'It could be lupus'

Identifying narrative event chains in clinical notes

19 June 2012

Phil Gooch

Centre for Health Informatics

City University London

Overview

- Brief introduction to the problem of coreference and anaphora resolution

- Why coreference resolution in clinical notes is important

  - Why standard approaches perform poorly

  - Protagonist theory, centering theory

  - Possible solutions and a few recent GATE plugins

- Analysis of results on ~1000 discharge summaries, progress reports, surgical, radiology and pathology reports

Along the way, we'll look at what can these guys ...

... and these guys ...

... can tell us about coreference and anaphoric relations

- **Coreference relation**: where two or more expressions refer to the *same* real-world entity. The semantic relation is identity
- **Anaphoric relation**: where a later expression (anaphor) has some semantic relation to an earlier expression (antecedent) and disambiguation of the anaphor is dependent on the antecedent.
  - Semantic relation may be identity, but not necessarily:
  - e.g. part—whole
- Often requires domain or world knowledge to resolve

Examples: pronouns and demonstratives

• **Starsky** pulled up in the Ford Torino. "Get in the car", **he** shouted to

  Hutch. "Gotta catch up with Huggy", *Hutch* yelled back, and *he*

  continued heading towards the club.

'mention pairs'

# Pronouns and demonstratives: mention-pair model

- **Pronominal coreference resolution**: usually consider gender and number agreement, plus recency, distance, grammatical role, person, position

- **Demonstrative and bridging coreference resolution**:

  - Same headword: 'lower extremity swelling … the swelling'

  - Hypernyms:

    'amoxicillin … the antibiotic'   'Staph bacteremia … the infection'

  - Synonyms: 'dyspnea … shortness of breath'

- Most work in the general domain has focused on resolving pronouns (e.g. Lappin & Leass 1994; Mitkov 1998), demonstratives (e.g. Vieira & Poesio 2001) or both (Raghunathan 2010) by considering mention pairs

Cardinal numbers as pronouns ...

<u>S</u>tarsky <u>o</u>ffered <u>H</u>utch **crisps** <u>a</u>nd <u>H</u>uggy <u>t</u>ook **one** <u>a</u>lso

opposite

hypotenuse
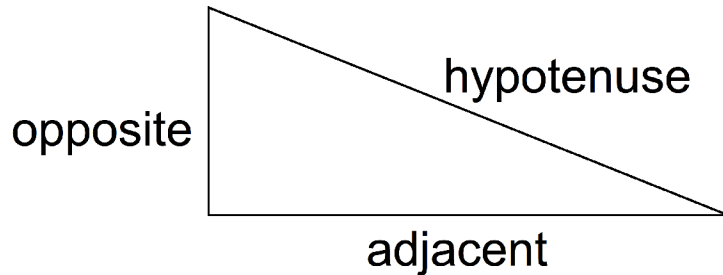
adjacent

sine = opposite/hypotenuse
cosine = adjacent/hypotenuse
tangent = adjacent/opposite

In addition to being a theorem mnemonic, resolving 'one' allows us to answer
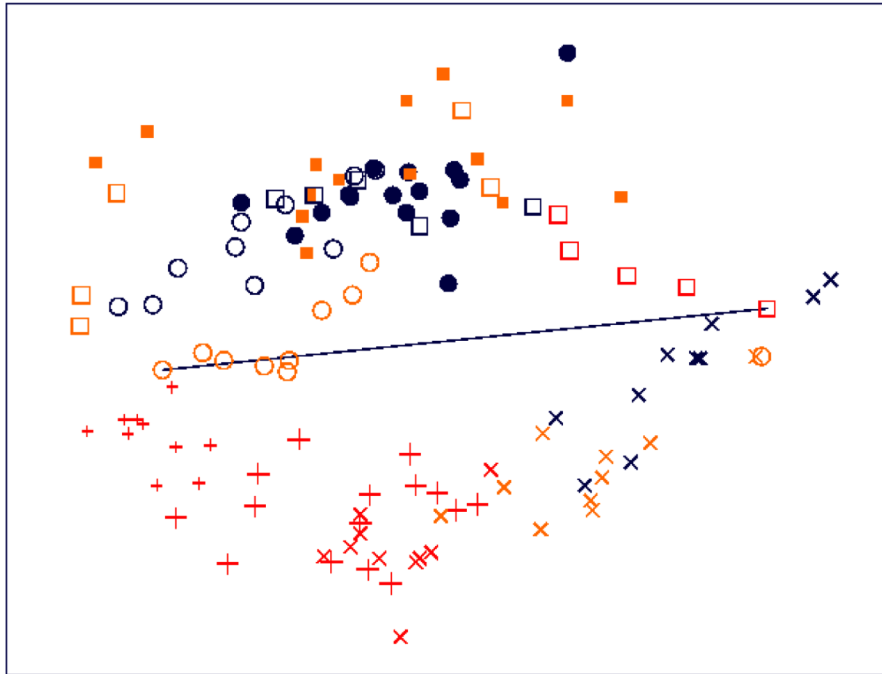
questions such as:

- What did Huggy take?

- Who got crisps?

Coreference resolution: Entity mention clustering model

Source: Wilhelm et al. 1999

- Considers clusters of terms and attempts to classify them into those that are coreferential and those that are not based on ranking (e.g. Rahman & Ng 2011)

Problems with it, that, this

- May not refer back to any specific mention

- Often used redundantly - 'pleonastic' *it*, *that, this*

    It is important to note *that* thresholds vary

    It would be difficult to treat the infection

- Can be difficult to distinguish from anaphoric 'it':

    Patient is taking **vancomycin**. **It** has been prescribed to treat the

        MRSA infection

    Patient is taking vancomycin. It has been difficult to treat the MRSA

        infection

'It could be lupus'. What is 'it' - diagnosis, symptoms, test results?

'*It* could be lupus ... increased platelet count can cause blood clots ...'

'Has he responded to **the latest lupus treatment**?'

'He's only been on **it** for a few hours...'

'ANA was negative, he hasn't responded because *it*'s not lupus'

Complex relations

- Clinical terms tend to be composed of multi-word expressions that can contain both coreferent and anaphoric relations:

[**the patient's** [<u>**head**</u> **wound laceration**]] … [**her** [<u>**scalp**</u> **laceration**]]

- *scalp* is anaphoric to *head* in a part—whole relation;

  wound is anaphoric to *laceration* in a synonym relation

- 'The patient's head wound laceration' and 'her scalp laceration' may be in a coreference relation

Coreference chains

- Neither the mention-pair model nor the entity cluster model solve the problem of generating narrative event *chains* of coreferential relations.

[G. House]_{Person}, a [[53 year-old]_{Age} male]_{Person}, [suffers]_{VP} from

[chronic thigh pain]_{Problem}. [The patient]_{Person} [takes]_{VP} [Vicoprofen]_{Treatment} for [this]_{Pronoun}

but [the medication]_{Treatment} [is not managing]_{VP}

[[his]_{Pronoun} discomfort]_{Problem}

directional narrative relationship
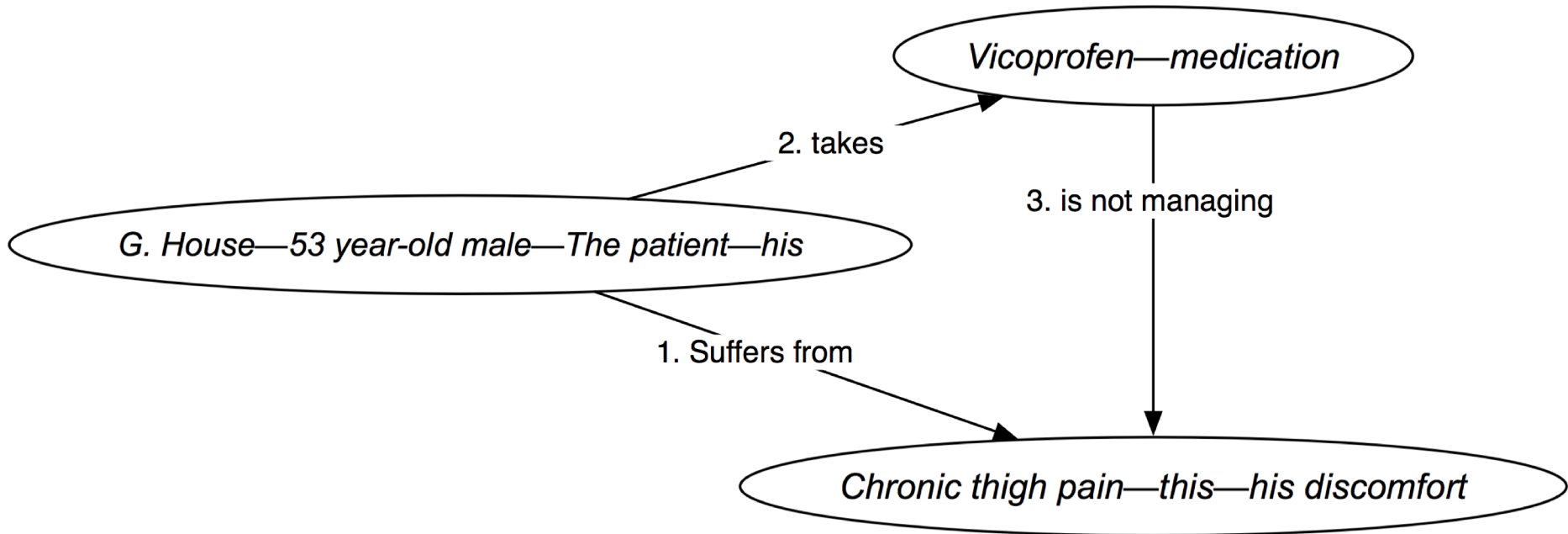
→

G. House—53 year-old male—The patient—his

Chronic thigh pain—this—his discomfort

Vicoprofen—medication

Narrative events

Vicoprofen—medication

2. takes

3. is not managing

G. House—53 year-old male—The patient—his

1. Suffers from

Chronic thigh pain—this—his discomfort

In this case the ordering of events isn't important, but they might be - e.g. dates of admission, discharge; order of tests, surgical procedures; addition of new medication regimes, etc.

Specific problems with clinical notes

- <u>Spelling inconsistencies</u> and errors (pateint, disciitis, pian)

- <u>Ambiguous abbreviations</u> without expansion (AMA, PT)

- <u>Anonymisation</u> - names replaced with random strings (can't infer gender from name)

- <u>Wide pronoun resolution scope</u> ('he' might refer to 'the patient' mentioned several sentences previously)

- <u>Domain knowledge</u> requirement ('abdomen' ... 'epigastric area')

- <u>Exact matches</u> might not be coreferent (family history of *hypertension* ... patient's *hypertension* managed with ACE inhibitor)
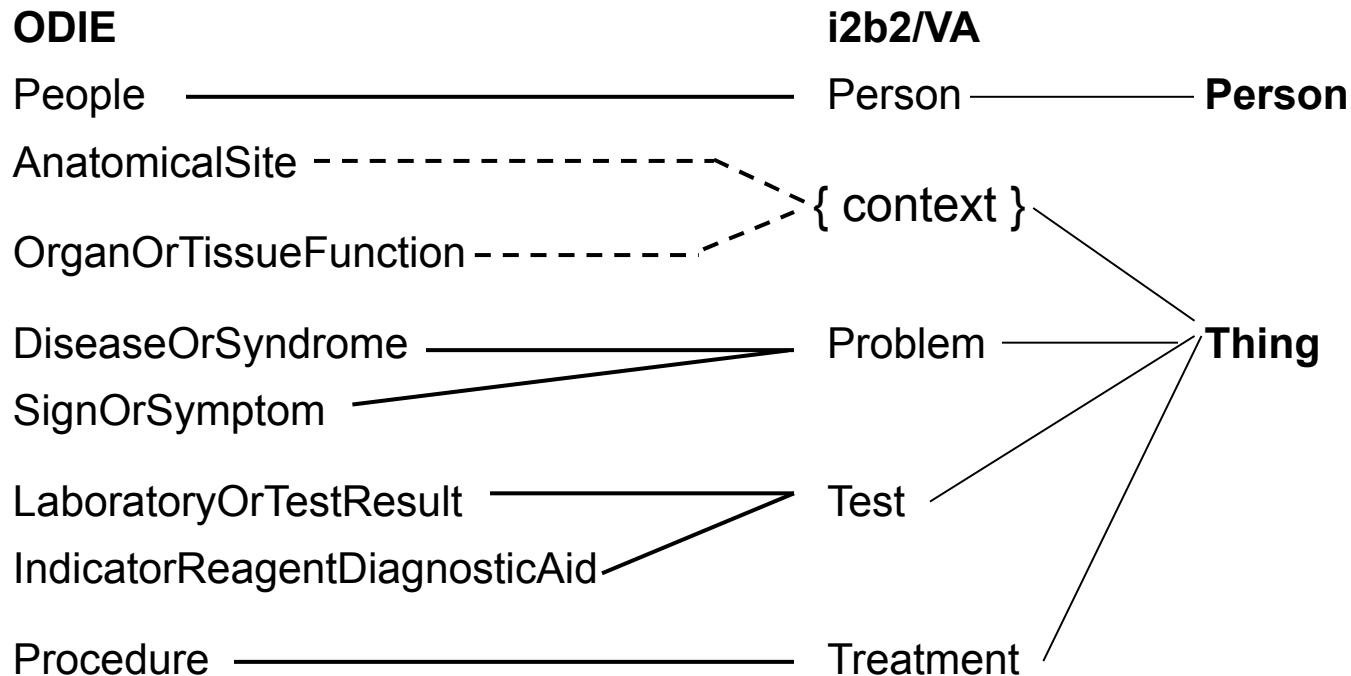
Coreference resolution in clinical notes

- Until recently, very little work in this area

- 2011 i2b2 NLP Challenge on Coreference Resolution

  - Release of manually annotated corpora (589 training, 388 test documents) of discharge summaries, progress notes, surgical, radiology and pathology reports from a range of institutions

  - General purpose, state-of-the art (Stanford NLP group) tools perform poorly on these corpora (F ~35%; Hinote 2011)

  - Corpus-specific tools not much better (F ~51%; Zheng 2012)

ODIE and i2b2/VA corpora

- Aim: same approach and code for both corpora and all document types

A lexical and knowledge-based approach to clinical coreference resolution

- Use techniques from previous research but tailored to specific entity classes

  - Combine ideas from both corpus-based and knowledge-based approaches to extract features for each entity class

  - *Centering* for resolution of personal pronouns; *protagonist* and associated verbs (Chambers & Jurafsky 2008)

  - Search in the direction of the narrative, i.e. forward, from mention to best pronoun and mention—mention (mention-pair approach) of the same class

  - Create coreference chains via dynamically pruned linked lists

Clinical coreference architecture
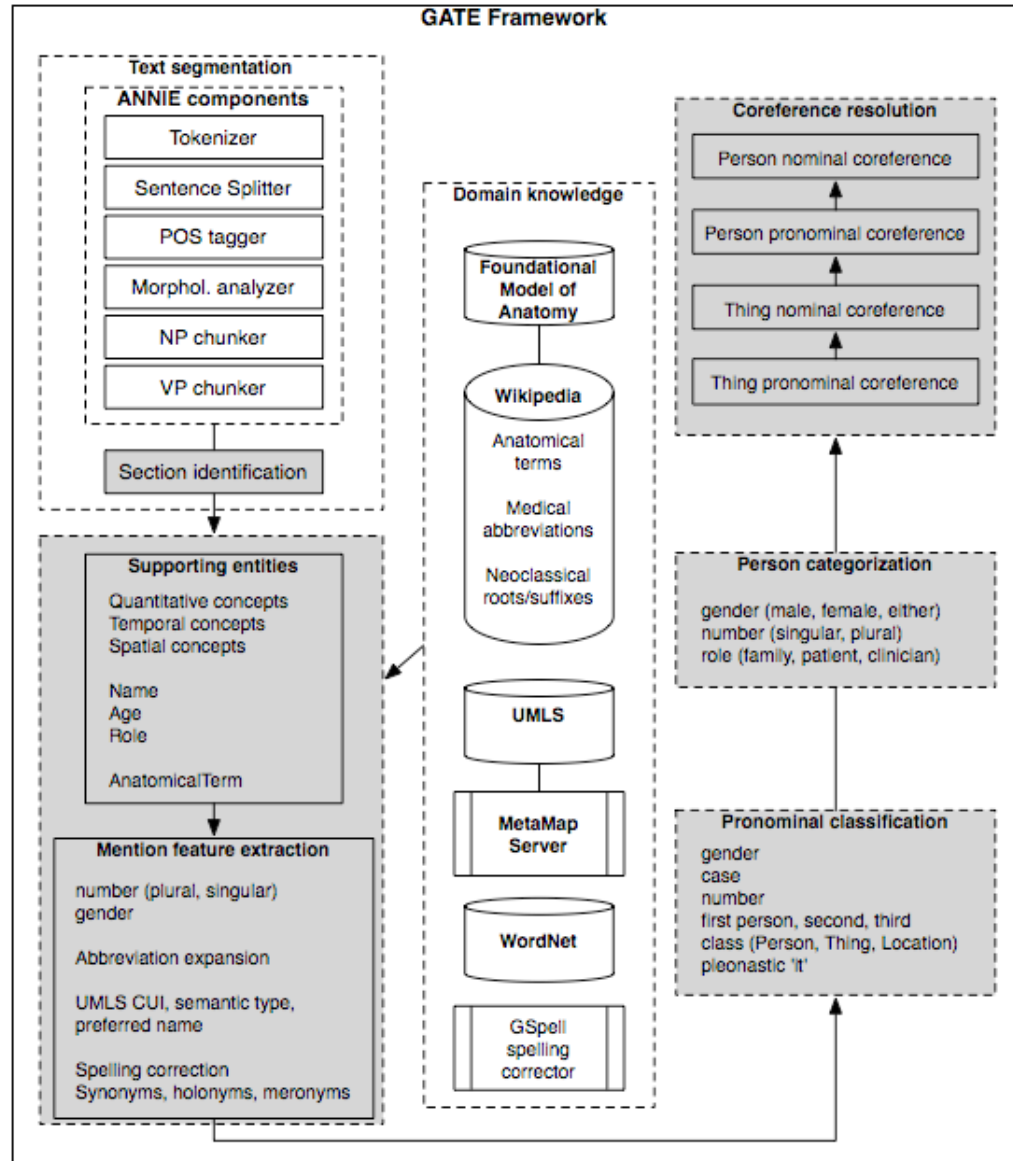
Source: Gooch & Roudsari 2012

Protagonist theory and classification of Person mentions

- Narrative events are centered on one or more key actors. Coreferring actors share congruent verbs, and distinct sets of verbs are typically associated with different actor types (Chambers & Jurafsky 2008)

- In our corpora, main protagonist is the patient (75% of all Person mentions, 86% of all personal pronoun mentions in training set)

- Narrative events are the admission, assessment, treatment and discharge planning processes which can be identified by certain verbs

- These verbs can help identify different protagonists and their roles (report author, report recipient, patient, family member)

Centering theory

- Mentions and their coreferent pronouns will occupy the same grammatical position in the sentence or clause

- E.g. given two potential mentions that could be the correct antecedent based on gender and number, if the pronoun is the subject of its clause, then select the mention that is also the subject of its own clause

- For Persons, only need to do this if the number of protagonists of different genders ≠ 1 or 2

Analysis of pronoun mentions in training set

Pronoun distribution in training set
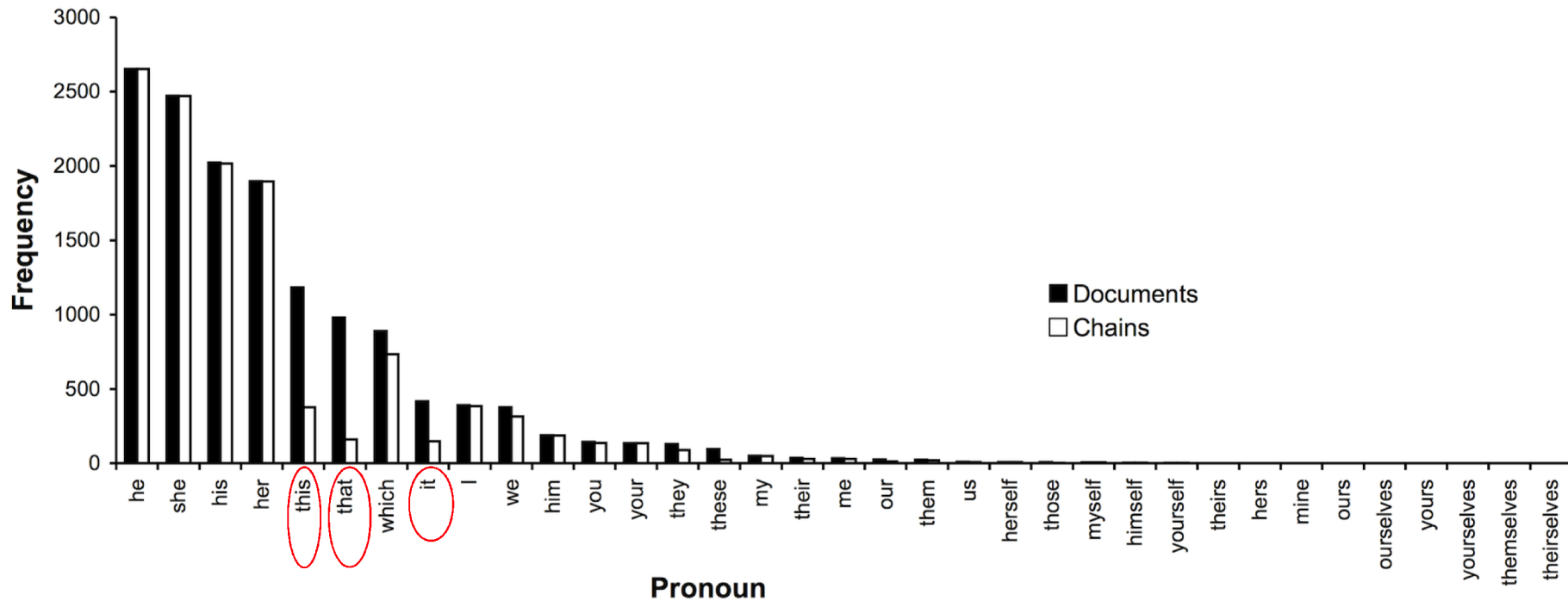
Analysis of pronoun mentions in training set

- Majority of 'this', 'that', 'it' mentions are *pleonastic* (dummy or expletive)

- All 'they' mentions are coreferenced against plural mentions - none used as gender-neutral singular

- Given that 86% of all he/she/his/her mentions refer to the patient, we can use the relative document frequency of these to infer the patient's gender, in the absence of all other cues.

Classification of nominal Person mentions

- Use gender, role and verb identifiers to classify Person mentions:

  - *Patient* - 'a 40 year old male' … 'was admitted'

  - *Patient's family or significant other* - 'his wife', 'her brother', 'the daughter'

  - *Clinician* - 'MD', 'attending', 'dictated', 'consulted'

    - Author

    - Recipient

      - Referred clinicians (social worker, external teams)

  - [Person]$_1$ [seen|treated|evaluated|treated…] by [Person]$_2$

Classification of Pronoun mentions

- Third person singular pronouns (he, she etc) that match the patient's gender have global, 'patient' scope by default

- First person pronouns are coreferenced to the report author

- Second person pronouns (you, your): either patient or recipient, depending on context ('your patient')

- Local exceptions:

   triggered by possessive pronoun ('his wife … she', 'his oncologist … he')

   appearance of a new protagonist ('social worker Barbara Cole … She')

- Other pronouns classified as anaphoric or pleonastic

GATE components for identifying potential bridging coreference

- **WordNet** (Miller 1995) Java API

    - Synonyms for Problem and SignOrSymptom classes: 'inflammation … swelling', 'chills … shivering'

    - Hypernyms for Problem, DiseaseOrSyndrome, Treatment, Procedure: 'bacteraemia … *infection*', 'biopsy … the *procedure*'

    - Meronyms for AnatomicalTerm: 'head … *scalp*'

- **MetaMap server** (Aronson 2001) + Java API

    - Headword identification

    - Term equivalence: 'dyspnea … shortness of breath'. Same CUI assignment or same UMLS preferred name

GATE components for identifying potential bridging coreference

- **GSpell Java API** (NLM Lexical Systems Group)

    - Adds correct spelling feature to mentions not picked up by MetaMap, e.g. 'disciitis→discitis'. MetaMap rematch attempted on mentions where spelling correction has occurred

- **SecondString Java API** (Cohen et al 2003)

    - Calculation of Monge-Elkan and Jaro-Winkler string similarity

    - E.g. 'Portacath' vs 'Port a catheter' - score = 0.92 (Jaro-Winkler)

- **Wikipedia** - identification and expansion of medical abbreviations

- **Foundational Model of Anatomy** (Rosse & Mejino 2003)

    - Identification of anatomical contexts of mentions

        [simple atheroma]$_{Problem}$ in the *aortic root*

        [simple atheroma]$_{Problem}$ in the *ascending aorta*

MetaMap plugin configuration

Name	Type	Required	Value
(?) annotateNegEx	Boolean	✓	false
(?) annotatePhrases	Boolean	✓	false
(?) excludeIfContains	ArrayList		[]
(?) excludeIfWithin	ArrayList		[Exclude]
(?) inputASName	String		
(?) inputASTypeFeature	String		normalizedString
(?) inputASTypes	ArrayList		[Mention.mentionClass==Problem
(?) metaMapOptions	String		-XQz -R SNOMEDCT,NCI
(?) outputASName	String		
(?) outputASType	String	✓	MetaMap
(?) outputMode	OutputMode	✓	HighestMappingOnly
(?) taggerMode	TaggerMode	✓	CoReference

GSpell plugin configuration

Name	Type	Required	Value
(?) excludeIfContains	ArrayList		[MetaMap]
(?) excludeIfWithin	ArrayList		[MetaMap, Exclude]
(?) filters	ArrayList		[[A-Z/\.\-]+, ^(\d+.+?)|(.+?\d+)$]
(?) inputASName	String		
(?) inputASTypeFeature	String		normalizedString
(?) inputASTypes	ArrayList		[Mention.mentionClass==Problem,
(?) maxCandidates	Integer	✓	2000
(?) maxEditDistance	Integer	✓	1
(?) mode	GSpell...	✓	PhraseTokens
(?) outputASName	String		
(?) outputASType	String		
(?) outputListFormat	GSpell...	✓	List

Contextual cue identification

- Pattern-based recognition of contextual cues

  - Document sections: history of present illness, family history, labs

  - Quantitative concepts: number, measurement

  - Temporal concepts: Date, time, duration, frequency, age

  - Spatial concepts: location and coordination (left, bilateral, anterior etc)

  - Role (family) and job title

- Ruling out coreference for same headword/phrase where contexts differ

  - 'chronic bilateral lower extremity swelling' *vs* 'right lower extremity swelling'

- Allowing coreference where the antecedent has more specific context than anaphor, but not vice versa

- Exclusion of coreference of non-Person mentions in family history sections, historical lab data

Feature enrichment with cues, UMLS, WordNet

bpocCUIs	[C1279572]
bpocPreferredNames	[Entire left knee]
bpocPreferredNamesHead	[knee]
context	Left Knee
coreferences	[]
form	singular
headCUIs	[C0024485]
headPreferredNames	[Magnetic Resonance Imaging]
headPreferredNamesHead	[Imaging]
mentionClass	Test
mentionString	MRI Left Knee w/o Contrast
normalizedString	MRI Left Knee without Contrast

antonyms	[hypotension]
hypernyms	[cardiovascular_disease]
hyponyms	[essential_hypertension]
synonyms	[high_blood_pressure, hypertension]
type	head

|	
mentionString	HTN
normalizedString	hypertension

UMLS features for 'MRI Left Knee w/o Contrast'   WordNet features for 'HTN'

Generation of coreference chains

- Taking the set of all mentions, create subsets according to mention class, and within each subset, compare pairs of mentions in document order.

- Once a coreference relation is established, the features of the antecedent are cloned to the anaphor and the antecedent is removed from the set of mentions to be compared

- The process then repeats from the previous anaphor which becomes the new antecedent

- This pruning, and limiting the sentence distance between candidate pairs, typically reduces the maximum number of comparisons per document from $\sim n^2/2$ to $\sim 2n$
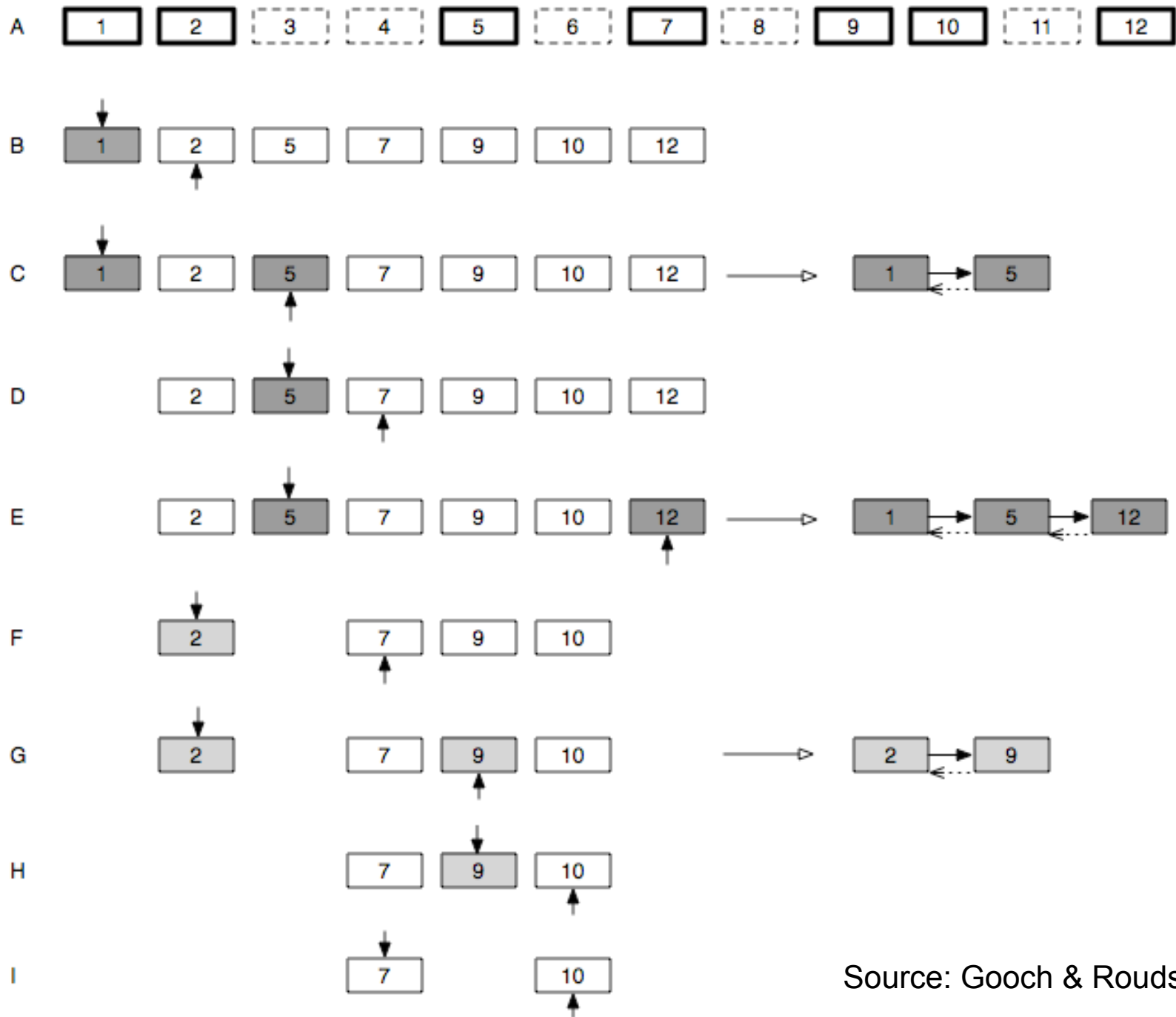
Source: Gooch & Roudsari 2012

Results - unweighted mean over evaluation metrics

	Training			Test		
Corpus	Precision	Recall	*F*	Precision	Recall	*F*
ODIE	0.771	0.828	0.796	0.765	0.827	0.792
i2b2	0.905	0.855	0.878	0.898	0.859	0.878

Performance analysis

- Deterministic rules cannot model discrepancies in the training data: inconsistencies in the coreferencing of names with their clinical role in both training sets

- Many synonyms not in WordNet or MetaMap, e.g. *confusion … delirium*; *ecchymosis … hematoma*; *carcinoma … tumor*; *unable to ambulate … bed bound*; *pins and needles from the knees ... neuropathic type pain*;

- Pathology reports difficult: *adenocarcinoma … exophytic mass*; *chemical stains … MLH1*

Future work

- Extend consideration of spatial, temporal and anatomical context of non-Person mentions to consider the events (verbs) surrounding them

  - Can be used to disambiguate singular, gender-neutral person 'they' *vs* plural thing 'they':

    'Patient's home **hypertensive medications** _were held_ while in the hospital. **They** _will be restarted_ on discharge'.

  - Might also help coreference terms where synonyms not available, e.g.

    'carcinoma … *forms* … mass'

- Improve identification of pleonastic 'it' - specific patterns (Dimitrov 2002)
- Use contextual features as input to a mention-pair classifier or cluster ranking model

Conclusions

- Semantically enriched mention-pairs with simple, forward-looking lexical rules can achieve good results with a variety of clinical reports

- GATE plugins developed for this task are available under GNU license from http://vega.soi.city.ac.uk/~abdy181/software/

  - Pronoun Classifier/Annotator/Coreferencer

  - WordNet Suggester

  - GSpell Spelling Suggester

  - BADREX Biomedical Abbreviation Expander

  - Baseline Coreferencer (for bridging coreference, coming soon :)

Postscript

- The process seems a bit 'heavyweight' - what happens if we remove some of the large knowledge base components (MetaMap/UMLS, WordNet, GSpell)?

- Results are a little surprising:

Corpus	Without domain knowledge			With domain knowledge		
	Precision	Recall	F	Precision	Recall	F
ODIE	0.743	0.836	0.781	0.765	0.827	0.792
i2b2	0.882	0.895	0.888	0.898	0.859	0.878

Postscript

- Overall, no significant effect on F-measure for either corpus when using these clinical domain knowledge resources

- Measurable increase in recall and reduction in precision in the i2b2 corpus but these cancel each other out, and in any case are reversed in the ODIE corpus

- Features that seem to be more important than synonym, meronym and hypernym matching are:

  - pronoun classification (gender, number, role, recency)

  - mention normalisation:

    - removal of leading determiners and pronouns

    - expansion of abbreviations

Postscript

- Plus:

  - temporal, spatial and anatomical contexts surrounding each mention

  - headword matching for mentions with similar contexts

  - approximate string matching

- The knowledge bases are probably not necessary for coreference resolution

  - The more interesting synonyms are not in there anyway!

    - See Performance Analysis

- But they are probably needed to identify and classify phrases as Disease, Symptom, Procedure in the first place