Natural Language Processing

Chapter 7 Pragmatic Analysis & Discourse

DR RAYMOND LEE
ASSOCIATE PROFESSOR, DST
BNU-HKBU UNITED INTERNATIONAL COLLEGE

Pragmatic Analysis & Discourse

1. Introduction to Pragmatics and Discourse
2. Discourse Phenomena
3. Coherence and Coreference
4. Discourse Segmentation
5. Discourse Coherence
6. Algorithms for Coreference Resolution
7. Evaluation
8. Summary

Introduction
Pragmatics Analysis & Discourse

- Pragmatics and Discourse Analysis involve the study of language in its contexts of use.
- Pragmatics focuses on the effects of context on meaning, and Discourse Analysis studies written and spoken language in relation to its social context.
- In a broder meaning, they refer to the “context” meaning of the sentence and utterance.
- As difference from the word-level and sentence-level meaning representation and analysis learnt in the previous chapters.
- Consists of collocated, structured, coherent groups of sentences
  - What makes something a discourse as opposed to a set of unrelated sentences?
  - How can text be structured (related)?
- In our daily life, there are two types of discourse.
- Monologue: a speaker (writer) and hearer (reader) with communication flow in one direction only
  - E.g. Reading (writing) this book, Watching a TV show or a play, listening to a speech, attending a presentation, attending a lecture (may or may not be, it depends).
- Dialogue: each participant takes turn being the speaker and the hearer (so 2-way participation, sometimes 3-way communication or more)
  - Human-human dialogue e.g. daily conversations, bargaining in the market, group discussions, etc.
  - Human-computer dialogue
    => HCI (Human-Computer Interaction) e.g. conversational agent, chatbot
    => CCI (Computer-Computer Interaction) e.g. future cross machine verbal communication in smart city and smart transportation, multi-agent based negotiation and bargaining, etc.

Monologue vs Dialogue

Discourse Phenomena
Coreference Resolution

- In daily life, we come across many discourse phenomena in which we can "naturally resolved" them as human, but it needs to take a lot of effort for machine to do that.
- One of the most fundamental and important discourse phenomena need to be handled is so-called "Coreference Resolution".
- So, what is Coreference Resolution (CR)?
- Coreference resolution is the task of finding all linguistic expressions (so-called mentions) in a given text that refer to the same real-world entity.
- After finding and grouping these mentions we can resolve them by replacing, as stated above all these pronouns with correct noun phrases.
- It seems to be simple (or human perspective), but always makes mistakes by machines.
- For example:

[7.1] Jack saw Michael in the examination hall. He looked nervous.

[7.2] Jack saw the student in the examination hall. He looked nervous.

In normal use of English, we (and machine)will most likely take the "first subject" of the pervious sentence as the reference to the Pronoun of next sentence.

That's, [7.1] will refer "He" as "Jack. However, in [7.2], coreference resolution in human perspective will say "He" might be more likely refer "the student" instead of Jack as we know "student" will take examination.

This example is more obvious:

[7.3] Jane talked to Janet about her examination result. She looked worry.

[7.4] Jane talked to Janet about her examination result. She felt sorry about her.

In [7.3], most probably, "She" should be referred to "Janet" as Janet is the one taking the exam, she instead of Jane should be worried.

In [7.4], most probably, "She" should be referred to "Jane" instead as Janet is the one taking the exam, she is more likely to feel sorry.

- Of course, all of these are driven by 1) Context; 2) World Knowledge (common-sense).
- For Human, it is totally natural thing, but how about computer. Can computer do that judgement?

Discourse Phenomena
Coreference Resolution

- Before we handle complex coreference resolution case, let's have a look on some simple and standard situation like this:

[7.3] Jack gives Tom 1000 dollars. He is generous. (original sentence)

[7.4] Jack gives Tom 1000 dollars. Jack is generous. (with coreference resolution)

Or more compact case that can be handled by computer nicely.

[7.5] "I voted for Obama as he is more aligned to my values", Tom said. (original sentence)

[7.6] "Tom voted for Obama as Obama is more aligned to Tom's values", Tom said. (with coreference resolution)

In The Adventures of Sherlock Holmes, we have:

[7.7] I was seized with a keen desire to see Holmes again, and to know how he was employing his extraordinary powers. (original sentence)

[7.8] Watson was seized with a keen desire to see Holmes again, and to know how Holmes was employing Holmes' extraordinary powers. (with coreference resolution)

Or more challenging sentence: the famous discourse in the story *A Scandal in Bohemia*:

[7.9] To Sherlock Holmes she is always "the woman". I have seldom heard him mention her under any other name. (original sentence)

[7.10] To Sherlock Holmes she is always "the woman". Watson have seldom heard Holmes mention Irene Adler under any other name. (original sentence)

This one is more challenging as the reference name "Irene Adler" is appeared not before, but after two sentences afterwards : It was not that he felt any emotion akin to love for Irene Adler.

In linguistics, we called such phenomena "Cataphora" - The word that gets its meaning from a subsequent word or phrase is called a cataphor. The subsequent word or phrase is called the antecedent or referent. Contrast to it is so-called "anaphora" – in which a rhetorical term for the repetition of a word or phrase at the beginning of successive clauses, which is normally use in many English sentence construction such [7.3], [7.5], [7.7] for which we mention the reference term before we repeat using it by using (says) a pronoun as replacement.

- Why is it important?
  - Information extraction
  - Text Summarization
  - conversational agents (Chatbots)
- Coreference resolution is an exceptionally versatile tool and can be applied to a variety of NLP tasks such as text understanding, information extraction, machine translation, sentiment analysis, or document summarization. It is a great way to obtain unambiguous sentences which can be much more easily understood by computers.

Coherence & Coreference
What is Coherence?

What is Coherence?

- In linguistics, the term "coherence" is used to refer to sense relations between single units (sentences or propositions) of a text.
- Due to these relations, the text appears to be logically and semantically consistent for the reader (or hearer).
- Text analysis focussing on coherence is primarily concerned with the construction and configuration of sense in the text i.e. how its single constituents are connected so that the text becomes meaningful for the addressee rather than being a random sequence of unrelated sentences and clauses.
- If something has coherence, its parts are well-connected and all heading in the same direction.
- Without coherence, a discussion may not make sense or may be difficult for the audience to follow.
- It's an extremely important in verbal language and writing.
- Coherence is relevant to every level of organization, from the sentence level up to the complete argument.

Example of short essay with Coherence:

[7.11] <u>History</u> shows that human beings have come a long way from where they started. <u>They</u> have developed new <u>technologies w</u>hich means that everybody can enjoy luxuries they never previously imagined. However, the <u>technologies</u> that are temporarily making this world a better place to live could well prove to be an ultimate disaster due to, among other things, the creation of <u>nuclear weapons</u>, increasing pollution, and loss of animal species.

- In this short essay, it uses coherence terms History->They->technologies->nuclear weapons with repeated terms and concepts to provide a "stream of flow" of ideas and knowledge, so that the hearer (or reader) is more easy to understand the message that being conveyed.

Coherence & Coreference
What is Coreference?

What is Coreference?

- To achieve coherence, one of the most important techniques is coreference.
- In linguistics, coreference (or co-reference) occurs when two or more expressions refer to the same person or thing; they have the same referent.
- For example:

[7.12] Jack said Helen would arrive soon, and she did. - the words *Helen* and *she* refer to the same person.

- In reality, coreference is not always trivial to determine.
- For example:

[7.13] Jack said he would join the team vs

[7.14] Jack told Jim to come, he smiled.

- [7.13] is rather trivial as there is only 1 subject noun in the previous clause, but how about [7.14], either he may be referred to Jack or Jim, which one is correct?
- Determining which expressions are coreferences is an important part of analysing or understanding the meaning, especially in NLP application such as Information Extraction, Text summarization and the understanding of the dialogues in chatbot systems.

Coherence & Coreference

Importance of Coreference Relations

Importance of CR to Text Summarization (TS) and Information Extraction (IE)

- **How can we summarize the following text or extract the key information from this text?**

[7.15] XYZ bank is continuing to struggle with severe financial problems. According to the Business Insider's report, their CEO, Michael Crowley, will announce to step-down during the press conference being hold early tomorrow morning.

- As a typical news article, such texts are mostly highly coherence with well-structured coreference terms and concepts
- The coherence concept terms involve, which is also the key information in terms of IE:

[XYZ bank] -> [Financial Problem] -> [CEO] -> [Michael Crowley] -> [step-down] -> [press conference] -> [tomorrow morning]

- Reasonable summary will be:

[7.16] The CEO of XYZ bank Michael Crowley will announce his step-down in tomorrow morning press conference.

- What you need to know: coherence relations between text segment – the first sentence is providing background for the more important 2nd sentence.
- Remarks: In a well-structural case of Text Summarization / Information Extraction, such extracted text and summary will be (and should be) matched with the Fillmore's Case Roles Theory with well defined: Agent, Patient, Location (Time), Purpose, Beneficiary, Possessor, Instrument, etc.
- Or, in other words: A well coherence text message and utterance will be well-structured in the sense that the first sentence is the "opening" of the "speech", the following sentences (utterance) will be the further elaboration of the "open statement" – so-called "elaboration" in terms of coreference relation and with "thematic relation" just like watching a movie of TV show.

Coherence & Coreference
Importance of Coreference Relations

Importance of CR for Discourse with Inference Statement

- In addition to elaborative and thematic type of Coreference Relation (CR), another important type of coherence and CR is the so-called inference-type discourse statement.
- A typical inference-type discourse statement usually with the first sentence (utterance) as the "claims" and the following sentences (utterance) is (are) the "explanation" of that claims sentence.
- In terms of inference argument, the first sentence is the "effect", the following sentence(s) is (are) the "cause".
- Example

[7.17] Jack keeps Jim's car key. He was drunk last night. (Highly coherence) vs

[7.18] Jack keeps Jim's car key. He wants to see movie tonight. (Not-coherence)

- In these two examples, [7.17] is highly coherence in the sense that the first sentence is highly related to the second sentence in terms of pragmatic meaning
- By using "common sense" (world knowledge), it is "most probably" the second statement is the explanation (cause) of the event described in the first sentence, in which Jack keeps Jim's car key because Jim is drunk last night. By simple inference, "He" should be "Jim" instead of "Jack.
- While in [7.18], the two sentences are not coherence at all, and we can't see any logical relationship (such as "cause-effect" relationship) between the first and the second statement.
- In that case, it is difficult to judge whether "He" in the second sentence should be referred to "Jack" or "Jim".
- In terms of standard use of English, we will treat the main subject in the previous statement, ie. "Jack" will be the referent of "He" in the second statement. Although it might be incorrect.

Coherence & Coreference
Entity-based Coherence

Entity based Coherence

[7.19] John went to his favourite music store to buy a piano.

[7.20] He had frequented the store for many years.

[7.21] He was excited that he could finally buy a piano.

[7.22] He arrived just as the store was closing for the day.

[7.23] John went to his favourite music store to buy a piano.

[7.24] It was a store John had frequented for many years.

[7.25] He was excited that he could finally buy that piano.

[7.26] The music played by it is great.

[7.27] It was closing just as John arrived.

- Entity-based coherence models measure this kind of coherence by tracking salient Centering Entities (CE) across a discourse.
- In the Centering Theory (Grosz et al., 1995), the most influential theory of entity-based coherence, keeps track of which entities in the discourse model are salient at any point (salient entities are more likely to be pronominalized or to appear in prominent syntactic positions like subject or object).
- As seen in the above example, CE from [7.19] to [7.23] is John, so naturally John will be the referent for "he" in these statement.
- While in [7.24] the CE is shifted to "favourite music store" and to "the piano" in [7.25] and [7.26].
- Make it more complex, the CE is shifted back to the "music store" in [7.26]
- We will discuss the Centering Theory in detail in the coming section.

Discourse Segmentation
What is Discourse Segmentation?

What is Discourse Segmentation?

- Discourse segmentation is the task of determining minimal non-overlapping units of discourse called elementary discourse units (EDUs).
- It can be further subdivided into sentence segmentation and sentence-level discourse segmentation.
- In short, the core purpose of Discourse Segmentation is to separating a document into a linear sequence of subtopics.
- This is often a simplification of a higher-level structure of a discourse
- For example: academic articles are usually segmented into Abstract, Introduction, Methodology, Implementation, Results, Discussion, Conclusion, etc for the ease of understanding and comprehension.
- Methods of Discourse Segmentation:
  - Unsupervised Discourse Segmentation
  - Supervised Discourse Segmentation
- Applications of automatic discourse segmentation:
  - Information Retrieval (IR) or Information Extraction (IE): Apply to an appropriate segment
  - Text Summarization (TS): Summarize each segment separately

Discourse Segmentation
Unsupervised Discourse Segmentation

Unsupervised Discourse Segmentation

- The class of unsupervised discourse segmentation is often represented as linear segmentation.
- Given raw text, segment it into multiple paragraph subtopics.
- Unsupervised: No training data is given for the task
- We can understand the task of linear segmentation with the help of an example.
- In the example, there is a task of segmenting the text into multi-paragraph units; the units represent the passage of the original text.
- These algorithms are dependent on cohesion that may be defined as the use of certain linguistic devices to tie the textual units together.
- Cohesion-based approach: Segment into subtopics in which sentences/paragraphs are cohesive with each other.
- On the other hand, lexicon cohesion is the cohesion that is indicated by the relationship between two or more words in two units like the use of synonyms.
- Cohesion: Links between text units due to linguistic devices
- Lexical Cohesion: Use of same or similar words to link text units, by using identical word, synonym or hypernym. For example:

[7.28] Yesterday was Jane's birthday. Betty and Janet went to the gift shop. They were going to get presents. Janet decided to get a purse. "Don't do that," said Betty. "Jane has a purse. She will make you take it back."

- Non-lexical Cohesion: For example, using anaphora. For example:

[7.29] Peel, core and slice the peaches and the pineapples. Then add these fruit to the skillet.

- Unsupervised Discourse Segmentation: Marti Hearst's TextTiling (done in early 90's)
- In the next section, we will use Hearst's classical works on TextTiling to illustrate how Discourse Segmentation works.

Discourse Segmentation

Hearst's TextTiling Method

Hearst's TextTiling Method

- Hearst's TextTiling (Hearst 1997) is a technique for automatically subdividing texts into multi-paragraph units (ie. Subtopics) that represent the original text passages.
- This discourse segmentation method is based on the idea of so-called subtopic shift, an algorithm for subdividing expository texts into multi-paragraph "passages" or subtopic segments.
- In her original work, by using the articles from a science magazine "Stargazers", TextTiling method tried to characterize the article text massages into various subtopics.
- For example, consider a typical 21-paragraph science news article from Stargazers, with the theme focus on the reporting of the existence of life on earth and other planets.
- Hearst's TextTiling method is a typical unsupervised segmentation method in the sense that no particular training dataset and prior knowledgebase is needed, the segmentation is achieved by the automatic clustering of related text passages into specific topics.
- The contents are being characterized into the following subtopic discussions (Hearst 1997):
  - [Para l- 3] Introduction - the search for life in space
  - [Para 4 – 5] The moon's chemical composition
  - [Para 6 – 8] How early earth-moon proximity shaped the moon
  - [Para 9 – 12] How the moon helped life evolve on earth
  - [Para 13] Improbability of the earth-moon system
  - [Para 14 – 16] Binary/trinary star systems make life unlikely
  - [Para 17 – 18] The low probability of nonbinary/trinary systems
  - [Para 19 – 20] Properties of earth's sun that facilitate life
  - [Para 21] Summary
- TextTiling is a method for partitioning full-length text documents into coherent multi-paragraph units.
- As shown in the above example which correspond to a sequence of subtopical passages.
- The algorithm assumes that a set of words is in use during the course of a given subtopic discussion, and when that subtopic changes, a significant proportion of the vocabulary will be altered accordingly.

Discourse Segmentation

Hearst's TextTiling Method

Hearst's TextTiling Method

- Fig. 7.1 Distribution of selected terms from the Stargazer text, with a single digit frequency per sentence number, with blanks indicate a frequency of zero (Hearst 1997)
- Several important findings can be seen from Fig. 7.1:
  1. Terms that appear frequently throughout the text such as "moon" and "planet" are often indicative of the main topic(s) of the text.
  2. Terms that are less frequent but more uniform in distribution, such as "scientists" and "form" which are too "generic" to be used to form the subtopic title.
  3. For example, terms "space" and "star" appear more frequent from sentences 5-20 and sentences 60-90, while terms "life" to "planet" appear more frequent from sentences 58 to 78 only, which might form two distinct "cluster" of sub-topic discussions.
  4. Similar phenomena occurs for terms "life" to "species" which form a natural cluster between sentences 35 to 55 and conform from the human judgement as the subtopic discussion of "How the moon helped life evolve on earth".
- The logic behind is: Sentences or paragraphs in a subtopic are cohesive with each other, but not with paragraphs in a neighbouring subtopic.

```
--------------------------------------------------------------------------------------------------------------
Sentence:        05   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85   90   95
--------------------------------------------------------------------------------------------------------------
14       form    1     111 1   1                               1 1     1   1       1     1     1    1
 8  scientist                   11            1   1               1          1      1  1
 5      space 11    1     1                                                            1
25       star    1             1                                       11 22  111112  1 1  1   11 1111     1
 5     binary                                                           11  1          1                    1
 4    trinary                                                            1   1         1                    1
 8 astronomer 1                1                                        1 1           1   1    1  1
 7      orbit   1                    1                                     12    1 1
 6       pull                          2     1 1                               1  1
16     planet   1    1       11               1           1                21  11111                  1     1
 7     galaxy    1                                                   1                1  11     1           1
 4      lunar            1  1     1       1
19       life 1  1  1                              1    11 1  11  1      1                 1 1    1 111  1 1
27       moon         13  1111   1 1 22 21  21    21           11 1
 3       move                                       1   1   1
 7  continent                                       2 1 1 2 1
 3  shoreline                                             12
 6       time                      1                1  1  1     1                                          1
 3      water                                     11            1
 6        say                                    1 1        1        11                1
 3    species                                        1  1  1
--------------------------------------------------------------------------------------------------------------
Sentence:        05   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85   90   95
--------------------------------------------------------------------------------------------------------------
```

Fig. 7.1 Distribution of selected terms from the Stargazer text, with a single digit frequency per sentence number (blanks indicate a frequency of zero)

Discourse Segmentation
TextTiling Algorithm

TextTiling Algorithm:

- The TextTiling algorithm for Discourse Segmentation and characterization of subtopic structure using term repetition consists of THREE main processes:
  1. Tokenization
     - convert words to lower case
     - remove stop words and stem words
     - group the words into pseudo-sentences of equal length (say 20 words)
  2. Lexical Score Determination
     - Compute lexical cohesion score at each gap between pseudo-sentences
     - Lexical cohesion score: Similarity of words before and after the gap (take say 10 pseudo-sentences before and 10 pseudo-sentences after)
     - Similarity: Compute the Cosine similarity between the word vectors (high if words co-occur)

$$\text{sim}_{\text{cosine}}(\vec{b},\vec{a}) = \frac{\vec{b}\cdot\vec{a}}{|\vec{b}||\vec{a}|} = \frac{\sum_{i=1}^{N} b_i \times a_i}{\sqrt{\sum_{i=1}^{N} b_i^2}\sqrt{\sum_{i=1}^{N} a_i^2}} \quad (7.1)$$

  3. Boundary Identification – assign a cut-off distance to identify a new segment.
     - Plot the similarity and compute the depth scores of the "similarity valleys", (a-b)+(c-b), as shown in Fig. 7.2
     - Assign segmentation if the depth score is larger than a threshold (Fig. 7.3) (e.g. one standard deviation deeper than mean valley depth)

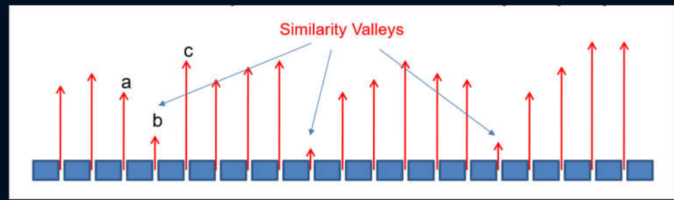

Fig. 7.2 Lexical score determination with Similarity Valleys

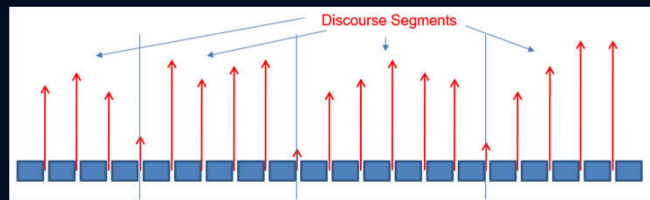

Fig. 7.3 Boundary Identification with Discourse Segments

Discourse Segmentation
Supervised Discourse Segmentation

Supervised Discourse Segmentation

- To be used when it is relatively easy to acquire boundary-labeled training data, for example:
  - News stories from TV broadcasts
  - Paragraph segmentation in texts and dialogues.
  - Useful to find paragraphs in speech recognition output
- Lots of different classifiers have been used
  - Feature set; generally a superset of those used for unsupervised segmentation
  - + discourse markers and cue words, generally domain specific
- Model as a classification task
  - Classify if the sentence boundary is a paragraph boundary
- Use any classifier SVM, Naïve Bayes, Maximum Entropy etc.
- Or model as a sequence labeling task:
  - Label a sentence boundary with "paragraph boundary" or "not a paragraph boundary label"
- Features:
  - Use cohesion features: word overlap, word cosine similarity, anaphoras etc.
  - Additional features: Discourse markers or cue word
- Discourse marker or cue phrase/word: A word or phrase that signal discourse structure
  - For example, "good evening", "joining us now" in broadcast news
  - "Coming up next" at the end of a segment, "Company Incorporated" at the beginning of a segment etc.
  - Either hand-code or automatically determine by feature selection
- Evaluation
  - Not a good idea to measure precision, recall and F-measure because that won't be sensitive to near misses
  - One good metric WindowDiff (Pevzner & Hearst, 2002)

Discourse Coherence
What makes a Text Coherent?

What makes a Text Coherent?

- Appropriate use of coherence relations between subparts of the discourse, so-called rhetorical structure and the whole theory – Rhetorical Structure Theory (RST)
  - Rhetorical structure theory (RST) is a theory of text organization that describes relations that hold between parts of text.
  - It was proposed by Mann et. al. (1988) in their influential paper "Rhetorical structure theory: toward a functional theory of text organization" published in Interdisciplinary Journal for the Study of Discourse in 1988.
  - This theory is developed as part of studies of computer-based text generation.
  - In fact, NLP researchers later began using RST in text summarization and other applications.
- Appropriate sequencing of subparts of the discourse, so-called discourse/topic structure
  - Discourse topic structure is also the key to the cohesion of the discourse and reflects the essence of discourse analysis.
  - Over the past decades, discourse topic structure has been widely studied and proven to be a critical cohesive element at the text analysis.
  - A linear segmentation of texts into proper topic structures may reveal valuable information on, for instance, not only the themes of segments but also the overall thematic structure of the text, and it can subsequently be applied to various text analysis tasks, such as text summarization, information retrieval and discourse analysis.
- Appropriate use of referring expressions
  - A referring expression (RE), in linguistics, is any noun phrase, or surrogate for a noun phrase, whose function in a text (spoken, signed or written on a particular occasion) is "pick out" some individual person, place, object, or a set of persons, places, objects, etc.
  - The technical terminology for "pick out" differs a great deal from one school of linguistics to another.
  - It plays an vital part in discourse analysis, we will discuss it in detail in the coming section.

Discourse Coherence
Coherence Relations

What is Coherence Relations (CR)?

- Coherence refers to a property of discourse that makes each instance of discourse “make sense” (or with appropriate “sense relation) in context.
- For instance, given that a text or a paragraph of a text, or even just only two sentences or clauses have a common theme, there must be some common denominator that identifies a sequence of propositions as pertaining to the same theme.
- Or possible connections between utterances in a discourse.
- Such sense relation in discourse analysis is what Hobbs named “Coherence Relations” in his works “Coherence and Coreference” published by Cognitive Science in 1979, which is further developed by other linguistics including Sander et al. (1992) and Kehler (2002: 3) into a well defined theory.
- These sense relations, which are also called "propositional relations" by Mann & Thompson (1986), are encoded in the text and identified by the reader who tries to “make sense” of the text and its constituents.
- It can be considered as: “Types of reasons why speakers or writers have added this particular sentence" (Meyer et al. 2005).
- In fact, Coherence Relations are sometimes called 'types of thematic development’, just like the narrative in watching a movie or TV shows with cause-and-effect story-type development in sense relations.

References:

Hobbs, J. R. (1979) Coherence and Coreference. Cognitive Science 3, 67-90.

Sanders, Ted J. M. et al., (1992) Toward a Taxonomy of Coherence Relations. Discourse Processes 15: 1-35.

Kehler, Andrew (2002) Coherence, Reference, and the Theory of Grammar. Stanford, Calif.: CSLI Publishers.

Mann, William C. & Sandra A. Thompson (1986) Relational Propositions in Discourse. Discourse Processes 9: 57-90.

Discourse Coherence

Types of Coherence Relations

Types of Coherence Relations

1. Parallel: Infer $p(a_1, a_2, \ldots)$ from the assertion of S_0 and $p(b_1, b_2, \ldots)$ from the assertion of S_1, where a_i and b_i are similar, for all i.
   [7.30] The scarecrow wanted some brains. The Tin Woodman wanted a heart.
   [7.31] Rich man wants more power. Poor man wants more food.
   - Commonly used when we would like to describe two sense relations with similar situation (meaning), but different in object, reference and scenario.
2. Elaboration: Infer the same proposition P from the assertions of S_0 and S_1.
   [7.32] Dorothy was from Kansas. She lived in the midst of the great Kansas prairies.
   [7.33] Nicolas Telsa was a genius. He invented over hundreds of things in his life.
   - Commonly used in discourse construction, the successive sentences (utterances) is the further elaboration of the previous one.
3. Cause-and-Effect: S_0 and S_1 are in cause-and-effect if S_1 infers S_0, i.e. $S_1 \rightarrow S_0$
   [7.34] Jack cannot afford to buy the car. He lose his job.
   [7.35] Nicolas Telsa invented over hundreds of things in his life. He is genius.
   - Cause-and-effect Discourse Relation can be referred to animate or inanimate subjects. In [7.35], the cause-and-effect discourse relation is just the reverse of Elaboration statement [7.33], but not always happen.
4. Contrast: S_0 and S_1 are in contrast if P_0 and P_1 infer from S_0 and S_1 with one pair of elements that are contrast with each other, where other elements are similar in context.
   [7.36] Hope for the best. Prepare for the worst.
   [7.37] Jack is meticulous while Bob is sloppy.
   - Contrast CR can be within a sentence, or in successive sentences (utterances).
   - Most often refer to two subjects, two event with contract sense relations.
5. Occasion: A change of state can be inferred from the assertion of S_0, whose final state can be inferred from S_1, or a change of state can be inferred from the assertion of S_1, whose initial state can be inferred from S_0.
   [7.38] Dorothy picked up the oil-can. She oiled the Tin Woodman's joints.
   [7.39] Jack fails in the exam. He starts to work hard.
   - state changes invoke new action

Discourse Coherence

Hierarchical structures

- The coherence of entire discourse can also be considered by hierarchical structure between coherence relations.
- For example, the following passage can be represented as hierarchical structure :

[7.40] Jack went to the town to buy a toy.

[7.41] He took the bus to the shopping mall.

[7.42] He needed to buy a toy for his child.

[7.43] It is Jane's birthday.

[7.44] He also wanted buy some books for weekend reading.

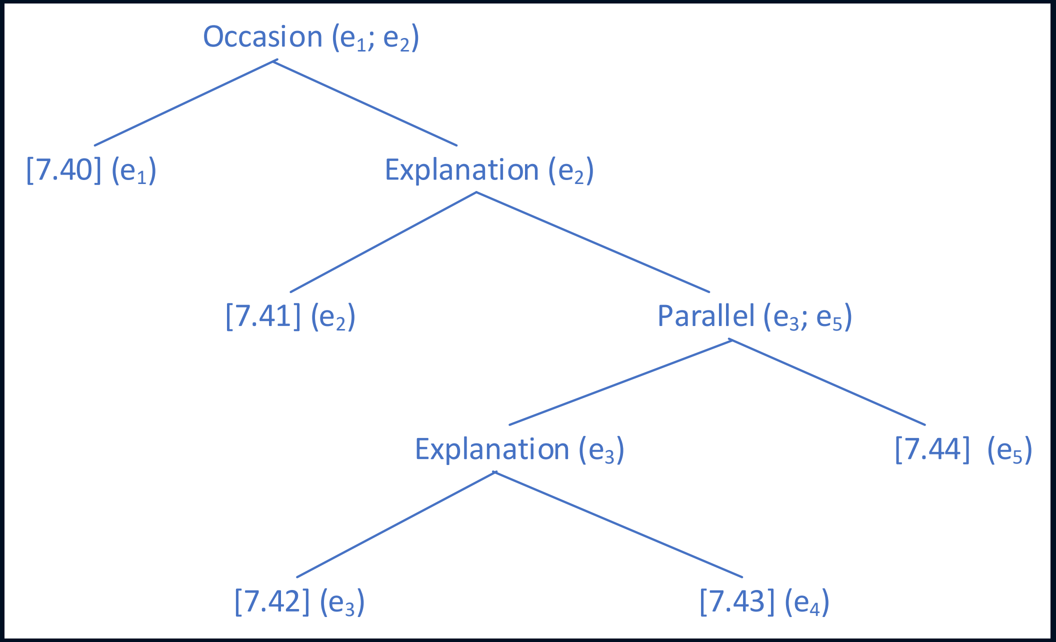

Fig. 7.4 Hierarchical structures in Discourse Coherence

Discourse Coherence
Types of Referring Expressions

- As mentioned, a referring expression (RE) is any noun phrase, or surrogate for a noun phrase, whose function in discourse is to identify some individual object.
- There are FIVE commonly used of Referring Expressions in Discourse Coherence:

1. Indefinite Noun Phrases: Introduces into discourse context entities that are new to the hearer.
   - Example: A man, some walnuts, this new computer

   [7.45] I go to the electronic store to buy a new notebook computer.
2. Definite Noun Phrases: refers to an entity that is identifiable to the hearer (e.g., been mentioned previously or well known, in set of beliefs about the world).
   - Example: "a big dog.... the dog...", the sun

   [7.46] Don't look at the sun directly in bare eyes, it will hurt yourself.
3. Pronouns: another form of definite reference, generally stronger constraints on use than standard definite reference.
   - Example: He, she, him, it, they...

   [74.47] I go to the electronic store to buy a new notebook computer. This computer is rather light and fast.
4. Demonstratives: demonstrative pronouns (e.g. this, that) can be alone or as determiners

   [7.48] That book seems to be very interesting and worth to buy.
5. Names: Common method of referring including people, organizations, and locations.

   [7.49] I go to KFC today to buy my lunch.

Discourse Coherence
Features for Filtering Potential Referents

- There are some common agreement and features for filtering potential referents in Discourse Coherence.
- Number Agreement: pronoun and referent must agree in number (single, plural)

  [7.50] The children are playing in the park. They look happy.
- Person Agreement: 1^{st}, 2^{nd}, 3^{rd}

  [7.51] Jane and Helen get up early this morning. They need to take morning exam.
- Gender Agreement: male, female, nonpersonal (it)

  [7.52] Jack looks tired. He hasn't sleep last night.
- Binding Theory Constraints: constraints by syntactic relationships between a referential expression and a possible antecedent noun phrase in the same sentence.
- Consider the following three sentences:

  [7.53] Jane bought herself a new iPad. ("herself" should be Jane)

  [7.54] Jane bought her a new iPad. ("her" should NOT be Jane)

  [7.55] She claimed that she bought Mary a new iPAD. ("She" and "she" should NOT be Mary)

Discourse Coherence

Preferences in Pronoun Interpretation

- Recency
  - Entities from most recent utterances more likely

[7.56] Tim goes to the clinic to see the doctor. He feels sick. It might be Covid-19.

- Grammatical Role
  - Salience hierarchy of entities that is ordered by the grammatical position of the expressions that denote them. [subject, object,...]

[7.57] Jane goes to Starbuck to meet Jackie. She orders a hot mocha. ("She" should be Jane)

[7.58] Jane discussed with Jackie about her exam result. She felt so nervous about it. ("She" should be "Jane" or "Jackie"?)

[7.59] Jane discussed with Jackie about her exam result. She felt so sorry about it. ("She" should be "Jane" or "Jackie"?)

Discourse Coherence

Preferences in Pronoun Interpretation

- Repeated Mention
  - keep talking about the same thing.

  [7.60] Jane went to superstore to buy some food. It turned out to be closed.
- Parallelism
  - subject to subject; object to object.

  [7.61] Mary went with Jane to the Starbuck. Jackie then went with her to the bookstore afterwards. ("her" should be "Jane" instead of "Mary")
- Verb Semantics
  - some verbs seem to place emphasis on one of their argument positions.

  [7.62] Jane warned Mary. She might fail the test.

  [7.63] Jane blamed Mary. She lost a lot in the investment.
- Selectional Restrictions
  - other semantic knowledge playing a role.

  [7.64] Mary lost her iPhone in the shopping mall after carrying it for the whole afternoon.

Algorithms for co-reference resolution
Introduction

- As mentioned, Coreference resolution (CR) is the task of finding all linguistic expressions (called mentions) in a given text that refer to the same real-world entity.
- After finding and grouping these mentions we can resolve them by replacing, as stated above, pronouns with noun phrases.
- In this chapter, we discussed the THREE fundamental algorithms for co-reference resolution:
  - Hobbs Algorithm
  - Centering Algorithm
  - Log-Linear Model (Learning Model)

Algorithms for co-reference resolution
Hobbs Algorithm

- Hobbs' algorithm (1978) was one of the earliest approaches to pronoun resolution.
- Hobbs (1978) presents two algorithms: a naive one based solely on syntax, and a more complex one that includes semantics in the resolution method.
- The naive one (henceforth, the Hobbs algorithm) is the one analyzed here.
- Unlike the other algorithms, the Hobbs algorithm does not appeal to any discourse models for resolution
- Rather, the parse tree and grammatical rules are the only information used in pronoun resolution.
- Although it is rather simple syntactical level and some sort of naïve-based pronoun resolution method, it works for many situations.
- Let's have a look on how it works.

Reference:
Hobbs, Jerry (1978) Resolving pronoun references. Lingua, 44:311–338.

Algorithms for co-reference resolution
Hobbs Algorithm

- The Hobbs algorithm assumes a parse tree in which each NP node has an N type node below it as the parent of the lexical object.
- The algorithm is as follows:

Hobbs Algorithm (Hobbs, 1978)

1. Begin at the noun phrase (NP) node immediately dominating the pronoun.
2. Go up the tree to the first NP or sentence (S) node encountered. Call this node X, and call the path used to reach it p.
3. Traverse all branches below node X to the left of path p in a left-to-right, breadth-first fashion. Propose as the antecedent any NP node that is encountered which has an NP or S node between it and X.
4. If node X is the highest S node in the sentence, traverse the surface parse trees of previous sentences in the text in order of recency, the most recent first; each tree is traversed in a left-to-right, breadth-first manner, and when an NP node is encountered, it is proposed as antecedent. If X is not the highest S node in the sentence, continue to step 5.
5. From node X, go up the tree to the first NP or S node encountered. Call this new node X, and call the path traversed to reach it p.
6. If X is an NP node and if the path p to X did not pass through the Nominal node that X immediately dominates, propose X as the antecedent.
7. Traverse all branches below node X to the left of path p in a left-to-right, breadth- first manner. Propose any NP node encountered as the antecedent.
8. If X is an S node, traverse all branches of node X to the right of path p in a left-to- right, breadth-first manner, but do not go below any NP or S node encountered. Propose any NP node encountered as the antecedent.
9. Go to Step4

Algorithms for co-reference resolution
Hobbs Algorithm

For illustration purpose, we are using the example from Hobbs (1978) to demonstrate how it works:
[7.65] The castle in Camelot remained the residence of the king until 536 when he moved it to London.

Fig. 7.5 shows the parse tree for the sentence [7.65].

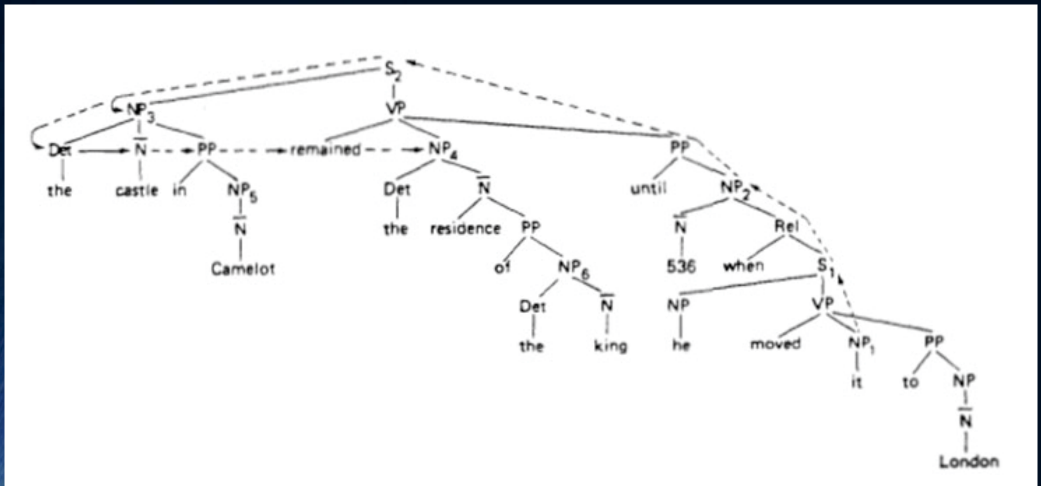

Fig. 7.5 Parse Tree for the sentence [7.65]

Ex1. What "it" stands for?

1. Beginning from node NP1, step 2 rises to node S1.
2. Step 3 searches the left portion of S1's tree but finds no-eligible NP node.
3. Step 4 does not apply.
4. Step 5 rises to NP2 which step 6 proposes as antecedent. Thus, '536' is recommended as antecedent of 'it'.
5. The algorithm can be improved somewhat by applying simple selectional constraints, such as
   Dates can't move;
   Places can't move;
   Large fixed objects can't move.
6. After NP2 is rejected, steps 7 and 8 turn up nothing, and control is returned to step 4 which does not apply.
7. Step 5 rises to S2, where step 6 does not apply.
8. In step 7, the breadth-first search first suggests NP3 (the castle), which selectional constraints reject.
9. It then continues to NP4 where it correctly settles upon 'the residence' as antecedent of 'it'.

Exercise: How to we check for the coreference resolution of "he" as "the king"?

Algorithms for co-reference resolution

Hobbs Algorithm

Performance of Hobbs Algorithm

- Hobbs analyzed, by hand, 100 consecutive examples from three "very different" texts.
- Assumed "the correct parse" was available.
- The algorithm was correct 72.7% (Hobbs, 1978).
- If the algorithm is integrated with syntactic constraints when resolving pronouns, such as the one we have illustrated in the example, the performance can be even higher.
- Hobbs algorithm suffers from TWO major problems:
  - When looking for a pronoun's antecedent within a sentence, it will go sequentially further and further up the tree to the left of the pronoun, and that failing will look in the previous sentence.
  - The algorithm does not assume a segmentation of discourse structure in this algorithm, it might go back arbitrarily far in the text to find an antecedent.
- Nevertheless, as Hobbs concluded in his original paper, Hobbs naïve-based approach on co-reference resolution do provides a high baseline, and it works in many usual situations in discourse analysis.

Algorithms for co-reference resolution

Centering Algorithm

- Centering Theory (CT) is proposed by Grosz and Sidner (1986) as a part of a larger theory of discourse analysis.
- As mentioned, Centering Theory is a theory of discourse structure that models the interrelationships between focus (so-called "centers") as the choice of referring expressions and perceived coherence of utterances.
- The basic idea is:
  - A discourse has a focus, or center.
  - The center typically remains the same for a few sentences, then shifts to a new object.
  - The center of a sentence is typically pronominalized.
  - Once a center is established, there is a strong tendency for subsequent pronouns to continue to refer to it.
- In Centering Algorithm, utterance within a discourse has one backward looking center (C_b) and a set of forward-looking centers (C_f).
- The C_f set for an utterance U_o is the set of discourse entities evoked by that utterance.
- The C_f set is ranked according to discourse salience; the most accepted ranking is by grammatical role.
- The highest-ranked element of this list is called the preferred center (C_p).
- The C_b represents the most highly ranked element of the previous utterance that is found in the current utterance.
- Essentially, it serves as a link between utterances.
- Abrupt changes in discourse topic are reflected by a change of Cb between utterances.
- Let's have a look on the Centering Algorithm.

Reference:
Grosz, Barbara J. and Candace L. Sidner. (1986) Attention, intentions, and the structure of discourse. Computational Linguistics, 12(3):175-204.

Algorithms for co-reference resolution
Centering Algorithm

Part I – Initial settings

- Let Un, Un+1 be 2 consecutive utterances.
- Backward looking center of Un, written Cb(Un), represents focus after Un interpreted.
- Forward looking centers of Un, written Cf(Un), forms ordered list of entities in Un that can serve as Cb(Un+1).
- Cb(Un+1) is highest ranking element of Cf(Un) mentioned in Un+1.
- Order of entities in Cf(Un): in which subject > existential predicate nominal > object > indirect object > demarcated adverbial PP
- Let Cp(Un+1) be highest ranked forward looking center

Part II – Constraints

For each utterance U_i (i = 1 … m) in a discourse segment D:

- There is precisely one C_b.
- Every element of the C_f-list for U_i must be realized in U_i.
- The center, $C_b(U_i, D)$, is the highest-ranked element of $C_f(U_{i-1}, D)$ that is realized in U_i.

Algorithms for co-reference resolution

Centering Algorithm

Part III – Rules and Algorithm

For each utterance U_i (i = 1 … m) in a discourse segment D:

Rule 1: If some element of $C_f(U_{i-1}, D)$ is realized as a pronoun in U_i, then so is $C_b(U_i, D)$.

Rule 2: Transition states (defined below) are ordered such that a sequence of Continues is preferred over a sequence of Retains, which are preferred over sequences of Smooth-Shift and then Rough-Shift.

The relationship between the C_b and C_p of two utterances determines the coherence between the utterances.

Centering theory ranks the coherence of adjacent utterances with transitions that are determined by the following conditions:

1. whether or not the C_b is the same from U_{n-1} to U_n;
2. whether or not this entity coincides with the C_p of U_n.

Fig. 7.6 shows the criteria for each transition in Centering Algorithm.

Based on these rules and conditions, the algorithm is defined as follows:

1. Generate all possible Cb - Cf combinations.
2. Filter combinations by constraints and centering rules.
3. Rank the remaining combinations by transitions.

	$C_b(U_{n+1}) = C_b(U_n)$ Or undefined $C_b(U_n)$	$C_b(U_{n+1}) \neq C_b(U_n)$
$C_b(U_{n+1}) = C_p(U_{n+1})$	**Continue**	**Smooth-Shift**
$C_b(U_{n+1}) \neq C_p(U_{n+1})$	**Retain**	**Rough-Shift**

Fig. 7.6 The criteria for each transition in Centering Algorithm

Algorithms for co-reference resolution
Centering Algorithm

Example: Use the following sentences as example to see how the Centering Algorithm works

[7.66] U1 : Jane heard some beautiful music at the CD store.

U2: She played it to Mary.

U3: She bought it.

By applying the grammatical role hierarchy to construct the Cf.

So, for U1 we have:

Cf(U1) : {Jane, music, CD store}

Cp(U1): Jane

Cb(U1): Undefined

As shown, U2 has two pronouns: She and it. She is compatible (in syntax) with Jane while it is compatible with either music and CD store.

As Jane is the highest ranked member of Cf(U1) , Cb(U2) should be referred to Jane.

By comparing the result transitions for every possible referent of it.

If we assume it refers to music, the result will be:

Cf(U2): {Jane, music, Mary}

Cp(U2): Jane

Cb(U2): Jane

Result: Continue (since Cp(U2) = Cb(U2) and Cb(U1) is undefined)

On the other hand, if we assume it refers to CD store, the result will be:

Cf(U2): {Jane, CD store, Mary}

Cp(U2): Jane

Cb(U2): Jane

Result: Continue (since Cp(U2) = Cb(U2) and Cb(U1) is undefined)

As they are both Continue – the same, we will set it refers to music instead of CD store.

Next, take a look on U3.

Algorithms for co-reference resolution

Centering Algorithm

Example: Use the following sentences as example to see how the Centering Algorithm works

[7.66] U1 : Jane heard some beautiful music at the CD store.

U2: She played it to Mary.

U3: She bought it.

For U3, She is compatible with either Jane or Mary, while it is compatible with music.

So, if we set she refers to Jane, i.e. Cb(U3) = Jane and the result will be:

Cf(U3): {Jane, music}

Cp(U3): Jane

Cb (U3): Jane

Result: Continue (since Cp(U3) = Cb(U3), and Cb(U3) = Cb(U2))

However, if we set she refers to Mary, i.e. Cb(U3) = Mary and the result will be:

Cf(U3): {Mary, music}

Cp(U3): Mary

Cb (U3): Mary

Result: Smooth-Shift (since Cp(U3) = Cb(U3) but Cb(U3) ≠ Cb(U2))

Since Continue is preferred to Smooth-Shift using Rule 2, Jane should be assigned as the referent.

The Centering Algorithm works in this situation.

Algorithms for co-reference resolution
Centering Algorithm

Performance of Centering Algorithm

- It is obvious to see that Centering Algorithm implicitly incorporates include the grammatical role, recency, and repeated mention preferences in pronoun interpretation.
- However, the manner in which the grammatical role hierarchy affects salience is indirect, since it is the resulting transition type that determines the final reference assignments.
- In particular, a referent in a low-ranked grammatical role will be preferred to one in a more highly ranked role if the former leads to a more highly ranked transition.
- In such situation, confusion will occur.
- For example:

[7.67] U1: Jane opened a new music store in the city.
U2: Mary enters the store and checks some music.
U3: She finally buy some.

In this case, by applying the Centering Algorithm, the algorithm will incorrectly assigns she to "Jane" because Cb(U2) = Jane so get continue, while "Mary" gets Smooth-Shift.

While by using Hobbs Algorithm, Mary will still be assigned as the referent.

One might ask: Is such case often happen?

It depends on situation and the thematic scenario.

Hobbs Algorithm vs Centering Algorithm:

- Marilyn A. Walker. 1989 manually compared a version of centering to Hobbs on 281 examples from three genres of text.
- Reported 81.8% for Hobbs and 77.6% for Centering Algorithm.

Reference:
Walker, Marilyn A. (1989). Evaluating discourse processing algorithms. In Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics, pages 251-261.

Algorithms for co-reference resolution

Log-Linear Model for Pronominal Anaphora Resolution

- Simple supervised machine learning approach
- Train classifier on a hand-labeled corpus in which are marked
  - Positive examples – antecedents marked with each pronoun
  - Negative examples (derived) – pairing pronouns with non-antecendent NPs
- Train on set of features
- Given a pro-antecedent pair predict 1 if they co-refer and 0 otherwise.

Algorithms for co-reference resolution
Features for Pronominal Anaphora Resolution

Commonly used features for anaphora resolution:

- Strict gender [true or false]
- Compatible number [true or false]
- Strict gender [true or false]
- Compatible gender [true or false]
- Sentence distance [0, 1, 2, 3,...] from pronoun
- Hobbs distance [0, 1, 2, 3,...] (noun groups)
- Grammatical role [subject, object, PP] – taken by potential antecedent
- Linguistic form [proper, definite, indefinite, pronoun] – of the pronoun

Algorithms for co-reference resolution

Example for Pronominal Anaphora Resolution

Using the following sentences (utterance) as example:

[7.68] U1: Jack saw beautiful Mercedes GLB300 at the used car dealership.

U2: He showed it to Jim.

U3: He bought it.

Fig. 7.7 shows the table of feature vector values for the sentence U2 [7.68]

Feature	He(U2)	it(U2)	Jim(U2)	Jack(U1)
Strict number	1	1	1	1
Compatible number	1	1	1	1
Strict gender	1	0	1	1
Compatible gender	1	0	1	1
Sentence distance	1	1	1	2
Hobbs distance	2	1	0	3
Grammatical role	subject	subject	PP	subject
Linguistic form	pronoun	pronoun	proper	proper

Fig. 7.7 The table of feature vector values for the sentence U2 [7.68]

Algorithms for co-reference resolution
Log-Linear Model for Pronominal Anaphora Resolution

Performance of Log-Linear Model

- Train on vectors.
  - Filter out pleonastic "it" as in "it is raining"
  - Results in weights for each of the features and combinations of features.
- Harder – must decide if any 2 noun phrases co-refer.

New Features:

- Anaphor edit distance
- Antecedent edit distance
- Alias [true or false] – use named entity tagger
- Appositive [true or false]
- Linguistic form [proper, definite, indefinite, pronoun]

Other Advanced Machine Learning Model

- Convolutional Neural Networks (CNN) (Auliarchman and Purwarianti, 2019)
- Recurrent Neural Networks (RNN) (Afsharizadeh et al., 2021)
- Long-Short Term Memory Networks (LSTM) (Li et al, 2021)
- BERT Model (Joshi et al., 2019)

In the following chapter, we will discuss the Transformer Technology and BERT Model in details.

Reference:

Afsharizadeh, M. et al. (2021). Automatic text summarization of COVID-19 research articles using recurrent neural networks and coreference resolution. Frontiers in Biomedical Technologies, 7(4)https://doi.org/10.18502/fbt.v7i4.5321

Auliarachman, T., & Purwarianti, A. (2019). Coreference resolution system for indonesian text with mention pair method and singleton exclusion using convolutional neural network. Paper presented at the 1-5. https://doi.org/10.1109/ICAICTA.2019.8904261

Joshi et al. (2019) BERT for Coreference Resolution: Baselines and Analysis. In Proc. of Empirical Methods in Natural Language Processing (EMNLP) 2019. https://doi.org/10.48550/arXiv.1908.09091

Li, Y. et al. (2021). Knowledge enhanced LSTM for coreference resolution on biomedical texts. Bioinformatics, 37(17), 2699-2705. https://doi.org/10.1093/bioinformatics/btab153

Evaluation

- Look at coreference chains as forming a set
- We represent the fact that A, B, and C corefer by having a class with A, B, and C in it.
- Reference Chain – True Chain – correct or true coreference chain an entity occurs in.
- Hypothesis chain – chain/class assigned to the entity by a coreference algorithm.
- Precision is evaluated according to:
  - weighted sum of correct # in hypothesis chain/# in hypothesis chain
- recall is evaluated according to:
  - # correct in hypothesis chain/# of elements in reference chain
- Intrinsic (using the prototype and model itself for evaluation) vs Extrinsic (task-based, end-to-end) Evaluation Scheme.

Summary

- Introduction to Pragmatics and Discourse
- Discourse Phenomena
  - Coreference Resolution
- Coherence and Coreference
  - Coherence vs coreference
  - Importance of CR
  - Entity-based coherence
- Discourse Segmentation
  - Unsupervised vs Supervised Discourse Segmentation
  - Hearst's TextTiling Method
- Discourse Coherence
  - Different Types of Coherence Relations
  - Different Types of Referring Expressions
- Algorithms for Coreference Resolution
  - Hobbs Algorithm
  - Centering Algorithm
  - Machine Learning -> Chapter 8 on Transfer Learning and Transformer Technology, include various technology such as RNN, LSTM and BERT model.
- Evaluation
- Summary

Next

Transfer Learning and Transformer Technology