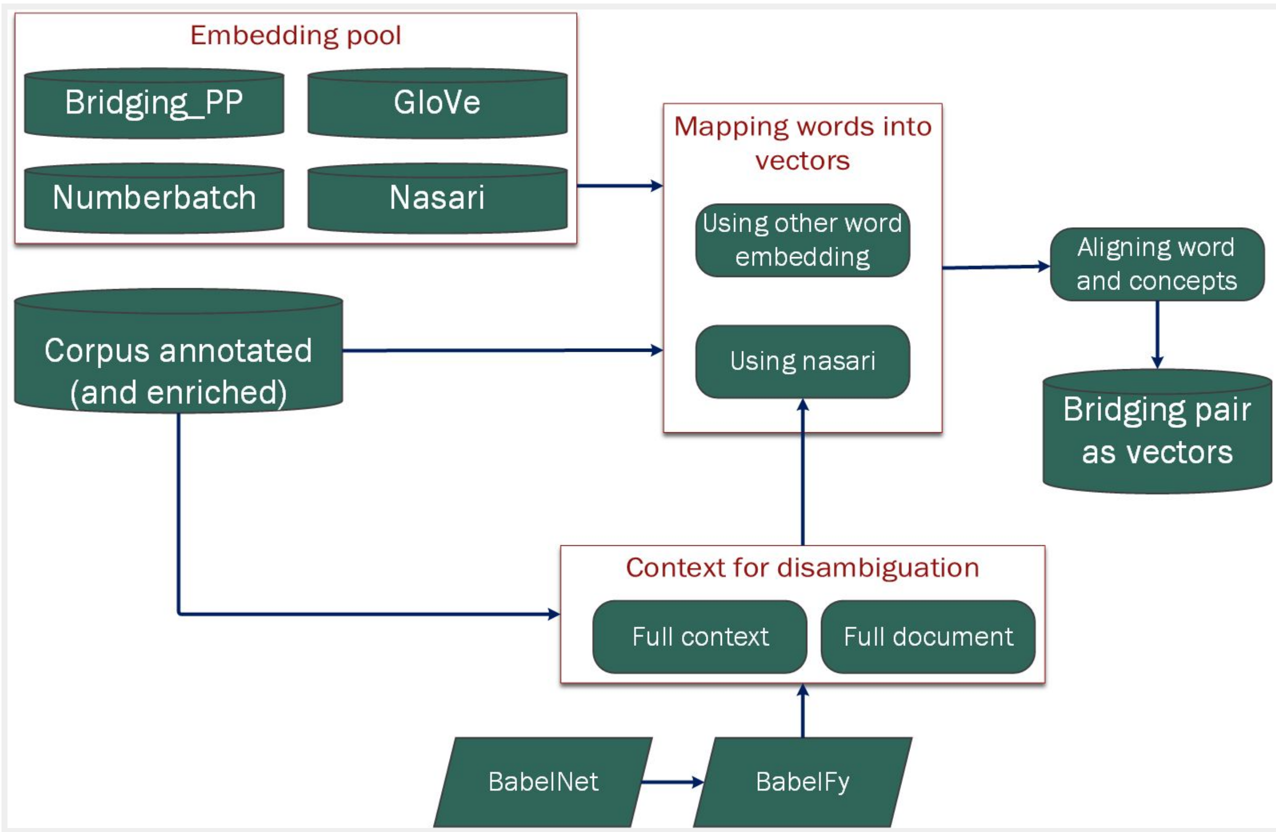

Figure 10. Model architecture: vector computation

Additionally this module allowed us to perform easy combination of word representations, therefore, some experiments were performed using a combination of sense embeddings with vanilla embeddings and specialized embeddings.

Finally, the strategy to perform predictions is calculate the distance between the candidate vector and the bridging anaphora vector for each bridging pair. To this end a distance matrix is built per each bridging pair, each matrix consists of the bridging anaphora in one axis and all its respective candidates in the other. We also filtered some candidates based on rules such as, bridging anaphora usually do not overlaps and DATE anaphoras refer to DATE candidates. Then, we proceed to measure the similarity for each bridging pair, the function by default to measure similarity between two vectors is Cosine distance. In this work we also tried Euclidean similarity and Manhattan Manhattan (Craw, 2017) similarity, as a way to test this stage of the pipeline too. In order to obtain these last two similarity measures we transformed the distance formula into similarity using this formula: $Similarity = 1 - Distance$. The respective formula are shown below. Results comparing the performance for this distance metrics are reported in Section 5.2.

Cosine similarity: $\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum\limits_{i=1}^{n} A_i B_i}{\sqrt{\sum\limits_{i=1}^{n} A_i^2}\sqrt{\sum\limits_{i=1}^{n} B_i^2}}$;

Manhattan distance: $d_1(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|_1 = \sum\limits_{i=1}^{n} |p_i - q_i|$;

Euclidean distance :

$$d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}.$$,

Where A and B are vectors for cosine similarity, and p and q are vectors for distance measures. The Figure 11 offers a graphical descriptions for these operations.

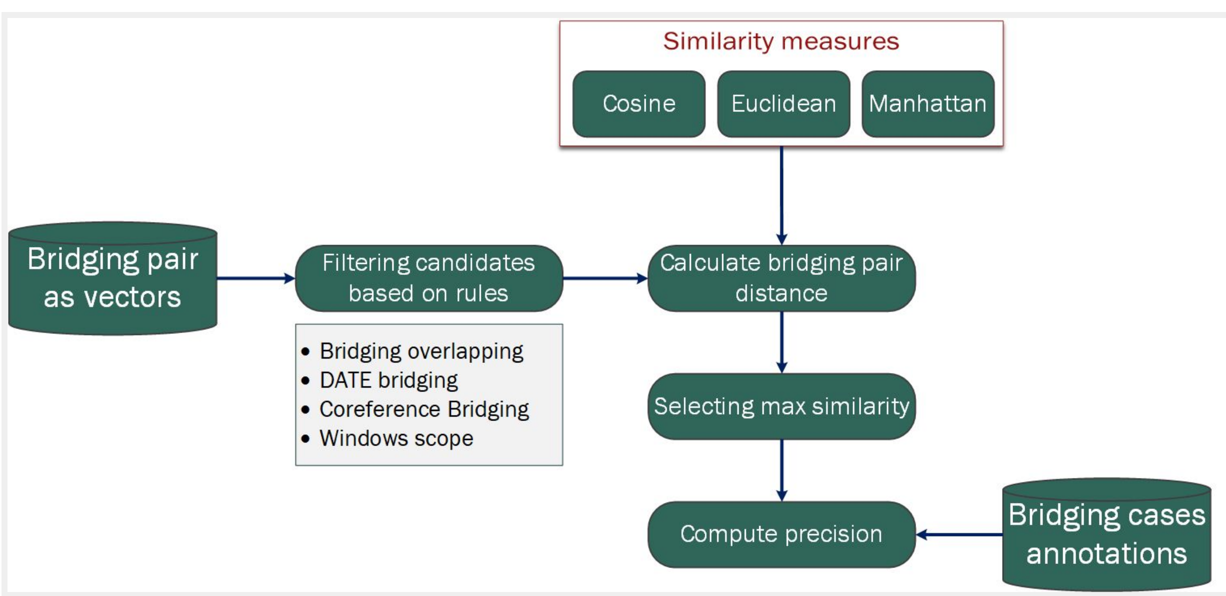

Figure 11. Model architecture: model predictions

5 Results

5.1 Annotation of bridging cases

Motivated by understanding the complexity of bridging phenomenon, and how good are embedding models to capture them, we perform a manual annotation over the 663 bridging anaphoras samples provided by ISNotes. This annotation consisted of three bridging cases: lexical, distributional and contextual which indicate the kind of relation, a given bridging anaphora holds with its antecedent. Table 3 presents our results of annotating bridging cases.

Table 3. Distribution of bridging cases in ISNotes

Case	Occurrences	Relative frequency
Knowledge	283	42%
Distributional	116	18%
Contextual	263	40%
total	663	100%

Our annotation of bridging cases for ISNotes corpus found the existence of an important quantity of “contextual cases”, (which between the three cases are the most difficult to infer, according to the arguments we gave in Section 4.1), just surpassed by the number of “knowledge cases”. On the other hand the number of “distributional cases” is small, we found it specially laborious annotating these examples, due to the difficulty to discriminate distributional from knowledge and contextual cases, because such examples usually exhibit feature from any or both of them, as we explained before. The analysis of interanottation agreement could improve this classification and discriminate better distributional cases. Additionally, this classification for ISNotes is more balanced in comparison to ARRAU corpus, which was judged by Rösiger et al. (2018a) because of its bias towards lexical case. Finally, in the subsequent Section, we plan to use this distribution to evaluate the strength or weakness of the models we considered.

5.2 Results analyses

Here we present the results for the main experiments we performed as well as a brief discussion of the findings derived from each of them. We point out that the kind of experiments we considered were focused on determining how much of the inclusion of lexical information can benefit a bridging anaphora solver, as well as determine what modules in the pipeline (from preprocessing the input throughout the prediction) are more sensitive to influence the performance according to several embedding representation, therefore, we made the following experiments:

- Enrich the bridging pair with coreference information using three different settings
- Test the influence of disambiguation module through more and less context input
- Verify the relevance of the similarity metric used through three metrics, namely Cosine, Euclidean and Manhattan.
- Quantify the contribution of each model to the bridging cases according to the proposed classification.

For the purposes of this experiments we designed a baseline which endeavored to replicate the results reported in Hou (2018b) and Hou (2018c) due to no source code of their models is provided. However, the performance obtained for each of these previous works is far from the reported in both cases. Hence, the baseline used for the experiments corresponds to the best result obtained from Hou (2018b) replication, since this got results closer to those reported than reported in Hou (2018c). The advantage of using a baseline for comparisons purposes instead of the state of the art results, is that it makes possible to experiment in more controlled and fair scenarios, since all the models are receiving the same input, i.e., no difference for preprocessing nor hyper-parameters, rather than those required specifically for the model.

We also point out that all the results are using accuracy, which is defined as the number of solved bridging anaphora over the total of bridging anaphoras. ISNotes has a total of 663 bridging anaphoras. Table 4 shows the first group of experiments, the inclusion of coreference information.

Table 4. Experiments using different bridging representation settings for coreference

Embedding	Without coreference	Union of tokens	Top n open words	The largest coreferent	Mention + Union of tokens
Baseline	19.30 [9]	17.95	19.16	19.46	**20.66**
GloVe	14.03	13.57	13.88	**14.48**	**14.48**
Numberbatch	14.93	**15.23**	14.48	13.88	**15.23**
Nasari	12.52	12.82	11.61	12.82	**13.27**
Nasari + Numberbatch	16.29	16.74	15.69	16.29	**17.35**
Nasari + Glove	15.08	14.78	14.03	15.54	**15.69**
Nasari + Baseline	13.12	11.76	10.86	13.12	**13.27**

In Table 4. we observe that all the models improved with the strategy called “mention + union of tokens”. Therefore, from this point forward we will continue using the best strategy to enrich bridging anaphora using coreference information. We also observed Nasari improves results of GloVe and Numberbatch when their vectors are concatenated, however it does not occur in combination with the baseline. Some other patterns can be evinced, such as, GloVe and the baseline, decrease the performance when using “Union of tokens”, but it increase for Nasari and Numberbatch. The opposite occurred, when “the largest coreference” was used. We noted that similar patterns persist over all the subsequent experiments for baseline-GloVe as a group and Nasari-Numberbatch as another.

The following table summarizes the result derived from experiment with the amount of context input for the disambiguation process. According to our definition, “full context” means the amount of text input for disambiguation encompasses all the words in the context between the elements of the bridging pair, and a window of words around anaphor (see Section 4.2.3). On the other hand, “full document” means all the document was introduced and each bridging pair extracted. Ultimately, “full context + full document” means both disambiguation process are done independently and the resulting senses are aggregated under union operation for mentions, and their corresponding vectors aggregating by average.

[9] Results in Hou (2018b) report 30.32, i.e., a marge of 11.02 for accuracy.

Table 5. Experiments using different context input for disambiguation

Embedding	Full context	Full document	Full context + full document
Nasari	13.27	13.42	**14.33**
Nasari + Numberbatch	17.35	17.65	**18.10**
Nasari +GloVe	**15.69**	14.33	15.08
Nasari + baseline	**13.27**	12.22	12.37

In Table 5 we observed that the considered strategies increase slightly the performance of Nasari, and Nasari in combination with Numberbatch, however they decrease it in combination with GloVe and the baseline. Consequently, the blending of both strategies in “full context + full document” performs similarly. A possible explanation for this behavior is about the source of the embeddings: Numberbatch and Nasari, are able to leverage global and local context maybe due to both are based on lexical information stored in a graph structure, while GloVe relies completely on distributional information, this argument holds for the baseline too, because “embeddings_PP” was built using GloVe algorithm. From this point forward we will continue using the best strategy of context disambiguation for each model concatenated with Nasari. i.e., in the subsequent experiments “full context + full document” will be employed for Nasari and Nasari combined with NumberBatch and we will use “Full context” when Glove or the baseline be combined with Nasari.

In this experiment, we also noted as expected, that results from full context disambiguation were more retrieved more senses, hence more multiword expressions and overlapping senses were found too. On the contrary, the full document provided more general and less amount of senses.

The next experiment presents alternatives for the similarity measure, since all the previous attempts were done using cosine similarity, we verify whether this metric is better for each of the models. Specifically, we tested with Euclidean and Manhattan metrics. The results confirm that for the models, cosine similarity performs better than the other tested measures, and Euclidean performs slightly better than Manhattan.

Tabla 6. Results comparing the performance for different similarity measures

Embedding	Cosine	Euclidean	Manhattan
Baseline	**20.66**	19.31	18.70
GloVe	**14.48**	13.88	13.88
Numberbatch	**15.23**	14.48	14.33
Nasari	**14.33**	14.18	14.18
Nasari + Numberbatch	**18.10**	17.50	17.19
Nasari + Glove	**15.69**	15.08	14.93
Nasari + Baseline	**13.27**	12.52	12.07

In Table 6, we show results for the experiment using similarity measures, here we quantify how much each model contributes to solve each type of bridging anaphora, according to the proposed classification. The first column correspond to the best result obtained per each model throughout the previous experiments, while the remaining 3 columns inform about how such results are distributed over the cases.

Table 7. Comparison of model performance for each bridging case

Embedding	Best result	Knowledge cases	Distributional cases	Contextual cases
Baseline	20.66	**11.46**	**4.98**	4.22
GloVe	14.48	8.90	3.02	2.56
Numberbatch	15.23	9.80	2.87	2.56
Nasari	14.33	6.49	3.32	4.52
Nasari + Numberbatch	18.10	8.14	**4.83**	**5.13**
Nasari + Glove	15.69	7.84	3.47	4.37
Nasari + Baseline	13.27	7.39	2.11	3.77

In Table 7 we observe the baseline performs better for Knowledge and distributional cases, while Nasari and those models combined with Nasari, exhibit a better result in contextual cases than the baseline, the best of them is Nasari + Numberbatch. We attribute the reason that Nasari had a good performance for contextual cases, to the fact that it uses the full context when the disambiguation step is performed.

Additionally, we observed that for all the experiments the baseline has the highest performance on average, but it presents a shortcoming in comparison with Nasari models for contextual cases. Nasari interestingly exhibit a good performance alone and in combination with other embeddings, however it exhibits a very bad performance every time it is combined with the baseline. Thus, they seem to be capturing opposite perspective of the bridging anaphora phenomenon, but they cannot complement each other by the simple means of concatenation.

6 Conclusions

6.1 Discussion

Bridging anaphora resolution is the challenging task of detecting the mentioned antecedent, a particular expression is indirectly referring to. Here in this work we investigate the benefits of using lexical information for unrestricted bridging anaphora resolution in unsupervised fashion. Specifically, we performed experiments using sense embeddings from Nasari (Camacho-Collados et al. 2015) and we combine it with vanilla word embedding such as GloVe and Numberbatch, and specialized word embeddings from Hou (2018b), which are specially designed for capturing bridging aspects by the use of prepositions patterns. The resulting model is able to improve the detection of bridging antecedents in some cases where the anaphor require a deeper understanding of the context in which the bridging anaphora takes place.

For comparison purposes, the results are reported using a baseline derived from the state-of-the-art unsupervised model, since no code is provided and our replication could not obtain the same performance which is claimed in that work. The results showed that lexical information derived from sense embeddings alone is not enough to solve bridging anaphora, but it can improve the detection of bridging antecedents in combination with vanilla word embeddings, which are capturing different linguistic aspects for bridging anaphora resolution. We also analyzed the specific cases in which this lexical information results beneficial to the task, resulting in that sense embedding information helps in cases where a higher context understanding is required. Additionally, we showed that a combination of sense embeddings with vanilla embeddings outperforms the baseline in “contextual cases”, and offers competitive results for detecting “distributional cases” However the performance decreases in combination with the specialized embeddings used for the baseline. We also show that the inclusion of coreference information (using gold labels) improve the performance for the baseline, therefore we hypothesize include this information in the original model would also improve the state of the art results.

We deem more research is necessary for solving bridging anaphora, and we encourage the community to develop more computational proposal for modelling this problem. We also recommend the inclusion of lexical resources in future research for bridging anaphora, because they. Below we summarize the contribution of this work and our ideas for future work on bridging anaphora are exposed.

6.2 Contribution

The outcome of this work offer a new sight about using lexical information for solving unrestricted bridging anaphora in unsupervise fashion in ISNotes corpus. The proposed model, although does not outperform our baseline derived from the state of the art model, is able to obtain a higher score for resolving special cases where the bridging anaphora involves a complex context. Additionally, the proposed model achieves competitive results for detecting distributional cases, without the need of specialized resources for bridging anaphora. Additionally, we propose a new classification scheme for bridging anaphora which divide bridging occurrences into 3 cases and that help to achieve a better understanding of the modelling capabilities of a bridging anaphora model. The classification is provided as additional annotation for ISNotes corpus.

6.3 Future Work

During the realization of this work, some relevant ideas to model bridging anaphora were postponed for being out of scope for the present research. Here we are presenting some of such ideas.

- Test the performance of the proposed model in other datasets, even it could be tried in non-english dataset, since babelnet is multilingua and Babelfy disambiguations can run in a language-agnostic scenario.
- Sense embeddings and specialized bridging embeddings seems to be capturing complementary information, yet a concatenation of them is not compatible, or at least it cannot increase the performance. Hence, we speculate an early fusion to train together specialized and senses embeddings could lead to better results.
- Improve the reliability of our bridging classification scheme by means of annotator agreement evaluations. This will also help to address issues like understanding the borders between distributional cases against lexical and complex cases.

- An alternative to improve our annotation reliability is to develop a model able to predict automatically the bridging case for each sample.
- Re-implement the model in a supervised fashion employing a set of feature engineering too.
- Propose a new model based on discursive information, able to leverage discursive features extracted from previous context. This idea could involve the analysis of trees structures provided by Rhetorical Structure Theory (RST) and Semantic Role Labelling (SRL) as auxiliary task.

References

Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., & Wiswedel, B. (2007). Studies in Classification, Data Analysis, and Knowledge Organization.

Camacho-Collados, J., & Pilehvar, M. T. (2018). From word to sense embeddings: A survey on vector representations of meaning. Journal of Artificial Intelligence Research, 63, 743-788.

Camacho-Collados, J., Pilehvar, M. T., & Navigli, R. (2015). Nasari: a novel approach to a semantically-aware representation of items. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.

Clark, H. H. (1975). Bridging. In Theoretical issues in natural language processing.

Collados, J. C. (2017). Semantic Vector Representations of Word Senses, Concepts and Entities and their Applications in Natural Language Processing.

Craw, S. (2017). Manhattan Distance. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning and Data Mining. Springer, Boston, MA.

Domingos, P., & Lowd, D. (2009). Markov logic: An interface layer for artificial intelligence. Synthesis Lectures on Artificial Intelligence and Machine Learning, 3(1), 1-155.

Gelbukh, A., & Sidorov, G. (1999). On Indirect Anaphora Resolution. In Proc. PACLING-99, Pacific Association for Computational Linguistics (pp. 181-190).

Guha, A. (2017). Data and Methods for Reference Resolution in Different Modalities (Doctoral dissertation).

Hawkins, J. A. (1978). Definiteness and indefiniteness: A study in reference and grammaticality prediction. London: Croom Helm.. 1991. On (in) definite articles: Implicatures and (un) grammaticality prediction. Journal of Linguistics, (vol)27, 405-442.

Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. In Proceedings of the 14th conference on Computational linguistics (vol)2, 539-545. Association for Computational Linguistics.

Honnibal, M., & Johnson, M. (2015). An improved non-monotonic transition system for dependency parsing. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (pp. 1373-1378).

Hou, Y., Markert, K., & Strube, M. (2013). Cascading collective classification for bridging anaphora recognition using a rich linguistic feature set. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (pp. 814-820).

Hou, Y. (2018). Enhanced Word Representations for Bridging Anaphora Resolution. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, (vol)2, 1-7.

Hou, Y. (2018). A Deterministic Algorithm for Bridging Anaphora Resolution.Extracted on (9/10/2019). arXiv preprint arXiv:1811.05721.

Hou, Y., Markert, K., & Strube, M. (2013). Global inference for bridging anaphora resolution. In Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies (pp. 907-917).

Hou, Y., Markert, K., & Strube, M. (2014). A rule-based system for unrestricted bridging resolution: Recognizing bridging anaphora and finding links to antecedents. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 2082-2093).

Hou, Y., Markert, K., & Strube, M. (2018). Unrestricted bridging resolution. Computational Linguistics, 44(2), 237-284.

Hovy, E., Navigli, R., & Ponzetto, S. P. (2013). Collaboratively built semi-structured content and Artificial Intelligence: The story so far. Artificial Intelligence, (vol)194, 2-27.

Irmer, M. (2011). Bridging inferences: Constraining and resolving underspecification in discourse interpretation (vol)11, pp.

Jurafsky, D., & Martin, J. H. (2009). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition. Prentice Hall series in artificial intelligence.

Lassalle, E., & Denis, P. (2011). Leveraging different meronym discovery methods for bridging resolution in French. In Discourse Anaphora and Anaphor Resolution Colloquium (pp. 35-46). Springer, Berlin, Heidelberg.

Markert, K., & Nissim, M. (2005). Comparing knowledge sources for nominal anaphora resolution. Computational Linguistics, 31(3), 367-402.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119).

Miller, G. A. (1995). WordNet: a lexical database for English. Communications of the ACM, 38(11), 39-41.

Mirkin, S., Dagan, I., & Padó, S. (2010). Assessing the role of discourse references in entailment inference. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (pp. 1209-1219).

Moro, A., Raganato, A., & Navigli, R. (2014). Entity linking meets word sense disambiguation: a unified approach. Transactions of the Association for Computational Linguistics (vol)2, 231-244.

Nand, P., & Yeap, W. (2012). A Framework for Interpreting Bridging Anaphora. In International Conference on Agents and Artificial Intelligence (pp. 131-144). Springer, Berlin, Heidelberg.

Navigli, R. (2009). Word sense disambiguation: A survey. ACM computing surveys (CSUR), 41(2), 10.

Navigli, R., & Ponzetto, S. P. (2010). BabelNet: Building a very large multilingual semantic network. In Proceedings of the 48th annual meeting of the association for computational linguistics (pp. 216-225). Association for Computational Linguistics.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).

Pilehvar, M. T., Jurgens, D., & Navigli, R. (2013). Align, disambiguate and walk: A unified approach for measuring semantic similarity. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, (vol)1, pp.1341-1351).

Poesio, M. (2004). The MATE/GNOME proposals for anaphoric annotation, revisited. In Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue at HLT-NAACL 2004 (pp. 154-162).

Poesio, M., Mehta, R., Maroudas, A., & Hitzeman, J. (2004). Learning to resolve bridging references. In Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04) (pp.143-150).

Poesio, Massimo, et al. The ARRAU Corpus of Anaphoric Information LDC2013T22. Web Download. Philadelphia: Linguistic Data Consortium, (2013). Extracted on (9/10/2019).

Poesio, M., Stuckardt, R., & Versley, Y. (2016). Anaphora resolution. Springer.

Prince, E. (1981). Towards a taxonomy of given-new information. In Peter Cole, ed., Radical Pragmatics, (pp. 223–256). Academic Press, New York.

Rösiger, I. (2016). Scicorp: A corpus of english scientific articles annotated for information status analysis. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) (pp. 1743-1749).

Rösiger, I., Riester, A., & Kuhn, J. (2018). Bridging resolution: Task definition, corpus resources and rule-based experiments. In Proceedings of the 27th International Conference on Computational Linguistics. (pp. 3516-3528).

Rösiger, I., Köper, M., Nguyen, K. A., & im Walde, S. S. (2018). Integrating predictions from neural-network relation classifiers into coreference and bridging resolution. In Proceedings of the First Workshop on Computational Models of Reference, Anaphora and Coreference. (pp. 44-49).

Saeboe, K. J. (1996). Anaphoric presuppositions and zero anaphora. Linguistics and Philosophy, 19(2), 187-209.

Sasano, R., & Kurohashi, S. (2009). A probabilistic model for associative anaphora resolution. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: (vol)3, 1455-146. Association for Computational Linguistics.

Sidorov, G., Gelbukh, A., Gómez-Adorno, H., & Pinto, D. (2014). Soft similarity and soft cosine measure: Similarity of features in vector space model. Computación y Sistemas, 18(3), 491-504.

Singer, M. (1979) Processes of inference during sentence encoding. Memory & Cognition, 7(3), pp.192–200.

Soricut, R., & Marcu, D. (2006). Discourse generation using utility-trained coherence models. In Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions (pp. 803-810).

Speer, R., Chin, J., & Havasi, C. (2017). Conceptnet 5.5: An open multilingual graph of general knowledge. In Thirty-First AAAI Conference on Artificial Intelligence.

Sukthanker, R., Poria, S., Cambria, E., & Thirunavukarasu, R. (2018). Anaphora and Coreference Resolution: A Review. Extracted on (9/10/2019). arXiv preprint arXiv:1805.11824.

Sun, W., & Uszkoreit, H. (2012). Capturing paradigmatic and syntagmatic lexical relations: Towards accurate Chinese part-of-speech tagging. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, (vol)1, 242-252. Association for Computational Linguistics.

Weischedel, R., Palmer, M., Marcus, M., Hovy, E., Pradhan, S., Ramshaw, L., & El-Bachouti, M. (2011). OntoNotes 4.0. Linguistic Data Consortium LDC2011T03. Extracted on (9/10/2019).

Glossary

Various technical terms used in this thesis are briefly defined below[10]:

- Textual entailment (TE): “it is a directional relation between text fragments. The relation holds whenever the truth of one text fragment follows from another text.”
- Natural language processing (NLP): “it is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.”
- word sense disambiguation (WSD): “it is an open problem concerned with identifying which sense of a word is used in a sentence.”
- Name Entity Recognition (NER): “it is a subtask of information extraction that seeks to locate and classify named entity mentions in unstructured text into pre-defined categories such as the person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc.”
- Markov Logic Networks (MLN): “it is a probabilistic logic which applies the ideas of a Markov network to first-order logic, enabling uncertain inference.”
- Entity linking (EL): “it is the task of recognizing (cf. Named Entity Recognition) and disambiguating (Named Entity Disambiguation) named entities to a knowledge base (e.g. Wikidata, DBpedia, or YAGO).”
- Probabilistic Graphical Models (PGM): “it is a probabilistic model for which a graph expresses the conditional dependence structure between random variables. They are commonly used in probability theory, statistics—particularly Bayesian statistics—and machine learning.”

[10] For simplicity of the explanations, all these definitions usings quotation marks are taken from Wikipedia.