Anaphora Resolution

Discussing 3 major papers in the Field

“There’s a pile of inflammable trash next to your car.

You have to get rid of it.”

Resolving Pronoun References

Jerry R. Hobbs
Dept. of CS, City College, CUNY, NY, USA

January, 1977

Idea

Traverse surface parse trees of sentences in a particular order looking for noun phrases of the correct gender and number.

Clearly, does NOT work in all cases.

Performs “remarkably well” - based on the results of “an examination of several hundred examples”

Simple, Naive & Efficient

Computational, Analytic point of view. Specified framework of analyzing texts.

Ideas floating around

Most work by linguists - Elucidate syntactic constraints on coreferentiality and non-corefernitality of 2 entities in SAME sentence

Winograd (1972) - procedures to locate Antecedents - collect all possible referents, rate plausibility based on syntactic knowledge

Rieger (1974) - find antecedent of a definite entity by creating, narrowing down a candidate set - based on ‘properties’ of antecedent

Charniak (1972) - Show how difficult it is - presented difficult cases in guise of children’s’ stories - NO solutions

Ideas floating around

Wilks (1974) - a nice partial solution

- Decide among competing plausible antecedents
- Use of selectional information to maximize redundancy
- Bidirectional search through a database of world knowledge

Surface Parse Tree

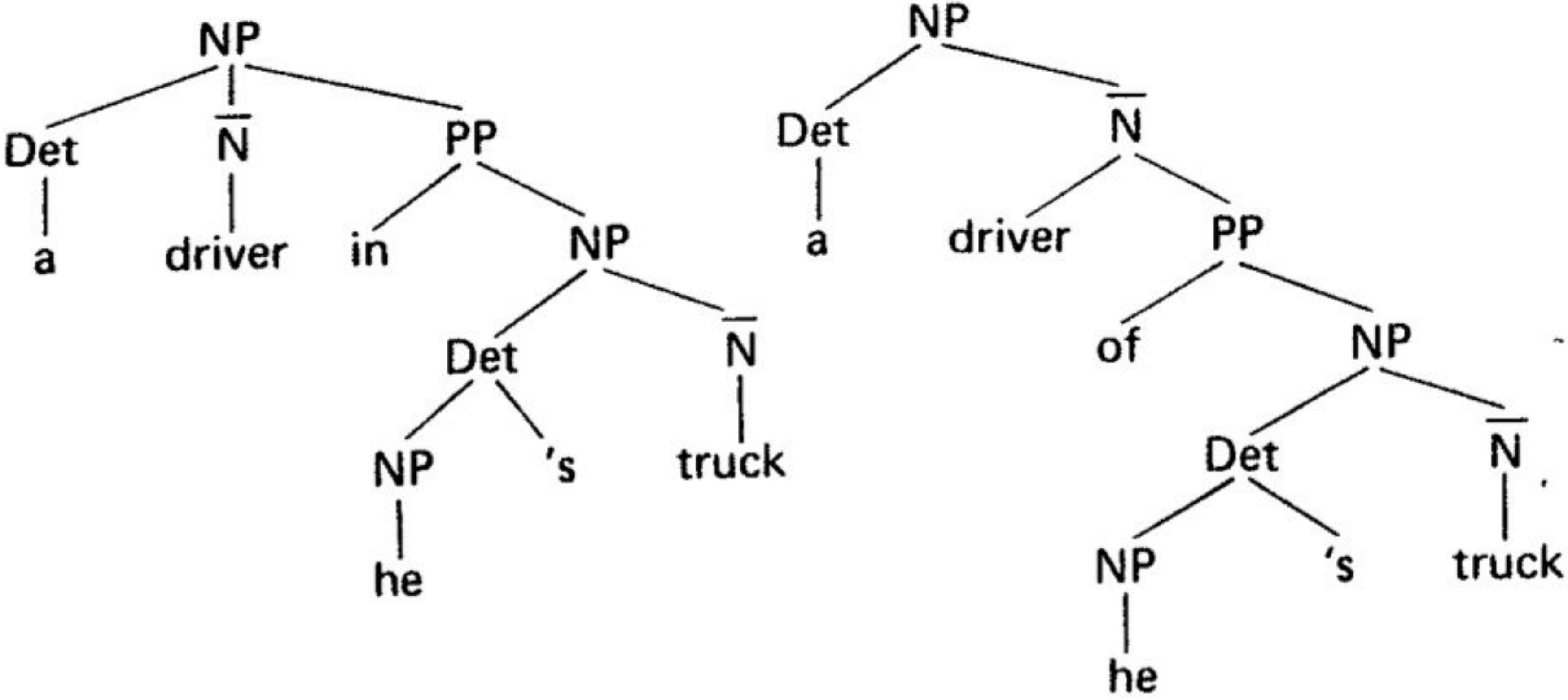
Exhibits Grammatical Structure - division into subject, verb, objects, adverbials

No permuting, omitting - terminal nodes from L to R → the sentence

NP must have N’ below [As proposed by Chomsky (1970)] to which a prepositional phrase containing argument of the head may be attached. Truly adjunctive prepositional phrases are attached to the NP node:

Necessary to distinguish between:

(a) Mr. Smith saw a driver in his truck
(b) Mr. Smith saw a driver of his truck


NP
Det
N̄
PP
a
driver
in
NP
Det
N̄
NP
's
truck
he
NP
Det
N̄
a
driver
PP
of
NP
Det
N̄
NP
's
truck
he

(a) (b)

CFG in use

S → NP VP
NP → (Det) N’ (PP)*
NP → (Det) N’ (Rel)*
NP → pronoun
Det → Article
Det → NP’s
N’ → noun (PP)*
PP → preposition NP
Rel→ *wh*-word S
VP → verb NP (PP)*

Algorithm in Detail

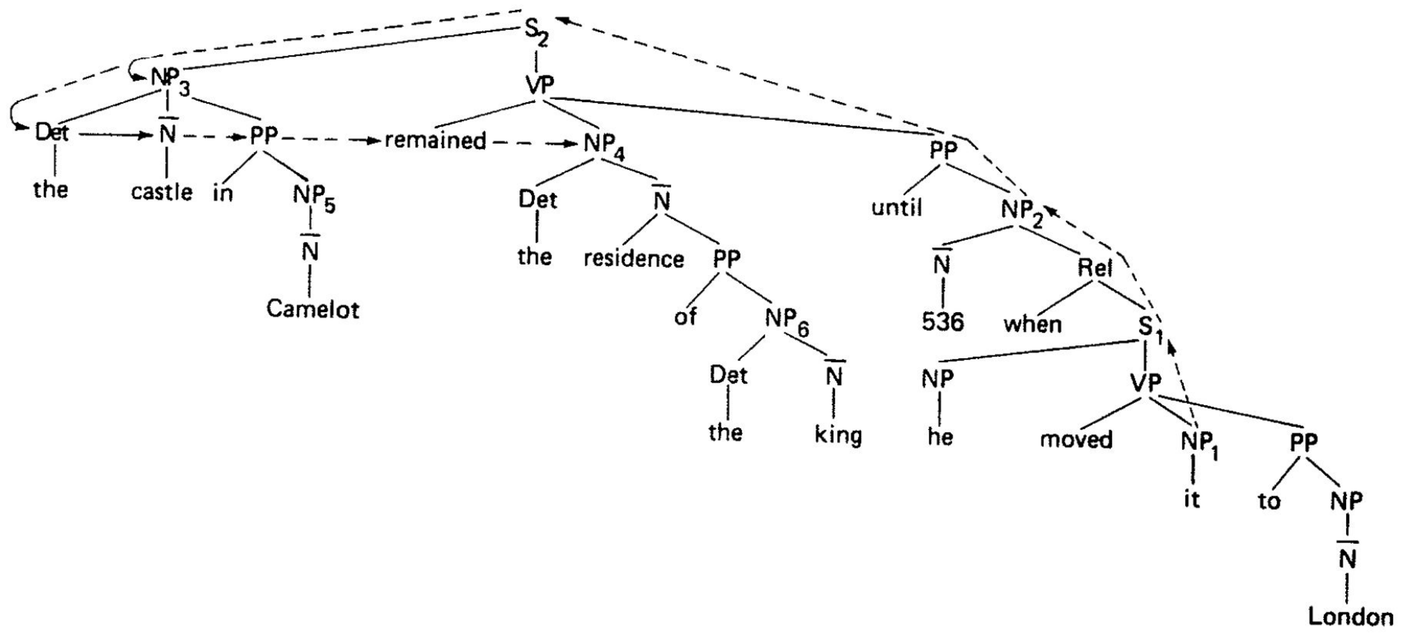
1. Begin at the NP node immediately dominating the pronoun.
2. Go up the tree to the first NP or S node encountered. Call this node X, and call the path used to reach it p.
3. Traverse all branches below node X to the left of path p in a left-to-right, breadth-first fashion. Propose as the antecedent 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 N 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 step 4.

See what Happens

“The castle in Camelot remained the residence of the king until 536 when he moved *it* to London.”

See what Happens

"The castle in Camelot remained the residence of the king until 536 when he moved *it* to London."



Vanilla Algorithm

Beginning from node NP1, step 2 rises to node S1. Step 3 searches the left portion of S1’s tree but finds no eligible NP node. Step 4 does not apply.

Step 5 rises to NP2 which step 6 proposes as antecedent. Thus, ‘536’ is recommended as antecedent of ‘it’.

Vanilla Algorithm

Beginning from node NP1, step 2 rises to node S1. Step 3 searches the left portion of S1’s tree but finds no eligible NP node. Step 4 does not apply.

Step 5 rises to NP2 which step 6 proposes as antecedent. Thus, ‘536’ is recommended as antecedent of ‘it’.

F

Simple Selectional Constraints*

Dates can’t move;
Places can’t move;
Large fixed objects can’t move.

After NP2 is rejected, steps 7 and 8 turn up nothing, and control is returned to step 4 which does not apply. Step 5 rises to S2, where step 6 does not apply. In step 7, the breadth-first search first suggests NP3 (the castle), which selectional constraints reject. It then continues to NP4, where it correctly settles upon ‘the residence’ as antecedent of ‘it’.

**The utility of these constraints is limited*

Generative Transformational Tradition

Conditions under which a noun or non-reflexive pronoun may not be coreferential with another element in the sentence. That is, the goal has been to state constraints of the form:

"A and B are necessarily non-coreferential if … "

Hence, take constraints into account

&

Have a mechanism which applies them.

Let's look at 2 constraints

Constraint 1:

Statement: A non-reflexive pronoun and its antecedent may not occur in the same simplex sentence (Lees and Klima 1963; Langacker 1969; Jackendoff 1972)

John likes him.
John’s portrait of him

‘him’ and ‘John’ cannot be coreferential. However, if an NP node precedes and is on a lower level than the pronoun, it is a possible antecedent, as in:

John’s father’s portrait of him.
After John robbed the bank, the police apprehended him.

This constraint is accommodated by steps 2 and 3 of the algorithm.

Pictures Pose Problems

John saw a picture of him.

Interpret ‘him’ as John

If they were coreferential, ‘himself’ would have been used.

Jackendoff has given an analysis of how reflexives are to be interpreted beyond the scope of simplex sentences.

The corresponding rule for how non-reflexives are not to be interpreted is incorrect. For there are cases where either the reflexive or non-reflexive pronoun may be used.

Pictures Still Pose Problems

John saw him.
John saw a picture of him.
John saw a picture of him hanging in the post office.
John saw that a picture of him was hanging in the post office.
John claimed that the picture of him hanging in the post office was a fraud.

‘Himself’ is perfectly acceptable in place of ‘him’ in all five sentences.

The more deeply the pronoun is embedded and the more elaborate the construction it occurs in, the more acceptable the non-reflexive becomes.

Yet there is no precise boundary between where it is acceptable and where it is not.

Rather than complicate the algorithm excessively

Let it FAIL on such cases.

Constraint 2:

Statement: Antecedent of a pronoun must precede or command the pronoun. A node NP1 is said to command node NP2 if neither NP1 nor NP2 dominates the other and if the S node which most immediately dominates NP1 dominates but does not immediately dominate NP2. (Langacker 1969)

Rationale: Take care of backward pronominalization*

After he robbed the bank, John left town.
That he was elected chairman surprised John.

Step 8 of the algorithm, which searches the tree to the right of the pronoun, handles such cases.

*Terms and Conditions Apply

Collateral Damage

Algorithm fails on several examples discussed in literature:

Mary sacked out in his apartment before Sam could kick her out.
(Lakoff 1968; Culicover 1976)

Girls who he has dated say that Sam is charming.
(Ross 1967)

This constraint never caused the algorithm to fail in the sample studied.
If it were lifted, the performance of the algorithm would degrade drastically.

Testing

One hundred consecutive examples of pronouns from each of three very different texts were examined to test the performance of the naive algorithm. The pronouns were ‘he’, ‘she’, ‘it’, and ‘they’.

‘It’ was not counted when referring to a syntactically recoverable ‘that’ clause or occurring in a time or weather construction. In applying the algorithm, it was assumed that the correct parse was available for each sentence.

The texts were:

1. William Watson’s Early Civilization in China, pp. 21-69
2. The first chapter of Arthur Haley’s novel Wheels, pp. 1-6
3. The July 7, 1975 edition of Newsweek,pp.13-19,beginning with the article ‘A Ford in High Gear’.

Results

Accuracy:

88.3% (without selectional constraints)
91.7% (with selectional constraints)

#	Conflicts before selection	Algorithm Correct	Conflicts after selection	Algorithm Correct
he	31	22	31	22
she	0	0	0	0
it	48	33	44	33
they	53	43	45	41
TOTAL	132	98	120	96

Propose a Semantic Approach

- understanding natural language requires a great deal of world knowledge
- assume this knowledge is available in the form of predicate calculus axioms
- several ‘semantic operations’ draw inferences selectively from the collection of axioms
- entities standing for anaphors are merged with the entities standing for their antecedents

Illustration:

The boy is on the roof of the building.

on(X1 | boy(X1), X2 | roof(X2, X3 | building(X3)))

Anaphora Resolution:
A Multi Strategy Approach

Jaime G. Carbonell
Ralf D. Brown
CS Dept & Center for Machine Translation, CMU

15 April, 1988

Idea

Anaphora Resolution requires integrated application of syntactic, semantic, and pragmatic knowledge.

Instead of attempting to construct a monolithic method for resolving anaphora, the combination of multiple strategies, each exploiting a different knowledge source, proves more effective, theoretically and computationally.

Syntactic information plays central role in establishing appropriate referents for intrasentential anaphora

Semantic and pragmatic information is required intersentential anaphora

Addresses problem of intersentential anaphora resolution, integrating caseframe semantics and more global dialog coherence structures

Semantics and Pragmatics Dominate

John took the cake from the table and ate it.
John took the cake from the table and washed it.
Tile robot pushed the box towards the conveyor belt. But, it goofed and dropped it on its way there.

Semantic preference constraints suffice to resolve the first example. The preferred object of ingestion is an edible substance.

Little more difficult to mechanize a process that excludes things such as cakes from being the object of washing.

The robot example requires sophisticated semantics.

The Various Strategies

Local Anaphor Constraints

Certain anaphors carry with them constraints (number, gender, case, etc.) which must be satisfied by the candidate referents.

John and Mary went shopping. He bought a steak. [he=John]

Eliminate from consideration all candidate referents that violate the local constraints of the anaphor its question.

Case-Role Semantic Constraints

Semantic constraints on the object case of "to eat" and "to wash" impose restrictions on the possible set of referents.

John took the cake from the table and ate it. [it = cake]
John took the cake from the table and washed it. [it = table]

Eliminate from consideration all candidate referents that violate any case-constraint imposed on the anaphor its question. Prefer those candidates that accord with typical case fillers, in the absence of hard constraints.

Precondition/Postcondition Constraints

John gave Tom an apple. He ate the apple. [he=Tom]

Here, "he" refers to Tom, as John no longer has the apple. The postcondition on give is that the actor no longer have the object being given, which conflicts with the precondition on eat that the actor have the item being eaten, if the actor is assumed to be John.

Eliminate from consideration all candidate referents associated with actions whose postconditions violate the preconditions of the action containing the anaphor.

Case-Role Persistence Preference

Mary gave an apple to Susan. John also gave her an orange. [her=Susan]
Mary gave an apple to Susan. She also gave John an orange. [she=Mary]

Case of structural parallelism. And, the semantic structure dominates over the syntactic one. For instance, in the first example, "Susan" is the object of the "to" prepositional phrase, whereas the corelerent anaphor is in the indirect object position: two different syntactic roles that map into the same semantic case, recipient. In the second example above, both syntactic and semantic structures coincide, and therefore the preference is stronger.

Search first for acceptable referents in the antecedent phrase (or phrases) that occur in the same semantic case role as the anaphor, if a match satisfying all constraints is found, look no further; else search the other case roles.

Semantic Alignment Preference

Mary drove from the park to the club. Peter went there too. [there=club]
Mary drove from the park to the club. Peter left there too. [there=park]

Sentences share the identical syntactic structure and the same basic underlying semantic case structure. However, discourse cohesion prefers to make the sentences coreferential (pragmatically parallel) with respect to the same underlying action (leaving the park and going to the club).

If the clause in which the anaphor is embedded aligns with a previous clause ("aligns" means that it can represent the same underlying action, perhaps with different instantiated case fillers), or with part of a previous clause, search first for referents of the anaphor in that clause. If there are no allowable referents in the semantically aligned clause, expand the search to other antecedent clauses; else halt the search.

Syntactic Parallelism Preference

The girl scout leader paired Mary with Susan, but she had paired her with Nancy last time. [she=leader, her=Mary]
The girl scout leader paired Mary with Susan, but she had paired Nancy with her last time. [she=leader, her=Susan]

There is no reason to prefer different referents for the pronoun "her" in each sentence above, other than retaining as much as possible the surface syntactic order from the first coordinate clause in the second clause.

In coordinated clauses, adjacent sentences or explicitly contrasted sentences, prefer the anaphoric referent that preserves the surface syntactic role from the first clause.

Syntactic Topicalization Preference

It was Mary who told Jane to go to New York. Why did she do it? [she=Mary]
It was Jane who went to New York at Mary's bidding. Why did she do it? [she=Jane]
It was Mary who told Peter to go to New York. Why did he do it? [he=Peter]
It was Peter who went to New York at Mary's bidding. Why did he do it? [he=Peter]

Exact same semantic and syntactic structures yield "Peter" both times as the referent of "he", because localized constraints so dictate, regardless of who is topicalized.

Search first a syntactically topicalized part of the candidate antecedent clause (or clauses) for the referent of the anaphor. If an acceptable referent is found, search no further; else search the rest of the clause(s).

Intersentential Recency Preference

At the paragraph (or dialog) level level, we advocate searching sentences in reverse chronological order, applying all the constraints and preferences to select among possible candidates within each sentence. If there are no satisfactory candidates in the previous sentence, then the one before that is considered, and so on.

Integrating Strategies

Principles

One needs to make a distinction between constraints (which cannot be violated), and preferences (which discriminate among candidates satisfying all constraints)

Apply the constraints first to reduce the number of candidate referents for the anaphor in question

If more than one preference applies, and each suggests different candidate referents for the anaphor in question all of which have passed the constraint tests, then we consider the anaphor to have a truly ambiguous referent.

Results

70 examples yielded 49 unique resolutions, 17 conflicting possibilities, and 4 anomalous cases.

Human judgements correlate very well in terms of identifying the same referent as that suggested by the system in the 49 unique cases. The majority of the 17 multiple-referent cases were judged ambiguous by the subjects (the rest required complex world knowledge to establish a unique referent).

Achieve “human-like performance” with the multi-strategy method of determining referents to anaphors using different sources of linguistic knowledge in a semi-modular fashion.

Implementation

In context of the Universal Parser (UP) project at Center for Machine Translation at CMU.

Modified form of lexical-functional grammar unifying syntactic and semantic knowledge sources to produce a complete parse of each sentence.

Operates post facto on the set of instantiated semantic case frames and syntactic trees, attempting to resolve anaphors in the parse of the newest sentence using earlier parses (semantic and syntactic) as context to mine for candidate referents

Bridge

A Statistical Approach to Anaphora Resolution

Niyu Ge, John Hale, Eugene Charniak

ACL 1998

A Mention-Ranking Model for Abstract Anaphora Resolution

Ana Marasović, Leo Born, Juri Opitz, and Anette Frank
Research Training Group AIPHES
Department of CL, Heidelberg University
21 July, 2017

Abstract Anaphora

Ever-more powerful desktop computers, designed with one or more microprocessors as their ‘brains’, are expected to increasingly take on functions carried out by more expensive minicomputers and mainframes. [Antecedent: “The guys that make traditional hardware are really being obsoleted by microprocessor based machines”], said Mr. Benton. [AnaphS: As a result of (this trend) AA, longtime powerhouses HP, IBM and Digital Equipment Corp. are scrambling to counterattack with microprocessor-based systems of their own.]

Establishes a relation between the anaphor embedded in the anaphoric sentence and its (typically non-nominal) antecedent

Task: Resolve a wide range of abstract anaphors, such as the NP ‘this trend’ above, as well as pronominal anaphors (this, that, or it)

Ideas

Exploit recent advancements in representation learning

Mention-ranking model that learns how abstract anaphors relate to their antecedents with an LSTM-Siamese Net

Overcome the lack of training data by generating artificial anaphoric sentence antecedent pairs

Difficult

In contrast to nominal anaphora, abstract anaphora is difficult to resolve, given that agreement and lexical match features are not applicable

Annotation of abstract anaphora is also difficult for humans

Few smaller-scale corpora have been constructed

1. ARRAU corpus (Uryupina et al., 2016) that contains abstract anaphors
2. Shell noun corpus used in Kolhatkar et al. (2013b).

Shell Nouns

Abstract nouns, such as fact, possibility, or issue, which can only be interpreted jointly with their shell content. Can be anaphoric (ASN) or cataphoric (CSN).

Embedded clause: Congress has focused almost solely on the fact that [special education is expensive - and that it takes away money from regular education.]

Antecedent: Environmental Defense notes that [Antec: Mowing the lawn with a gas mower produces as much pollution as driving a car 172 miles.] [AnaphS: This fact may [...] explain the recent surge in the sales of old-fashioned push mowers].

Existing Approaches

Kolhatkar et al. (2013b):

- Resolve six typical shell nouns following the observation that CSNs are easy to resolve based on their syntactic structure alone
- Assumed that ASNs share linguistic properties with their embedded (CSN) counterparts.
- Manually developed rules to identify the embedded clause (i.e. cataphoric antecedent) of CSNs and trained SVM rank on such instances.
- The trained SVM rank model is then used to resolve ASNs

Existing Approaches

Kolhatkar and Hirst (2014):

- Generalized method to be able to create training data for any given shell noun
- Heavily exploited specific properties of shell nouns
- Does not apply to other types of abstract anaphora

Clark and Manning (2015), Clark and Manning (2016a):

- Mention-ranking neural coreference model
- Integrates a loss function which learns distinct feature representations for anaphoricity detection and antecedent ranking.

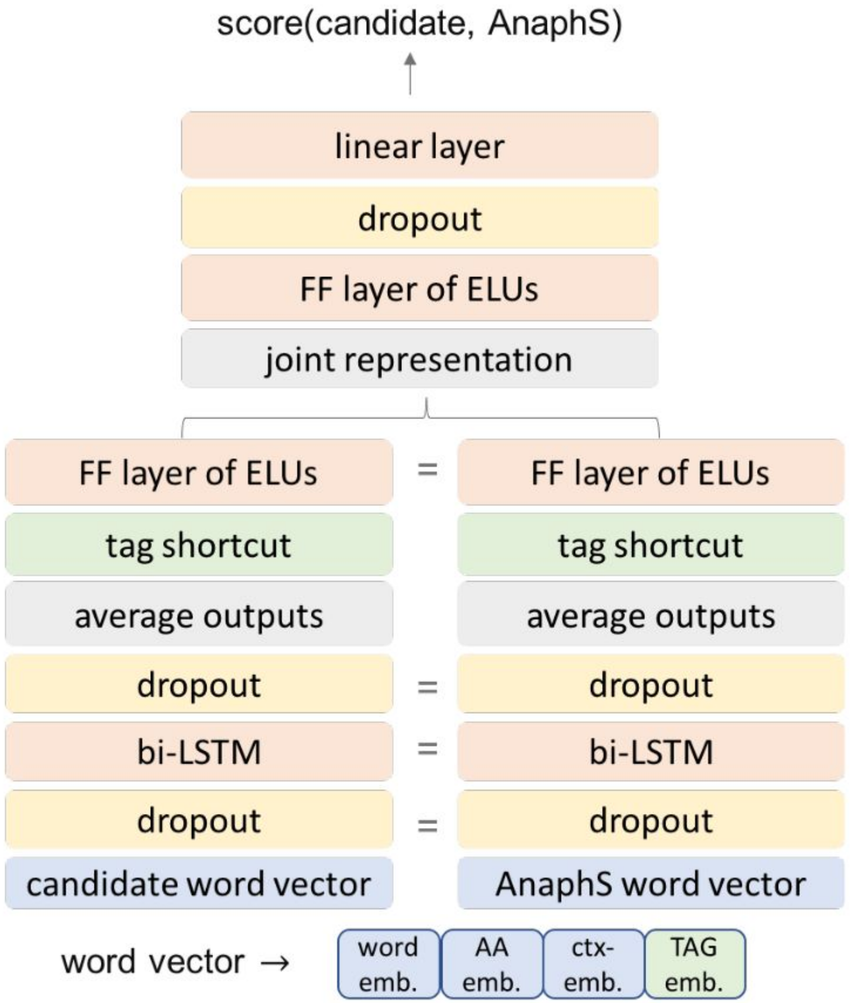
Solution

LSTM-Siamese Net learns representations for the candidate and the anaphoric sentence in a shared space

Representations combined into a joint representation used to calculate a score that characterizes the relation between them

Learned score used to select the highest-scoring antecedent candidate for the given anaphoric sentence and hence its anaphor

One anaphor at a time - provide the embedding of the context of the anaphor and the embedding of the head of the anaphoric phrase to the input to characterize each individual anaphor



Learn representations of an anaphoric sentence ‘s’ and a candidate antecedent ‘c’ using a bidirectional Long Short-Term Memory.

One bi-LSTM is applied to the anaphoric sentence s and a candidate antecedent c, hence the term siamese.

Compute the score for the pair (c, s) that represents relatedness between them, by applying a single fully connected linear layer to the joint representation.

Results

Report success@n (s@n)

Measures whether the antecedent, or a candidate that differs in one word, is in the first 'n' ranked candidates

n ∈ {1, 2, 3, 4}

In terms of s@1 score, outperforms previous state-of-the-art without even necessitating HP tuning*

		s @ 1	s @ 2	s @ 3	s @ 4
fact (train: 43809, test: 472)	MR-LSTM	**83.47**	**85.38**	**86.44**	**87.08**
	KZH13	70.00	86.00	92.00	95.00
	TAG_{BL}	46.99	-	-	-
reason (train: 4529, test: 442)	MR-LSTM	71.27	77.38	80.09	80.54
	+ tuning	**87.78**	**91.63**	**93.44**	**93.89**
	KZH13	72.00	86.90	90.00	94.00
	TAG_{BL}	42.40	-	-	-
issue (train: 2664, test: 303)	MR-LSTM	**88.12**	**91.09**	**93.07**	**93.40**
	KZH13	47.00	61.00	72.00	81.00
	TAG_{BL}	44.92	-	-	-
decision (train: 42289, test: 389)	MR-LSTM	**76.09**	**85.86**	**91.00**	**93.06**
	KZH13	35.00	53.00	67.00	76.00
	TAG_{BL}	45.55	-	-	-
question (train: 9327, test: 440)	MR-LSTM	**89.77**	**94.09**	**95.00**	**95.68**
	KZH13	70.00	83.00	88.00	91.00
	TAG_{BL}	42.02	-	-	-
possibility (train: 11874, test: 277)	MR-LSTM	**93.14**	**94.58**	**95.31**	**95.67**
	KZH13	56.00	76.00	87.00	92.00
	TAG_{BL}	48.66	-	-	-

**in most cases*

Observation

Omitting syntactic information boosts performance in ARRAU

When the model is provided with syntactic information

- It learns to pick S-type candidates
- But does not continue to learn deeper features to further distinguish them

Not able to point to exactly one antecedent, resulting in a lower s@1 score, but does well in picking a few good candidates, which yields good s@2-4 scores.

Thank You.