Generation of Anaphoric Noun Phrases and their coreferential properties

Parma Nand
School of Computing and
Mathematical Sciences
Auckland University of Technology
Auckland, New Zealand
Email: pnand@aut.ac.nz

Wai Yeap
School of Computing and
Mathematical Sciences
Auckland University of Technology
Auckland, New Zealand
Email: wyeap@aut.ac.nz

***Abstract*—This paper investigates the generative processes involved in the formation of noun phrases (NPs) from an existing clause in a discourse. A speaker or writer is able to generate a range of NPs from a stated clause to co-refer to the different semantic units of the clause by combining the various primitive components. These generative processes can be grouped into either predicate-deletion or normalization. We use relations defined by Levi [1] as a basis to define constraints used in the formation of NPs by predicate-deletion and empirically determine the conformance level of these constraints with natural discourses. For the normalization process, we argue that the inflicted morpheme's anaphoric properties can be characterized by their suffixes. We also show that, these anaphoric properties are independent of the choice between subjective and objective modifiers. The anaphoric properties of suffix-based morphemes is also empirically investigated for conformance levels, and the results are encouraging.**

I. Introduction

An integral task of any nlp system is the resolution of anaphora. An Anaphor is an noun phrase[1](NP) that is used to co-refer to a previously mentioned entity in a discourse and anaphora resolution is the process of identifying the entity an NP co-refers to (see [2] and [3] for a discussion on types of anaphora). An entity can be co-referred to by either a pronoun or an NP. Pronouns are predominantly used to refer to a **recently mentioned** entity (see [4], [5] and [**?**] for an overview), and since there are only a limited number (34) of them, they have to be reused for different entities through the course of a discourse. NPs on the other hand, can be generated by permutating various *properties* and *relations* of an entity which enables one to have enough unique "variables" available to co-refer to the required entities in a discourse. This permutation can lead to generation of a large number of NPs, some of which may also involve covert semantic based relations bridging between the generated NP and the entity it refers to. The coreferencial properties of an NP anaphor is not as focussed as pronominal anaphora. Pronominal anaphora is almost always[2] co-refers to an entity, or a more then one entity in the case of plural pronouns (they, them and their). NP anaphora on the other hand, can also refer to non-entity components of an event, in addition to potentially co-referring to one or more entities. A large proportion of NP anaphora does not co-refer to a specific entity, but to a semantic unit, which can is formed based on relations and properties of the participating entities. Semantic units can be formed by entities sharing a subset of their properties/relations such as a car and a truck, two entities participating in a similar action such as two people involved in an accident or a fork and a spade which have a similar function. Similarly, entities engaged in an action can form semantic units that correspond to neither the action nor the entities participating in the action, eg. "industry produces milk" forms the NP industrial production which co-refers to the act of producing milk. To be able to interpret these anaphoric NPs an nlp system needs to be able to construct these semantic units so that NPs can be correctly interpreted to resolve to these units. Although NPs can co-refer to units formed as a result of several common relations and properties between discourse entities, in this study we examine anaphoric NPs generated as a result of *predicate-deletion* and *normalization* from a single clause. The study involved content analysis of over 3000 sentences from New Zealand online newspaper articles, by isolating clauses specifying the relation as well as the clauses containing the inflicted anaphoric NPs co-referring to some part of the clause specifying the relation.

To be able to identify the various components of a clause an anaphoric Np can co-refer to, we need to define a clause in terms of these components. Hence, we will define a clause to consist of a *subject* and a *predicate*. The subject is the argument that refers to the origin or the undergoer of the state as determined by the predicate. The subject usually manifests itself as the actor or agent of the predicate, however this might not be always true, as in the case of unaccusative verbs such as arrive, die, fall etc. The predicate consists of the main verb of the clause as well as one or more objects, if it/they exist. We will define a clause to consist of the semantic units that correspond to each of the participating *entities*, the *clause* as a whole, *predicate* and the *action*.

Section II examines the formation of NPs that result from deletion of predicates and their interpretation in terms of semantic units that they can co-refer to. Section III discusses

[1]An NP can be defined to have the form <determiner/s> <modifier/s> <head noun>

[2]except sometimes the pronoun *it* which can be coreferential to components other then an entity

the NPs that can be generated by normalization of the verbs and the effect of modifiers originating from the subject or the object of the clause. The conclusion states some of the additional observations resulting out of the content analysis done, together with the main conclusion from this research.

II. NPs formed from predicate-deletion

Predicate-deletion is the process of combining the subject and the object participating in an event into an NP, where the predicate of the clause is implicit in the NP by virtue of the *relation* between the two combined entities forming the NP. In the process, the predicate gets deleted hence the name, predicate-deletion. The process of combining entities to form an NP may involve either juxtaposing the two entities, as in "milk industry" or concatenating them as in "girlfriend". Let us consider the clause in (1). The NPs in sentences (1a) and (1b) are formed from predicate-deletion discussed in this section while the ones in (1c) and (1d) are formed from normalization which is discussed in section III.

(1) *The industry processes milk.*

a) The industry/industrial milk...
b) The milk industry...
c) The industrial processor...
d) The milk processor...

NPs are generated by combining nouns or their adjective morphemes resulting in coreferential properties determined by their source as well as combined use of other syntactical components. The generative process of constructing NPs from a clause is constrained, in terms of, the order and the primitive nouns that can be combined. We cannot randomly juxtapose or concatenate nouns to form new NP's. From a phrase such as "The battle caused fatigue" we can generate "battle fatigue" but not "fatigue battle". Linguists working in this area have come up with theories to explain the NP generative process. These are based on factors such as entity properties, verb classification and entity taxonomies. An influential theoretical framework that uses verb classification together with entity properties was presented by Levi [1] which is based on similar framework as Lees [6]. Levi outlined an initial list of seven categories of relations which is used in the process of NP generation from clauses. Whether this list is complete or not, is contestable, and has been argued by researchers such Downing [7]. In spite of the contests, the researchers do concur that the framework is able to explain the majority of generated NPs. Levi's list of these semantic relations consists of CAUSE, HAVE, MAKE, USE, BE, IN, and FOR, which account for the example NPs shown in table I adapted from Downing [7]. Table I constrains us to only form NPs by way of concatenation and juxtaposing nouns **only** if they fall into any one of these seven categories. For example, in the case of the first deletable predicate, CAUSE, a clause such as "The battle gave us fatigue" can generate the NP "battle fatigue". In this clause the the predicate "gave us fatigue" is classified as CAUSE hence the the predicate can be deleted and replaced with a compound NP consisting of the subject and object of the clause. The entity "battle" causes fatigue enables us to generate "battle fatigue" but not "fatigue battle", however this is not valid in the case of "The industry produces milk" which generates "milk industry" as well as "industry milk". Since we are able to generate a NP using both the subject and the object as the head-noun, it makes the CAUSE deletable bi-directional. In addition, Levi classifies HAVE and MAKE as bi-directional as well, with the rest (USE, BE, IN/ON, FOR) being unidirectional. For unidirectional deletables, the **object** survives as the head-noun and the subject or one of its morphemes becomes the modifier.

Predicate	N + N	ADJ + N
CAUSE	battle fatigue	viral infection
	disease germ	malarial mosquito
HAVE	snake poison	reptilian scales
	apple cake	musical comedy
MAKE	cellblock	floral wreath
	silk worm	sebaceous glands
USE	steam engine	electric drill
BE	girlfriend	professorial friends
IN/ON	fieldmouse	polar bear
	hay truck	sugary cake
FOR	bird sanctuary	avian sanctuary

TABLE I
TABLE SHOWING THE LIST OF SEVEN DELETABLE PREDICATES AS DEFINED BY LEVI ([1]). WITH A FEW ADDITIONS

The objective of this study was to determine the level of conformance of NP formation from predicate-deletion to the seven relations. This study forms part of a larger nlp project which deals with resolution of NP anaphora hence before developing rules based on the seven relations we had to test the conformance of the theory on natural discourses.

A. Method and results for predicate-deletion

We used a locally developed text processor to search for NPs consisting of constituents of previous nouns from 120 articles (of mixed genre) from New Zealand Herald, The Dominion Post and The Press, which are three major online newspapers in three different cities in New Zealand. The choice of articles were not completely random. The corpora was developed to serve as the input data for anaphora resolution system which is the parent project of this study. Hence, the corpora was developed from the articles which had more then 20 sentences, exhibited use of variety of anaphoric uses and had been written by different writers. This resulted in 246 predicate-deleted NPs which were further filtered for NPs that got formed only from nouns or their morphemes that participate in the **same event**. This reduced the list to 155 predicate-deleted test cases from more than 3000 sentences. These test cases were then manually classified into either one of the seven relation in table I or into an unclassified category. Table II shows the breakdown of the results.

The classifications were done manually by a single person in order to avoid any differences in interpretation. Some of the above relation categories can be based entirely on verbs while others depend on the properties of participating nouns. For

Relation	No.	Percent
CAUSE	41	26.5
HAVE	15	9.7
MAKE	13	8.4
USE	10	6.5
BE	28	18.1
IN/ON	13	8.4
FOR	17	11.0
Unclassified	18	11.6
Totals	**155**	**100**

TABLE II
BREAKDOWN OF RESULTS FOR CONFORMANCE OF DELETABLE PREDICATES TO SEVEN RELATION CATEGORIES.

example verbs such as cause, generate, lead to, and bring about are causative irrespective of the participating nouns while other verbs such as kill, melt, provoke, and poison are ambiguous, their causativeness being dependent on the participating nouns. For example the clause "the heat melted the ice" was classified as causative while "He melted the ice" was classified as non-causative. Table II shows that the greatest proportion (26%) of NPs formed by predicate-deletion were from the CAUSE relation followed by the BE relation (18.1%). The BE relation is also referred to as "is-a" relation by some researchers and includes sentences which are copula such as "John has been a teacher," and "Car is a vehicle." More importantly, it was established that only 11.6% of NPs that were formed as a result of predicate- deletion were outside the constraints of the seven relations defined in table I, giving us a conformance level of 88%. For a theory describing an aspect of a natural language a conformance level of approximately 88% can be taken to be reasonably acceptable.

Although the results of manual classification is reasonably convincing, it maybe sometimes difficult to computationally determine relation as not all relations are overt in the syntactic or surface structures. Currently, studies using WordNet 1.7 [8] is being used to automatically determine the relations is being conducted, and will be reported in forth coming papers. In addition, rules based on the NP formation using predicate-deletion based on the seven relations is being studied as part of a wider anaphora resolution system. The results of this integration is also encouraging and will be reported in forth-coming publications. The next section examines generation of anaphoric NPs from *normalization*.

III. NPs FORMED FROM NORMALIZATION

Normalization is the process of changing a verb or an adverb into either a the head-noun or a modifier of the head-noun in a subsequent NP. Some examples of normalizations for each of the categories are :
verb to head-noun – drive normalized to driver.
adverb to head-noun – accidently drove a car into... normalized to accident.
adverb to modifier – accidently drove a car into... normalized to accidental act.
verb to modifier – communicate normalized to communicative act...

The noun resulting from the normalization process is referred to as the an *inflicted form* or a *morpheme* of the originating adjective/verb. Out of the two forms of normalizations, verb-normalization is predominant and hence has received most attention from researchers in linguistics (eg. Downing [7], Abbott[9] and Heim [10]) and computational linguistics (eg. Vieira et al. [11], Ng et al. [12] and Fraurud [13]). In this section we will examine verb as well as the adjective inflictions from the perspective or their effect on *anaphoric* properties.

Most computational linguists dealing with verb-normalization implement verb inflictions as in "John drives a car", hence John can be co-referred to as "the car driver". Linguists such as Levi ([1]) studying normalization from the point of view of NP generation propose two forms; one in which a morpheme of the subject survives as the modifier, referred to as subjective infliction, and the other in which a morpheme of the the object survives as the modifier, referred to as the objective infliction. For example, the phrase "John massaged the heart", can have an objective infliction as "The heart massage" where the object survives as the modifier of the NP. We will consider both forms of verb normalization from the point of view of their co-referential properties.

There can be more then one NP generated by the process of normalization and these generated NPs have varying coreferential properties. In addition, we can add either subjective or objective modifiers for NPs in each of the NP generated by the normalization process. To illustrate this lets consider the phrase in (2). The normalized NPs are "production" and "product" which belong to two *forms of morpheme* distinguished by the suffix "ion" and "uct". The two sentences in (2a) and (2b) are two *morpheme types* characterized by use of subjective or objective modifier. Our objective is to ascertain the difference in the coreferential properties between different **forms** of morphemes ((2a) and (2b) compared to (2c) and (2d) and different **types** of morphemes (2a) compared to (2b).

(2) *The industry produces useful bacteria.*
   a) The bacterial production...
   b) The industrial production...
   c) The industrial product...
   d) The bacterial product...

In the rest of the discussion we will examine:

- The range of morpheme forms for verbs and their **functional role**.
- The co-referential properties of these morphemes with respect the morpheme **type** and morpheme **form**.

The granulate components of a clause that an anaphor can co-refer to are:

- CLAUSE
- PREDICATE
- AGENT
- ACTION

- OBJECT/S

A verb used in a phrase can inflict a morpheme, which can take several different forms characterized by the suffix that is allowed by the language lexicon. Not all suffixes can be used with all verb inflictions. The range of suffixes that can be used with a particular verb is determined by the language lexicon[3] The suffix of a word determines its *morpheme type*, and these types have co-referential properties which correspond to their anaphoric usage. The morpheme types are associated with underlying semantics of the root word, which is the associative verb in our case. The inflicted forms of the verb "produce", for example, has the morpheme types "producer", "product", "producible" and "producing" while the verb "govern" has the forms "governor", "government", "governable" and "governing". The different morphemes of a word are meant to communicate the associated semantics of a word which in the case of verbs would be the related arguments of the action related by the verb. The verb "produce" has the morpheme "product" which forms the object argument while "producer" forms the agent argument. The complication in natural languages arises from the fact that not all the same morpheme types communicate the same associated semantics, neither do all root words have consistent morpheme forms. In the case of "govern" the agent argument is communicated by the morpheme "governor" whereas the verb "produce" has the morpheme "producer" serving the same purpose. In yet another case, the verb "abortion" has the morpheme "abortionist" for the agent argument. Hence we can not determine the **morpheme type** on the basis of the suffix used. In the same token, a single morpheme type is not consistently used for the same argument. The example in (2) shows that the same form of morpheme (ion suffix) is used for the purpose of two different arguments, that is, referring to two different components of the clause. We next look at the morpheme types of a verb, which are inflicted by addition of suffixes as well as the factors that influence the associated semantics which the morpheme can represent.

Table III gives a list of identified suffixes of morphemes which can result from adverb or verb normalizations. The classification of morphemes into adjectives and nouns is not straight forward since a some adjectives can be used as nouns and vice versa, eg. glamorous and honorable. The classifications in table III is based on majority usage from our content analysis of corpora. From a computational point of view any cross-usage can be easily determined using the role in which the morpheme is used, that is, either as a modifier or as the head-noun.

The last two columns of in table III show the component of the clause that a verb morpheme with specific suffixes can co-refer to. A verb can have inflicted forms which can function only as a noun or only as an adjective or in some cases both. The verb "capture" for example has the inflicted form "captive" which can function both as a modifier (eg. captive man) or as a a head-noun (eg. The captive). Similarly the other

[3]Originating from combining the basic sounds like vowels and consonants.

Suffix	Part-of-speech	Examples	Mod	H-N
-er, -or, -ie	noun	teacher, director, truckie	–	AGT
-ant, -ent	noun	applicant, president	–	AGT
-ess (not ness)	noun	waitress	–	AGT
-ist	noun	plagiarist, abortionist	–	AGT
-ian	noun	politician, beautician	–	AGT
-ee	noun	employee, trainee	–	OBJ
-uct	noun	product, conduct	–	OBJ
-ness	noun	abusiveness, sharpness	–	ACT
-ity	noun	security, stability	–	ACT
No suffix	noun	run, talk	–	ACT
-ment	noun	achievement, astonishment	–	PRED
-ance, -ence	noun	violence, attendance	–	PRED
-ion	noun	information, expression	–	CLS
-ism	noun	criticism, terrorism	–	CLS
-ship	noun	leadership, dictatorship	–	CLS
-able, -ible	adj	reliable, valuable	PRD	–
-al	adj	arrival, theoretical	PRD	ACT
-ful	adj	wonderful, helpful	PRD	–
-ish	adj	greenish, girlish	PRD	–
-ive	adj	creative, captive	PRD	OBJ
-ous, -ious	adj	glamorous,flirtatious	PRD	–
-ic	adj	hypnotic, theoretic	PRD	PRD
-ed, -en	adj	played, written	OBJ	–
-ing	adj	running, thinking	OBJ	–
No suffix	adj	wet, silent	ACT	–

TABLE III
TABLE SHOWING THE POSSIBLE SUFFIXES INFLICTED BY FINITE VERBS THAT CAN FUNCTION EITHER AS A HEAD-NOUN OR A MODIFIER.

morpheme types, which can function both as modifier as well as a head-noun are those with suffixes *al* and *ic*. The rest of the suffixes function either as the modifier or as the head-noun. The largest range of suffixes are reserved for the agent of a clause with some also exhibiting the gender property, eg. waiter and waitress. There are a handful of verbs which inflict morphemes for both the agent as well as the object, eg. "employ" inflicts "employer" for the agent and "employee" for the object. Apart from the specific suffixes which can co-refer to certain components of a clause, some verbs also allow non-inflicted forms to be used for either the noun or the adjective. The table contains the example for the verb "wet" which can be used as "The wet suit..." after the clause "John wet the suit". In this case the verb is used in an un-inflicted form as a modifier to refer to the object of the clause in which it was used as a verb. The corresponding example for the case of a noun is given for the verb "run". For instance, the phrase "The run will..." can grammatically correctly follow the clause "John will run the marathon". The verb "run" is used as a noun in exactly the same form as the one used in the clause but in this case it is coreferential to the action of the clause where it was first used.

Verb normalizations can be coreferential to different components of a clause. All verbs have a set of normalized forms of morphemes which vary depending on the semantics of the verb. The normalized forms can be categorized into different groups according to their coreferential properties and each of the groups have a range of suffixes which are constrained by the language lexicon. We have so far looked at the various components that the morpheme types can refer to. Next we

will look at the effect of use of either a subjective or an objective modifier to the co-reference properties of a verb inflicted morpheme. Lets look at the clause in (3). The verb "direct" has the inflicted form "director" which according to table III should be coreferential to the agent. Parts (3a) and (3b) show subjective modifiers used on the verb infliction while (3c) shows use of an objective modifier. In all the cases the co-reference property of the verb inflicted anaphor still refers to the agent of the clause, however the granularity of the semantics has slightly changed with the origin of the modifier. In the case of the objective modifier, the semantics is biased towards the predicate which involves direction of the play by the director. On the other hand, with the subjective modifier, the bias is towards the agent itself, with the modifier specifying further properties and/or relations pertaining to the agent which might be unrelated to the the predicate. In the case of (3a) and (3b) the modifiers "sergeant" and "Brent" merely further specify the *name* and *title* properties of the agent which don't have a direct relation to the predicate of the clause.

(3) *Sergeant Brent directed the play.*

a) The sergeant director...
b) Brent, the director...
c) The play director...

Table IV shows further examples of anaphora from verb normalization in the rest of the categories. The verb inflicted NP anaphora from each of the clauses are presented with a subjective and an objective modifier for categories coreferential to different components of a clause. The examples show that use of either a subjective or an objective modifier **does not** shift the coreferential properties of the anaphor. However, there is a change in bias of the underlying semantics associated with the anaphora. In the first case, for example "The opposition criticism" represents a bias towards the "opposition" criticizing the government, while "The government criticism" implies that the government was the object in the act of criticism (by someone). Hence in subjective modifier case, the bias is towards the agent while in the objective case it is towards the predicate. However in both cases the target of both the referents is still the complete act of "opposition criticizing the government". The last example of "The prisoner employee" and "The company employee" is perhaps more illustrative. Here again, the change from objective to subjective modifier has no effect on the coreferential properties of the anaphor even though there are covert effects on the semantics of the two referents.

From the discussion in this section two things are apparent. Firstly, there are no constraints governing use of either subjective or objective modifiers with verb inflicted referents. Use of either of the modifiers with verb inflicted referents does not shift the coreferential properties of the referent, however it shifts the associated semantics towards either the subject or the object. The second pertinent point established earlier is that the component of the clause to which a NP refers to, is determined by the type of morpheme which is characterized by the suffixes used which can be grouped to refer to different components of a clause. Furthermore, not all suffixes can be grammatically correctly used with all verbs. There are no overt language rules that govern this, hence the only tool available to computational systems is to cross check with the language lexicon. Section III-A presents the the empirical results used to arrive at the theory in table III.

Suffix	Mod	conf	H-N	conf
-er, -or, -ie	–		AGT	17 (1)
-ant, -ent	–		AGT	10 (0)
-ess (not ness)	–		AGT	6 (0)
-ist	–		AGT	9 (1)
-ian	–		AGT	13 (2)
-ee	–		OBJ	14 (3)
-uct	–		OBJ	7 (0)
-ness	–		ACT	12 (3)
-ity	–		ACT	16 (2)
No suffix	–		ACT	9 (2)
-ment	–		PRD	9 (1)
-ance, -ence	–		PRD	12 (3)
-tion	–		CLS	13 (0)
-ism	–		CLS	14 (2)
-ship	–		CLS	8 (1)
-able, -ible	PRD	5 (0)	–	
-al	PRD	6 (1)	ACT	12 (2)
-ful	PRD	6 (4)	–	
-ish	PRD	9 (2)	–	
-ive	PRD	12 (3)	OBJ	5 (1)
-ous, -ious	PRD	6 (1)	–	
-ic	PRD	8 (2)	PRD	4 (1)
-ed, -en	OBJ	2 (0)	–	
-ing	OBJ	9 (2)	–	
No suffix	ACT	11 (3)	–	
Total		74 (18)		190 (25)
Percentage		80.4 (19.6)		88.4 (11.6)

TABLE V
TABLE SHOWING THE CONFORMANCE LEVEL FOR NORMALIZATIONS. THE NUMBERS IN THE BRACKETS SHOW NON-CONFORMANCE

A. Method and results for normalization

The input data used was again the the corpora developed using 120 newspaper articles described in section II-A, however, in addition, manual searches were done for normalized usage of words with some suffixes that did not exist in the corpora. For this purpose we mainly used Corpus of Contemporary American English [14] which is an extensive corpora containing some 410 million words and it allows you to search for words with any pattern including the surrounding context in which the word appears. The results are detailed in table V.

The results in table V are derived from approximately 3000 sentences searched automatically by a text processor as well as approximately 500 sentences searched manually. The manual search was done using the Corpus of Contemporary of American English in order to search of examples that were missing in the in-house corpus used as input data for the larger nlp project. The conformance level of normalizations where the morpheme is a modifier is 80.4% which is slightly lower then when it is used as the head noun which has a conformance level of 88.4%. There were a total of 307 normalized forms examined, out of which 86% conformed to the anaphoric properties described in table III. Again this is encouraging

The opposition criticized government.	[The opposition criticism...]	CLAUSE
	[The government criticism...]	CLAUSE
The industry produces milk.	[The industrial production...]	PREDECATE
	[The milk production...]	PREDECATE
Durnham maintains roads.	[The Durnham maintenance...]	PREDECATE
	[The road maintenance...]	PREDECATE
The teacher announced sports.	[The The teacher announcement...]	PREDECATE
	[The sports announcement...]	PREDECATE
The teacher welcomed the parents.	[The teacher welcome...]	ACTION
	[The parent welcome...]	ACTION
The company employs the prisoners.	[The prisoner employees...]	OBJECT/S
	[The company employees...]	OBJECT/S

TABLE IV
EXAMPLES OF CLAUSES AND THE ANTECEDENT-COMPONENT OF THE CLAUSE BEING REFERRED TO BY THE VERB INFLICTED ANAPHORA CORRESPONDING TO EITHER SUBJECTIVE AND OBJECTIVE MODIFIERS.

result for a theory dealing with an nlp task and has been integrated with the anaphora resolution system, the results of which will be reported in forth coming publications. [15], [16], [17], [18], [19], [20], [21]

IV. CONCLUSION

NPs generated from single clauses can be accounted for by the generative processes, *predicate-deletion* and *normalization*. The process of predicate-deletion process can be constrained by the relations CAUSE, HAVE, MAKE, USE, BE, IN/ON, and FOR. We empirically determined that the entities connected by a relation other then these seven, can not be combined to form a NP with a confidence level of 89%. Similarly the NPs generated by normalization were examined to have anaphoric properties consistent with their suffixes, which was also empirically determined to have a conformance level of 86%. Additionally, the choice of subjective or objective modifier does not alter the anaphoric properties of the NP although it has subtle effect on the semantics of the NP. These are significantly useful results for studies engaged in developing rule based anaphora resolution systems in terms of implementation of rules as well as for accuracy of resolution targets.

REFERENCES

[1] J. Levi, "Where do all those other adjectives come from," in *Chicago Linguistic Society*, vol. 9, 1973, pp. 332–354.

[2] A. Garnham, *Mental models and the interpretation of anaphora*. Psychology Pr, 2001.

[3] J. Tetreault, "Analysis of syntax-based pronoun resolution methods," in *Annual meeting–Association for Computatinal Linguistics*, vol. 37. Association for Computatinal Linguistics, 1999, pp. 602–605.

[4] S. Lappin and H. Leass, "An algorithm for pronominal anaphora resolution," *Computational Linguistics*, vol. 20, no. 4, pp. 535–561, 1994.

[5] P. C. Gordon and K. A. Scearce, "Pronominalization and discourse coherence, discourse structure and pronoun interpretation." *Memory and Cognition*, vol. 23, pp. 313–323, 1995.

[6] R. B. Lees, *On a Transformational Analysis of Compounds: A Reply to Hans Marchand Lees, Robert B.*, W. P. S. ans Krahe, Ed. Berlin, New York (Walter de Gruyter): Indogermanische Forschungen, 1966.

[7] P. Downing, "On the creation and use of english compound nouns," *Language*, vol. 53, no. 4, pp. 810–842, 1977. [Online]. Available: http://www.jstor.org/stable/412913

[8] C. Fellbaum, *WordNet: An Electronic Lexical Database*. Bradford Books, 1998.

[9] B. Abbott, "Definite and indefinite," *Encyclopedia of language and linguistics, 2nd ed. New York: Elsevier*, 2005.

[10] I. Heim, "The semantics of definite and indefinite noun phrases," *Electronic Doctoral Dissertations for UMass Amherst*, 1982.

[11] R. Vieira and M. Poesio, "An empirically based system for processing definite descriptions," *Computational Linguistics*, vol. 26, no. 4, pp. 539–593, 2000. [Online]. Available: http://www.mitpressjournals.org/doi/abs/10.1162/089120100750105948

[12] V. Ng and C. Cardie, "Identifying anaphoric and non-anaphoric noun phrases to improve coreference resolution," in *Proceedings of the 19th international conference on Computational linguistics-Volume 1*. Association for Computational Linguistics, 2002, pp. 1–7.

[13] K. Fraurud, "Definiteness and the processing of noun phrases in natural discourse," *Journal of Semantics*, vol. 7, no. 4, pp. 395–433, 1990.

[14] M. Davies, "Corpus of contemporary american english," http://www.americancorpus.org/, 2010.

[15] S. N. Kim and P. Nakov, "Large-scale noun compound interpretation using bootstrapping and the web as a corpus." in *The Proceeding of Conference on Empirical Methods in Natural Language Processing, EMNLP.*, Edinburgh, UK, 2011.

[16] S. Tratz and E. Hovy, "A taxonomy, dataset, and classifier for automatic noun compound interpretation," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ser. ACL '10. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 678–687. [Online]. Available: http://portal.acm.org/citation.cfm?id=1858681.1858751

[17] P. Nakov, "Noun compound interpretation using paraphrasing verbs: Feasibility study," 2008. [Online]. Available: http://www.cs.berkeley.edu/~nakov/selected_papers_list/aimsa2008.pdf

[18] C. Butnariu and T. Veale, "A concept-centered approach to noun-compound interpretation," in *Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1*, ser. COLING '08. Stroudsburg, PA, USA: Association for Computational Linguistics, 2008, pp. 81–88. [Online]. Available: http://portal.acm.org/citation.cfm?id=1599081.1599092

[19] S. N. Kim and T. Baldwin, "Interpreting semantic relations in noun compounds via verb semantics," in *Proceedings of the COLING/ACL on Main conference poster sessions*, ser. COLING-ACL '06. Stroudsburg, PA, USA: Association for Computational Linguistics, 2006, pp. 491–498. [Online]. Available: http://portal.acm.org/citation.cfm?id=1273073.1273137

[20] V. S.-S. Nastase, J. Sokolova, and S. M. Szpakowicz, "Learning noun-modifier semantic relations with corpus-based and wordnet-based features," *Proceedings of the National Conference on Aritficial Intelligence.*, vol. 21, no. Part 1, pp. 781–787, 2006.

[21] K. Barker and S. Szpakowicz, "Semi-automatic recognition of noun modifier relationships," 1998.