

In-Depth Analysis of Anaphora Resolution Requirements

Helene Schmolz, David Coquil, Mario Döller

Presentation by: Helene Schmolz
helene.schmolz@uni-passau.de

TIR,
09/04/2012

What are anaphors?

- Example:

Tom plays the piano. He likes music.

antecedent

t

anaphor

- Anaphora:
“the use of a word which refers to, or is a substitute for, a preceding word or group of words”
(Simpson & Weiner 1989: 436)
- Anaphora resolution



Related work

- Existing corpora (examples)
 - Syracuse study
(600 abstracts from psychology and computer science)
 - Mitkov & Hallett
(technical manuals, newswire texts, literary texts)
- Existing anaphora resolution frameworks or methods (examples)
 - Hobbs's algorithm
 - Lappin & Leass's RAP



Research desiderata

- Thorough analysis of types of anaphors for English
- Examination of the frequency of each anaphor type
- Distribution of anaphor types in hypertexts
- Annotated corpora



Anaphor definition: 6 conditions

- 1) Anaphors refer back as well as forward, i.e. the antecedent either precedes or follows the anaphor.
- 2) Anaphors must have an explicit antecedent, i.e. an antecedent which occurs in the same text.
- 3) Anaphors are interpreted in relation to their antecedent.
- 4) The relation between anaphor and antecedent is coreferential, substitutional, or shows other, miscellaneous features.
- 5) The use of anaphors leads to a reduction of the text and/or avoids excessive repetition.

Examples: a) After she had come home, **Susan** answered her e-mails.

 b) Tom bought a blue **shirt**. Simon bought a green one.

Types of anaphors

Anaphor type	Examples
Central pronouns	<i>she, her, herself</i>
Reciprocal pronouns	<i>each other, one another</i>
Demonstrative pronouns	<i>this, those</i>
Relative pronouns	<i>who, which</i>
Adverbs	<i>when, there</i>
Noun phrases with “the”	<i>(the girl)</i>
Proper names	<i>(Mr Miller)</i>
Indefinite pronouns	<i>one, other, many</i>
Other forms of coreference and substitution	<i>the same, such, so</i>
Verb phrases with “do” and combinations	<i>do, do it, do this</i>
Ellipses	
Non-finite clauses	<i>-ing, -ed, to</i>

Types of anaphors – examples

Anaphor type	Example sentences
CENP	Tom plays the piano. <u>He</u> likes music.
RECP	Tom and Susan looked at <u>each other</u> .
DEMP	The mountains in Austria are higher than <u>those</u> in Germany.
RELP	The boy <u>who</u> is playing with the ball is my nephew.
ADV	Dad was in New York . He came back from <u>there</u> yesterday.
NPT	Betty repaired the lamp. <u>The girl</u> is only twelve years old.
PROPN	Bob Miller is at a meeting in Berlin today. In urgent cases you can call the secretary there – just ask for <u>Mr Miller</u> .
INDP	Tom bought a blue shirt . Simon bought a green <u>one</u> .
OCS	He is reliable and a team player . <u>Such qualities</u> are important for a career in computer science.
VPDO	Mary speaks French perfectly . At least, I think she <u>does</u> .
ELL	Betty left early in the morning and ___ forgot to wake them up.
NFC	The goals <u>scored</u> by the team were impressive.



Corpus of hypertexts – design

Classification by Rehm (2007) –
adaption:

- Wikipedia texts
- Blog texts
- Texts from traditional websites
 - homepages of companies
 - institutional homepages
 - personal homepages
 - online newspapers

In total: ~75,000 words



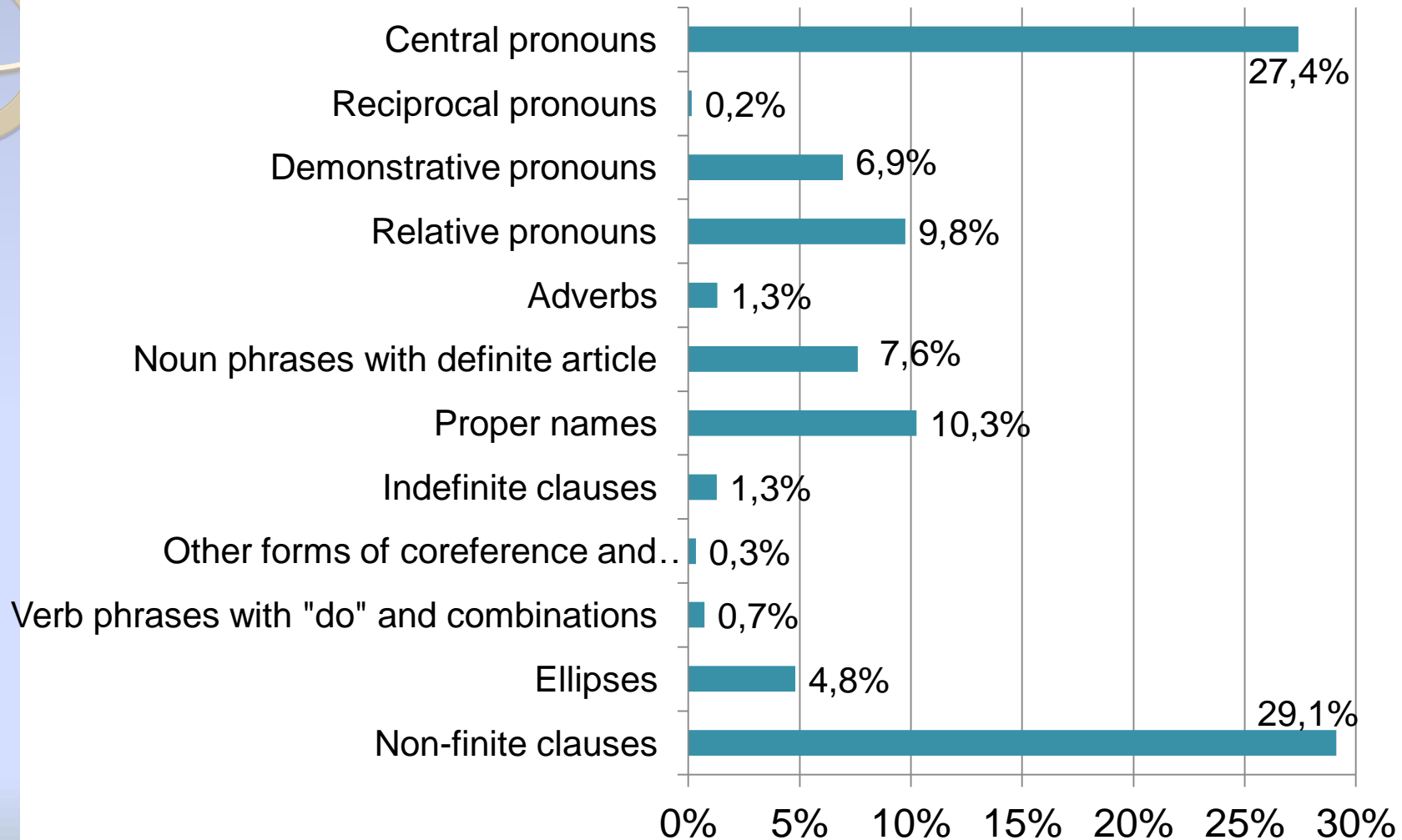
Corpus annotation

- Anaphor and its type
- Antecedent(s) of each anaphor
- Relationship between anaphor and antecedent

Website:

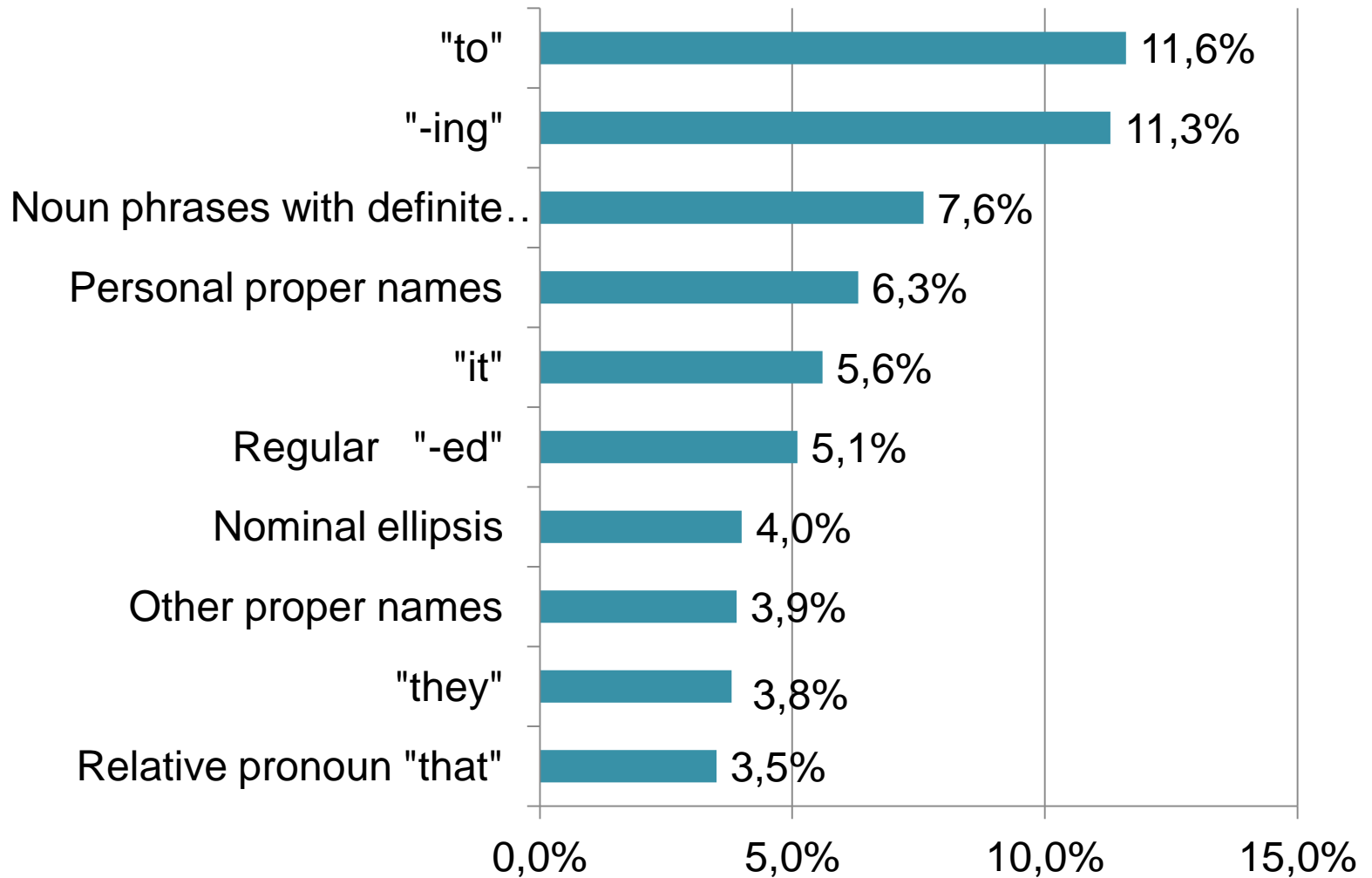
<https://www.dimis.fim.uni-passau.de/iris/index.php?view=anares>

Frequency of anaphor types



In total: 6,115 anaphors (80.5 anaphors / 1,000 words)

The 10 most frequent anaphor items



Ratio of anaphoric and non-anaphoric items

	Items in total	Central pronouns	Reciprocal pronouns	Demonstrative pronouns	Relative pronouns	Adverbs
Hypertexts (anaphoric)	6,115 (26.7 %)	1,676 (68.6 %)	10 (83.3 %)	424 (62.4 %)	596 (59.8 %)	80 (15.1 %)
Hypertexts (non-anaphoric)	16,798 (73.3 %)	769 (31.5 %)	2 (16.7 %)	256 (37.7 %)	401 (40.2 %)	450 (84.9 %)
	Noun phrases with definite article	Proper names	Indefinite pronouns	Other forms of coreference and substitution	Verb phrases with "do" and combinations	Non-finite clauses
Hypertexts (anaphoric)	466 (12.3 %)	627 (12.7 %)	78 (5.3 %)	21 (9.0 %)	44 (13.4 %)	1,780 (24.7 %)
Hypertexts (non-anaphoric)	3,314 (87.7 %)	4,295 (87.3 %)	1,394 (94.7 %)	212 (91.0 %)	285 (86.6 %)	5,420 (75.3 %)



Framework and future work

- Framework “Anares”: identifying potential anaphors and antecedents
- Expand the framework “Anares”:
e.g. knowledge about genus,
animate/inanimate
- Expand the corpus



Thank you for your attention!



Sources

- Hobbs, Jerry R. (1986, first published 1978), “Resolving Pronoun References”, in Barbara Grosz, Karen Sparck Jones & Bonnie Lynn Webber, eds. (1986), *Readings in Natural Language Processing*, Los Altos, California: Morgan Kaufmann, pp. 339-352.
- Lappin, Shalom & Herbert J. Leass (1994), “An Algorithm for Pronominal Anaphora Resolution”, *Computational Linguistics*, 20 (4), pp. 535-561.
- Liddy, Elizabeth DuRoss (1990), “Anaphora in Natural Language Processing and Information Retrieval”, *Information Processing & Management*, 26 (1), pp. 39-52.
- Mitkov, Ruslan (2001), “Outstanding Issues in Anaphora Resolution”, in Alexander Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing*, Berlin – Heidelberg: Springer, pp. 110-125.
- Mitkov, Ruslan & Catalina Hallett (2007), “Comparing Pronoun Resolution Algorithms”, *Computational Intelligence*, 23 (2), pp. 262-297.
- Quirk, Randolph *et al.* (2012, first published 1985), *A Comprehensive Grammar of the English Language*, London – New York: Longman.
- Rehm, Georg (2007), *Hypertextsorten. Definition – Struktur – Klassifikation*, Norderstedt: Books on Demand.
- Simpson, John A. & Edmund S.C. Weiner, eds. (1989, 2nd ed.), *The Oxford English Dictionary*, Oxford: Clarendon Press.
- Stirling, Lesley & Rodney Huddleston (2010), “Deixis and Anaphora”, in Rodney Huddleston & Geoffrey Pullum, eds., *The Cambridge Grammar of the English Language*, Cambridge: Cambridge University Press, pp. 1449-1564.

Knowledge used for rules

Non-anaphoric uses of non-finite clause items and other forms looking like non-finite clause items:

	nouns	gerunds	adjectives	prepositions	prepositional adverbs	as subject	in extraposition	part of verbs and other fixed expressions	simple finite verb phrases		complex finite verb phrases		
									present forms	past forms	be	have	modal verbs
“to”	x	x	x	✓	✓	✓	x	✓	x	x	x	x	x
“-ing”	✓	✓	✓	✓	x	✓ ¹	✓	x	✓	x	✓	✓	✓
“-ed”	✓	x	✓	x	x	✓	x	x	✓	✓	✓	✓	✓
irregular “-ed”	x	x	✓	x	x	✓	x	x	x	✓	✓	✓	✓

¹ anaphoric use is possible, but rare



Example of implemented rules

Processes for identifying potential anaphoric items of non-finite clauses for anaphora resolution:

1. Search for “to”-items, and all items ending in “-ing” and “-ed”.
2. Exclude non-anaphoric uses of these items as outlined above.
3. The remaining items are potentially anaphoric.

Example:

Text (from Wikipedia: Australia (continent), date of last access: 06/01/2012):

When the last ice age *ended* in about 10,000 BC, *rising* sea levels *formed* Bass Strait, separating Tasmania from the mainland. Then between about 8,000 and 6,500 BC, the lowlands in the north were *flooded* by the sea, separating New Guinea and Australia.

Identifying potential anaphors:

1. The items “ended”, “rising”, “formed”, “separating”, “flooded”, “separating” are found.
2. The items “rising” (adjective), “ended”, “formed” (past forms, simple finite verb phrases), “flooded” (“be”, complex finite verb phrases) are non-anaphoric and consequently excluded.
3. The two remaining items “separating” are anaphoric here.